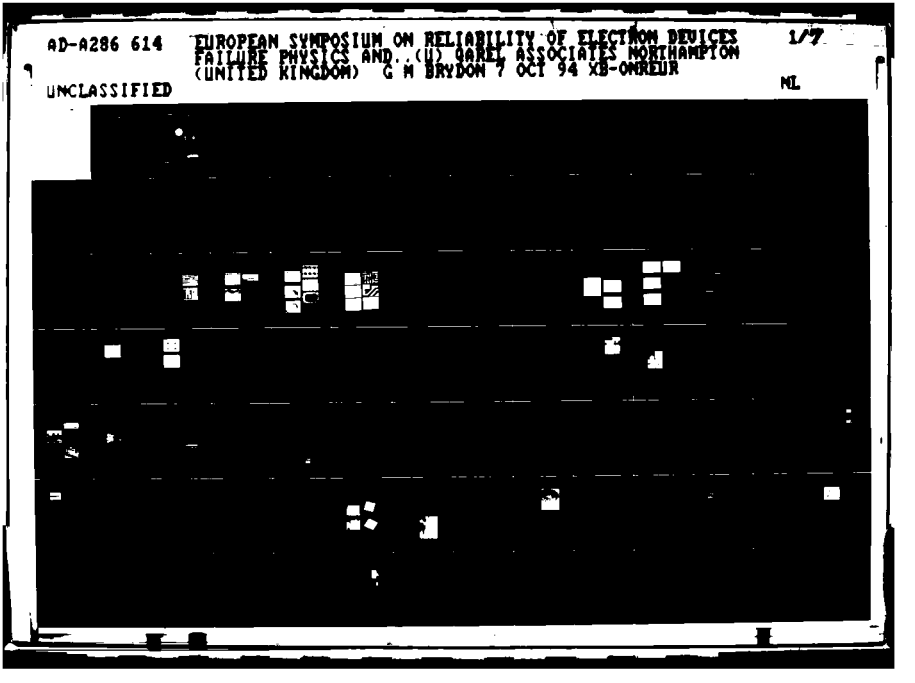
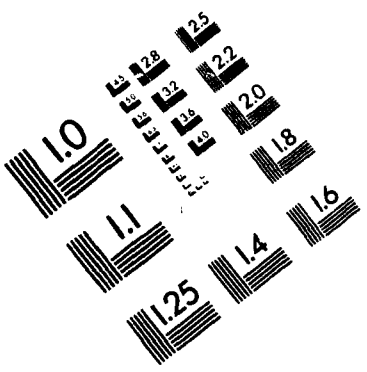
AD-A286 614 EUROPEAN SYMPOSIUM ON RELIABILITY OF ELECTRON DEVICES FAILURE PHYSICS AND..(U) OAKEL ASSOCIATES NORTHAMPTON (UNITED KINGDOM) G M BRYDON 7 OCT 94 XB-ONREUR

1/7

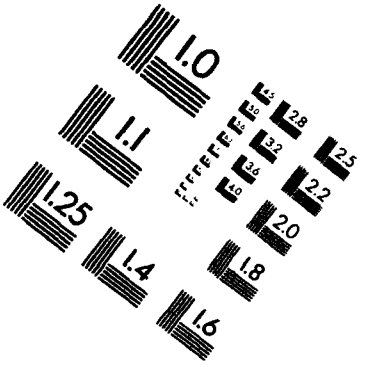
UNCLASSIFIED

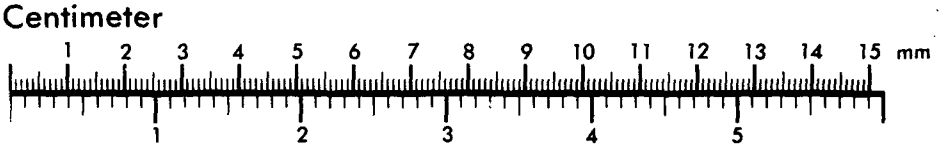
NL

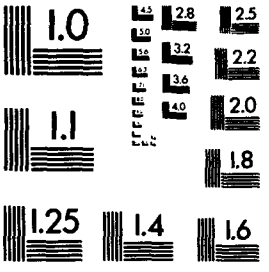


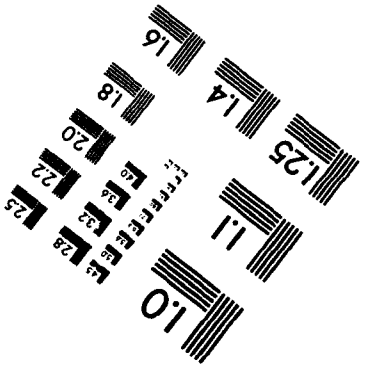


AIIM
Association for Information and Image Management
1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202

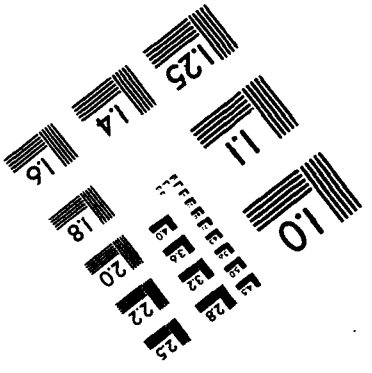








MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.



AD-A286 614

5th EUROPEAN SYMPOSIUM ON RELIABILITY OF ELECTRON DEVICES, FAILURE PHYSICS AND ANALYSIS

Glasgow, Scotland

4 - 7 October 1994

DTIC
ELECTE
NOV 2 9 1994
S G D

organised by
QaRel Associates

in collaboration with
IERI - Loughborough University of Technology

in co-operation with
AEI - CCTE - Italy
CECC
EC - IT R & D PROGRAMME, Microelectronics
GME - Germany
IEE - United Kingdom
IOP - United Kingdom
IQA - United Kingdom
RSG/DRA - United Kingdom
SEE - France

94-36295
5590
94 1128 052

sponsored by
The USA Office of Naval Research European Office

DTIC QUALITY INSPECTED 1

DRAFT SF 298

1. Report Date (dd-mm-yy) 7 Oct 94	2. Report Type conference proceedings	3. Dates covered (from... to) 4-7 October 1994
4. Title & subtitle 5th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis		5a. Contract or Grant # N00014--94-J-9015
		5b. Program Element #
6. Author(s) Mr. G. M. Brydon, editor		5c. Project # 5202
		5d. Task # EUR
		5e. Work Unit #
7. Performing Organization Name & Address QaRel Associates 12 Favell Way Weston Favell Northampton NN3 3BZ, United Kingdom		8. Performing Organization Report #
9. Sponsoring/Monitoring Agency Name & Address Office of Naval Research Europe PSC 802 BOX 39 FPO AE 09499-0700		10. Monitor Acronym ONREUR
		11. Monitor Report #
12. Distribution/Availability Statement A		
13. Supplementary Notes		
14. Abstract		
15. Subject Terms electron devices, failure physics		

Security Classification of			19. Limitation of Abstract	20. # of Pages	21. Responsible Person (Name and Telephone #)
16. Report Unclassified	17. Abstract Unclassified	18. This Page Unclassified	Unlimited	590	Mike Shear, 011-44-0171-514-4921

STEERING COMMITTEE

M.Barré	Matra Defense	France
G.M.Brydon	QaRel	United Kingdom
Y.Danto	University of Bordeaux	France
F.Fantini	University of Parma	Italy
W.Gerling	Siemens	Germany
O.Hallberg	Ericsson Telecom	Sweden
M.Hugen		EC
H.Maes	IMEC	Belgium
E.Montero		Spain
J.F.Verwey	University of Twente	The Netherlands

ORGANISING COMMITTEE

Conference Chairman

G.M.Brydon	QaRel	United Kingdom

Conference Vice Chairmen

A.Bendell	University of Trent	United Kingdom
R.F.B.Conlon	GMMT	United Kingdom

Technical Programme Chairman

J.A.Hayes	Loughborough University	United Kingdom

Technical Programme Vice Chairmen

K.Beasley	GMMT	United Kingdom
B.K.Jones	Lancaster University	United Kingdom

Conference Secretary

D.Brydon	QaRel	United Kingdom

Poster Paper Session Co-ordinators

Poster Session 1	J.A.Hayes (IERI, UK)
Poster Session 2	R.F.B.Conlon (GMMT, UK)
Poster Session 3	E.Montero (Spain)

Best Paper Award Chairmen

H.Maes	IMEC	Belgium
K.Beasley	GMMT	UK

Accesion For	
NTIS CRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

TECHNICAL PROGRAMME COMMITTEE

L.Adams	ESTEC	The Netherlands
W.Anderson	Naval Research Laboratory	USA
V.V.Aristov	USSR Academy of Science	USSR
E.Auvray	Thomson CSF	France
X.Aymerich	Barcelona University	Spain
A.Balogh	Scientific Soc. of Telecoms.	Hungary
J.Barrett	NMRC	Ireland
H.H.Berger	Technische Universität	Germany
A.Birolini	ETH	Switzerland
B.Bonati	Technopolis Csata	Italy
F.Borri	SGS Thomson	Italy
P.U.Calzolari	Bologna University	Italy
A.Christou	Maryland University	USA
M.Finetti		EC
M.Fukuda	NTT	Japan
G.Gregoris	Alcatel Espace	France
G.Groeseneken	IMEC	Belgium
F.Jensen	Reliability Consultant	Denmark
J.Kolodziejski	Inst. Elec. Technology	Poland
N.Kornilios	Hellas Inst. of Electronics	Crete
H.Koyama	Mitsubishi	Japan
J.Lloyd	Digital	USA
I.Lonzi	SGS Thomson	Italy
F.Magistrali	Pirelli Cavi	Italy
J.Sikula	Brno University	Czechoslovakia
N.Stjadinovic	University of Nis	Yugoslavia
R.W.Thomas	Technology Experts Network	USA
A.Touboul	Bordeaux University	France
L.K.Vandamme	Eindhoven University	The Netherlands
E.Wolfgang	Siemens	Germany
R.Woltjer	Philips	The Netherlands

SUPPORTING ORGANISATIONS

AEI/CCTE	Associazione Elettrotecnica e Elettronica Italia Circuiti Componente Tecnologia Elettroniche
CECC	CENELEC Electronic Components Committee
EC	The Commission of European Communities Information Technology Research and Development Programme, Microelectronics
GME	VDE/VDI - Gesellschaft Mikroelektronik
IEE	Institution of Electrical Engineers
IOP	Institute of Physics
IQA	Institute of Quality Assurance
PCS/DRA	Professional Component Services /Defence Research Agency
SEE	Société des Electriciens et Electroniciens

ESREF 93 AWARDS

On October 7th 1993 at the closing session of ESREF 93 in Arcachon, France, the following authors were awarded for their outstanding contributions.

BEST PAPER AWARD

Presented to:

W.Clayes, V.Quintard, S.Dilhaire, A.Hijazi and Y.Danto
(University of Bordeaux, France)

For the paper entitled:

Early Detection of Ageing in Solder Joints Through Laser Probe - Thermal Analysis of Peltier Effect.

BEST POSTER PAPER AWARD

Presented to:

C.Russ, H.Gieser, P.Egger and S.Irl
(Technical University of Munich, Germany)

For the paper entitled:

Compact Electro-thermal Simulation of ESD-protection Elements.

Authors' biographies appear with their papers in the ESREF Special Issue of Quality and Reliability Engineering International

CONFERENCE PROGRAMME

TUESDAY 4th OCTOBER

Tutorial 1	Field Data Collection and Analysis
Tutorial 2	Failure Analysis
Workshop 1	Reliability with Respect to SMEs
Opening Session	
Session A	Failure Analysis
Session B	Characterisation and Modelling

WEDNESDAY 5th OCTOBER

Session C	Hot Carriers
Session D	Oxide States
Session E	Power Devices
Workshop 2	Power Devices
Session F	Opto-electronics
Session G	Microwave
Standardisation Forum	

THURSDAY 6th OCTOBER

Session H	Reliability Methodologies
Session I	Electromigration
Session J	Packaging
Session K	ESD

FRIDAY 7th OCTOBER

Session L	Design for Reliability
Session M	Reliability Evaluation

Poster Session 1	AP, BP, CP, DP
Poster Session 2	EP, FP, GP, HP, IP
Poster Session 3	JP, KP, LP, MP

The Content of these Proceedings does not necessarily reflect the position or the policy of the United States Government and no official endorsement should be inferred.

CONTENTS

TUTORIALS

T1 **Reliability of Electronic Equipment Assessed by Field Data Collection and Analysis** 15
J.Mølitoft
The Danish Engineering Academy, Lyngby, Denmark

T2a **Advanced Techniques for Integrated Circuit Failure Analysis** 16
K. van Doorselaer
Alcatel Bell, Antwerp, Belgium

T2b **Advanced Failure Analysis of Semiconductor Devices** 16
M.Vanzi
University of Cagliari, Cagliari, Italy

WORKSHOPS

W1 **Reliability with Respect to Small to Medium Enterprises** 17

W2 **Power Devices Workshop "Reliability of Power Semiconductors for Traction Applications"** 18

OPENING SESSION

1 **Reliability, Which Strategy? (INVITED PAPER)** 19
P.D.T.O'Connor
British Rail Research, Derby, England

SESSION A Failure Analysis

A1 **Back SEM Inspection of Metallisation: A Technique for Failure Analysis and Reliability** 27
J.Barbero, E.Cordero, J.Soria
Alcatel, SESA, Madrid, Spain

A2 **Voltage Contrast Studies on 0.5µm Integrated Circuits by Scanning Force Microscopy** 33
C.Böhm, J.Sprengepiel, E.Kubalek
Gerhard-Mercator-Universität-GH-Duisburg, Duisberg, Germany

AP1 **EBIC Applications for Complex Devices** 39
A.Jurg
Philips Semiconductors, Nijmegen, The Netherlands

AP2 **Characterisation and Application of FIB/SIMS for Microelectronic Materials and Devices** 43
A.J.Leslie
Institute of Microelectronics, Singapore

AP3 **"Express" Failure Analysis of Electronic Components** 51
C.P.Strudwick
ERA Technology Ltd, Leatherhead, UK

SESSION B Characterisation and Modelling

B1 **A Method for the Calculation of the Soft-Error Rate of Sub-µm Dynamic Logic CMOS Circuits** 55
T.Juhnke, M.-P.Bringmann, H.Klar
Technische Universität Berlin, Berlin, Germany

B2 **Simulating Single Event Upset Rate in Large Digital Circuits** ... 61
P.Pavan, E.R.Minami, R.H.Tu, P.K.Ko, C.Hu
University of California, Berkeley, U.S.A.

B3 **Elementary Test Structures for Reliability Prediction** 65
V.Riviere [1], A.Touboul [1], S.Ben Amor [2], G.Gregoris [2], J.L.Stevenson [3], P.S.Yeung [3]
[1] IXL, University of Bordeaux, Bordeaux, France
[2] Alcatel Espace, Toulouse, France
[3] INTELSAT, Washington DC, USA

BP1 **Characterisation of Al(Cu)/TaAl Thin Films Used for Thermal Ink Jet Devices Using Atomic Force Microscopy and Transmission Electron Microscopy** 71
K.L.Pey [1], Y.Chen [2], W.K.Chim [3], W.Qin [1],
[1] Institute of Microelectronics, Singapore
[2] Fudan University, Shanghai, People's Republic of China
[3] National University of Singapore, Singapore

BP2 **Evaluation of Low-Voltage ZnO Varistors Quality by Low-Frequency Noise Characterisation** 77
F.Verdier, S.Gouverneur, C.Lucat, A.Touboul
IXL, University of Bordeaux, Bordeaux, France

BP3 **Semiconductor Structure Characterisation by Focus Electron and Optical Beams** 83
V.Aristov, E.Rau, E.Yakimov
Russian Academy of Sciences, Chernogolvka, Moscow District, Russia

BP4 **An Embedded Characterisation Structure for ASIC Failure Analysis and Performance Assessment** 85
C.Nouet [1], G.Rolland [2], M.Barré [1]
[1] Matra Defense, Velizy, France
[2] CNES, Toulouse, France

BP5 **Ion Implant Charging and Machine Fingerprinting Methodology** .. 91
C.Messick [1], S.Reno [1], J.Shideler [1], W.Lukaszek [2], W.Dixon [2], M.Vella [3], J.Reedholm Snr [4]
[1] National Semiconductor, West Jordan, USA
[2] Stanford University, Stanford, USA
[3] Lawrence Berkeley Labs., Berkeley, USA
[4] Reedholm Instruments, Freemont, USA

BP6 **The Use of DAVINCI for the Measurement of the Trap Energy Level and Surface State Density: A New Approach for Quality Evaluation of IC's** 93
I.Hachicha, P.Fouillat, J.P.Dom, H.Levi
IXL, University of Bordeaux, Bordeaux, France

BP7 **Analysis of Avalanche Breakdown of n-Channel MOSFETs Using Spectroscopic Emission Microscopy** 97
D.S.H.Chan, L.S.Koh, W.K.Chim, J.C.H.Phang
National University of Singapore, Singapore

SESSION C Hot Carriers

C1 **Understanding of the Hot-Carrier Degradation in Submicron MOSFET's from Uniform Injection Towards the Real Operating Conditions (INVITED PAPER)** 103
G.Groeseneken, R.Bellens, G. van den Bosch, H.E.Maes
IMEC. Leuven, Belgium

C2 **Pre-Breakdown Charge Trapping in High Field Stressed Oxides** 117
L.Dong, M.J.Tunnicliffe, V.M.Dwyer
Loughborough Univ. of Technology, Loughborough, U.K.

C3 **Investigation of Accuracy of Worst-Case Hot-Carrier Reliability Lifetimes Predicted by Berkeley's Model** 121
A.Meehan, P.O'Sullivan, P.Hurley, A.Mathewson
National Microelectronics Research Centre, Cork, Ireland

C4 **Evaluation of the Hot-Carrier-Induced Offset Voltage of Differential Pairs in Analog CMOS Circuits** 127
R.Thewes [1,2], M.Kivi [3], K.Goser [2], W.Weber [1]
[1] Siemens AG, München, Germany
[2] Universität Dortmund, Dortmund, Germany
[3] University of Liverpool, Liverpool, UK

CP1 **Normalised 1/f Noise: A More Sensitive Diagnostic Tool for Hot-Carrier Degradation in Submicron MOSFETs** 131
X.Li, L.K.J.Vandamme
Eindhoven University of Technology, Eindhoven, The Netherlands

CP2 **Hot Carrier Degradation in Submicron CMOS Technologies: A Comparative Study**..... 135
J.Vandenbroeck [1], H.Rempp [2], R.C.G.Swann [3]
[1] Alcatel Bell, Antwerpen, Belgium
[2] Alcatel Sel, Stuttgart, Germany
[3] Alcatel ITS, Zug, Switzerland

SESSION D Oxide States

D1 **Defect Monitoring and Layout Related Yield and Reliability Prediction for VLSI Interconnects** 143
A.Dreizner, H.Kück, K.Lukat, J.Pahlitzsch
Fraunhofer Institute of Microelectric Circuits and Systems, Dresden, Germany

D2 **Write/Erase Degradation and Disturb Effects in Source-Side Injection Flash EEPROM Devices** 147
D.Wellekens [1], J.Van Houdt [1], L.Faraone [2], G.Groeseneken [1], H.E.Maes [1]
[1] IMEC, Leuven, Belgium
[2] University of Western Australia, Australia

D3 **Modelling of the Threshold Voltage-Current Gain Relation in a Degraded I^2L Test Cell. Application for Improving the Reliability of the Technological Process** 155
O.Bonnaud [1], A.Seidou Maiga [1], L.Rolland [1,2], M.Taurin [2]
[1] University of Rennes, Rennes, France
[2] SGS Thomson Microelectronics, Rennes, France

D4 **Floating Gate p-MOSFET for the Mobile Ion Contamination Test** 159
J.C.Mitros, J.B.Finn
National Semiconductor Co., West Jordan, USA

D5 **Comparison of Inter-Polysilicon Oxide Lifetime Using Ramped and Constant Voltage Measurements** 163
A.Martin, P.O'Sullivan, A.Mathewson
National Microelectronics Research Centre, Cork, Ireland

DP1 **Frequency Dependence of Degradation and Breakdown of Thin SiO_2 Films** 169
M.Nafria, D.Yélamos, J.Suñé, X.Aymerich
Universitat Autònoma de Barcelona, Spain

DP2 **Gate Oxide Shorts in nMOS Transistors: Electrical Properties and Lifetime Prediction Method** 175
S.A.Dababneh [1], C.F.Hawkins [1], J.M.Soden[1 2]
[1] Univ. of New Mexico, Albuquerque, New Mexico, USA
[2] Sandia National Laboratories, Albuquerque, New Mexico, USA

SESSION E Power Devices

E1 **Reliability Issues in New Technology Implementation (INVITED PAPER)** 183
R.J.Francey
GEC Plessey Semiconductors, Lincoln, England

E2 **The 3-D Turn-Off Process of GTO Thyristors. Description of Failure Using Measurements and Simulation** 185
H.Bleichner [1], K.Nordgren [2], M.Rosling [1], P.Jonsson [1], E.Nordlander[1 2]
[1] Uppsala University, Uppsala, Sweden
[2] Gävle/Sandviken University College, Sandviken, Sweden

E3 **A Contactless Analysis Scenario for the Investigation of the Dynamic Behaviour of Power Semiconductors: Internal and External Laser Probing in Comparison with Computer Simulations** 189
G.Deboy [1], H.Brunner [1], G.Sölkner [1], W.Claeys [2], S.Dilhaire [2], V.Quintard [3], F.Koch [3]
[1] Siemens AG,München, Germany
[2] IXL, University of Bordeaux, Bordeaux, France
[3]Technische Universität München, München, Germany

E4 **Contribution to Failure Prediction Induced by Ageing due to Thermal Fatigue for Power Electronic Devices** 193
M.Khelif, G.Rojat, G.Clerc
CEGELY, Ecully, France

E5 **Identification of Performance Critical Steps in Power Semiconductor Processing** 199
C.M.Johnson [1], P.R.Palmer [2], D.A.Hinchley [2]
[1] University of Newcastle, Newcastle upon Tyne, UK
[2] University of Cambridge, Cambridge, UK

EP1 **Failure Mechanisms in Smart Power IC's** 205
L.Lonzi, M.Piccoli, F.Speroni
SGS Thomson Microelectronics, Cornerado (MI), Italy

EP2 **High Power Semiconductor Reliability Evaluation Based on Traction Vehicle Failure Analysis** 211
G.Bifulco [1], F.Fasce [2], M.Pasqualetti [2], M.Portesine [2], P.Zani [2]
[1] Ansaldo Trasporti, Napoli, Italy
[2] Ansaldo Trasporti, Genova, Italy

EP3 **Noise Diagnostics of PN Junction Power Devices** 217
J.Sikula, B.Koktavy, P.Vasina, Z.Chobola, P.Schauer, M.Korenska
Technical University of Brno, Brno, Czechoslovakia

EP4 **Accelerated Test of Power MOSFET** 221
M.E.Jimenez, E.Nogueira, Q.Perez, J.E.Vila
Telefonica Investigacion y Desarrollo, Madrid, Spain

SESSION F Opto-electronics

F1 **Behaviour of Lasing Wavelength Change During Degradation in Buried-Heterostructure (BH)-Type DFB Laser** 227
M.Fukuda, F.Kano, T.Kurosaki, J.Yoshida
NTT Opto-electronics Laboratories, Kanawaga, Japan

F2 **Reliability Assessment of a 1480nm MQW Pump Laser Diode for Erbium Doped Fibre Amplifiers** 231
A.Hawkridge, A.Carter
Hewlett-Packard Fibre-Optic Components Operation, Ipswich, U.K.

F3 **The Thermal Resistance: A Reliability Monitor for Laser-Diode Modules** 237
P.Malberti, M.Ciappa
Swiss Federal Institute of Technology, Zürich, Switzerland

F4 **Blue Shift and Mirror Degradation in InGaAs/GaAs Strained Quantum Well Lasers** 245
L.Serra, G.A.Azzini, P.Montangero
CSELT SpA, Torino, Italy

F5 **Reliability Results on 980nm High Power Lasers** 251
F.Magistrali [1], A.Righetti[1], G.Salmini[1], W.J.Gignac[2]
[1]Pirelli Cavi, Milan, Italy
[2] SDL, San Josè, CA, USA

F6 **Influence of Package Atmosphere on 980nm High Power Chip Reliability** 257
P.Whitney[1], D.Darby [1], L.Scheffel [1], D.Flanders [1], R.Pusch [2]
[1]Lasertron Inc., USA
[2]Alcatel SEL, Stuttgart, Germany

SESSION G Microwave

G1 **Hot-Electron-Induced Effects, Light Emission, Breakdown and Reliability Problems Phenomena in GaAs MESFET's, AlGaAs/GaAs HEMTs and AlGaAs/InGaAs Pseudomorphic HEMTs (INVITED PAPER)** 261
E.Zanoni, A.Noviani, G.Meneghesso, E.De Bortoli, L.Vendrame, A.Rizzato
University of Padova, Padova, Italy

G2 **Quality Evaluation of AlGaAs/GaAs and AlGaAs/InGaAs/GaAs HEMTs by LF Excess Noise Analysis** 273
N.Saysset [1], C.Maneux [1], N.Labat [1], A.Touboul [1], Y.Danto [1], J-M.Dumas [2]
[1] IXL, University of Bordeaux, Bordeaux, France
[2] ENSIL, Limoges, France

G3 **GaAs MMIC Switch Reliability** 279
W.T.Anderson, J.A.Roussos, K.A.Christianson
Naval Research Laboratory, Washington DC, USA

G4 **Picosecond Electro-optic Sampling System with Improved Spatial Resolution using the Technique of a Near Field Scanning Optical Microscope** 285
C.Böhm, J.Bangert, W.Mertin, E.Kubalec
Gerhard-Mercator-Universitat, Duisberg, Germany

GP1 **Pseudomorphic HEMTs Reliability with Scattering and Noise Parameters** 291
P.Conti, P.Audagnotto
CSELT S.p.A., Torino, Italy

SESSION H Reliability Methodologies

H1 **Reliability in Automotive Electronics, A Review (INVITED PAPER)** 295
R.Rathbone
Elmos, Dortmund, Germany

H2 **In-Situ Failure Detection in Thick Film Multilayer Systems** 309
J.Manca [1], L.De Schepper [1 2], W.De Ceuninck[1 2], M.D'Olieslaeger [1], L.M.Stals [1]
[1]Limburg University Centre, Diepenbeek, Belgium
[2]DESTIN N.V.,Diepenbeek,Belgium

H3 **Optical Ammeter for Integrated Circuit Characterisation and Failure Analysis** 315
W.Claeys [1], S.Dilhaire [1], D.Lewis [1], V.Quintard [1], T.Phan[1], J.L.Aucoutourier [2]
[1] CPMOH, University of Bordeaux, Bordeaux, France
[2] IXL, University of Bordeaux, Bordeaux, France

H4 **Process Maturity Grids Used as a Decision Support Tool** 321
R.G.J.Arendsen[1 2], A.C.Brombacher[1 2 3], B.Habraken[2], O.E.Herrmann[1]
[1] Twente University, Enschede, The Netherlands
[2] Philips Consumer Electronics, Eindhoven, The Netherlands
[3] Technical Univ Eindhoven, Dept. Mechanical Eng., Eindhoven, The Netherlands

HP1 **Reliability Assessment of Potentiometers by 1/f Noise Analysis** . 327
E.P.Vandamme, L.K.J.Vandamme
Eindhoven University of Technology, Eindhoven, The Netherlands

HP2 **Reliability Test Results on Coated and Uncoated PCBs** 331
J.A.Adell, E.Nogueira, J.E.Vila
Telefonica Investigacion y Desarrollo, Madrid, Spain

HP3 **Observation of Simultaneous Extrusions and Voids, in Tungsten-Filled Sub-Micron Vias, Induced During High-Temperature Processing** 337
J.A.Walls
Digital Equipment Corporation, South Queensferry, U.K.

SESSION I Electromigration

I1 **Computer Analysis of Microscopic Electromigration Failure Models** 343
J.T.Trattles, A.G.O'Neill, B.C.Mecrow
University of Newcastle, Newcastle upon Tyne, U.K.

I2 **Shape Changes of Voids in Bamboo Lines: A New Electromigration Failure Mechanism** 349
O.Kraft. U.E.Möckl E.Arzt
Universität Stuttgart, Stuttgart, Germany

I3 **Activation Energy of the early stages of Electromigration in Al-1% Si/TiN/Ti Bamboo Lines** 355
I.De Munari [1], A.Scorzoni [2], F.Tamarri [2], F.Fantini [1]
[1] Universtità di Parma, Parma, Italy
[2] CNR - Instituto LAMEL, Bologna, Italy

I4 **Electromigration Behaviour of On-Chip Interconnects: Influence of Stress Parameters Studied by Early Resistance Changes** 361
V.D'Haeger[1], W.De Ceuninck[1 2], H.Stulens[1],L.De Schepper[1 2], L.M.Stals[1]
[1]Limburg University Centre, Diepenbeek, Belgium
[2] DESTIN N.V., Diepenbeek, Belgium

I5 **Failure Locations in Different Via Structures due to Electromigration** 365
K.Weide, W.Hasse
University of Hannover, Hannover, Germany

IP1 **The Evolution of the Microscopic Damage in Electomigration Studied by Multiple Electrical Measurements** 371
B.K.Jones, Y.Z.Xu, P.Zobbi
Lancaster University, Lancaster, U.K.

IP2 **Electromigration in Al Based Stripes: Low Frequency Noise Measurements and MTF Test** 377
P.E.Bagnoli, C.Ciofi, B.Neri, G.Pennelli
Elettronica, Informatica, Telecomunicazioni, Pisa, Italy

IP3 **Electromigration Modelling as a Tool for Prediction of the Failure Behaviour of Metallisation Lines** 383
H.Hoebbel, V.Breternitz
Technical University of Ilmenau, Ilmenau, Germany

SESSION J Packaging

J1 **Plastic Packaging is Highly Reliable (INVITED PAPER)** 387
N.Sinnadurai
Middlesex University Microelectronics Centre, UK

J2 **Optical Current Density Probing at Interfaces in Bondings and Solder Joints: Investigation of Ageing Mechanisms** 399
W.Claeys [1], S.Dilhaire [1], D.Lewis [1],V.Quintard [1], T.Phan [1],A.Fenech [2], W.Hasse [3], K.Weide [3]
[1] CPMOH, University of Bordeaux, Bordeaux, France
[2] IXL, University of Bordeaux, Bordeaux, France
[3]University of Hannover, Hannover, Germany

J3 **Determination of Thermomechanical Behaviour of Microelectronics Packaging Based on Microstructural Analysis** ...405
A.Fenech, A.Hijazi, Y.Danto
IXL, University of Bordeaux, Bordeaux, France

J4 **Focus on Failure Modes of Plastic Devices after Temperature and Humidity Accelerated Sequential Tests** 411
N.Barlassina, M.F.Bedos, M.Brizoux, G.Deleuze
Thomson - CSF/SCTF, Orsay, France

J5 **The Demise of Plastic Encapsulated Microcircuit Myths** 417
E.B.Hakim [1], R.K.Agarwal [2], M.Pecht [3]
[1] Army Research Laboratory, Fort Monmouth, NJ, USA
[2] DELCO Electronics, Kokomo, IN, USA
[3] CALCE, University of Maryland, MD, USA

JP1 **Fracture Mechanical Characterisation of IC-Device-Interfaces** 423
K.Burger
TEMIC TELEFUNKEN Microelectronic, Heilbronn, Germany

JP2 **Ionizing-Radiation and High-Temperature-Induced Oxidation of the Electronic Packaging Materials of GE-HDI Electronic Board** 435
M.Al-Sheikhly, A.Christou
University of Maryland, Maryland, MD, USA

JP3 **Chip Surface Damage Induced by Internal Stress of Lead-On-Chip (LOC) Packages** 441
M.Amagai
Texas Instruments Japan, Oita, Japan

SESSION K ESD

K1 **Survey on Electrostatic Susceptibility of Integrated Circuits (INVITED PAPER)** 447
H.Gieser[1], I.Ruge[2]
[1] Fraunhofer Institute for Solid State Technology, Munich, Germany
[2] Technical University of Munich, Munich, Germany

K2 **ESD-Protection Elements During HBM Stress Tests - Further Numerical and Experimental Results** 457
C.Russ[1], H.Gieser [2], K.Verhaege [3]
[1] Technical University of Munich, Munich, Germany
[2] Fraunhofer Institute for Solid State Technology, Munich, Germany
[3] IMEC, Leuven, Belgium

K3 **A Physical Model for the Creation of Latent Gate Oxide Defects by Very Fast Electrostatic Discharges** 467
J.C.Reiner
ETH, Zürich, Switzerland

KP1 **Comparison of Transient Suppressor Devices** 473
E.Mino, J.E.Vila
Telefonica Investigacion y Desarrollo, Madrid, Spain

SESSION L Design for Reliability

L1 **Component Robustness; the Key to Reducing Field Failures (INVITED PAPER)** 479
F.Jensen
Finn Jensen Reliability Consultancy ApS, Holte, Denmark

L2 **Critical Area Analysis for Design Based Yield Improvement of VLSI Circuits** 487
D.Schmitt-Landsiedel [1], D.Keitel-Schultz [1], J.Khare [2], S.Griep [1], W.Maly [2]
[1] Siemens AG, München, Germany
[2] Carnegie Mellon University, Pittsburgh, U.S.A.

L3 **Influence of Gate and Tunnel Oxide Defectivity on EEPROM Yield and Reliability** 493
E.Ciantar [1,2], P.Boivin [1], M.Burle [1], J.Oualid [2]
[1] SGS-Thomson Microelectronics, France
[2] ENSPM, University of St-Jérôme, Marseille, France

L4 **Reliability Test Structures for ASIC Technology Assessment** 499
P.Weiss [1], B.Johlander [2], J.King [3]
[1] Swiss Centre for Electronics and Microtechnology, Neuchatel, Switzerland
[2] ESTEC, The Netherlands
[3] GEC Plessey Semiconductors, UK

L5 **Failure-Analysis-Based Test Chip Design for Quick Yield Improvement** 505
C.Hashimoto, S.Nakajima
NTT LSI Laboratories, Kanagawa, Japan

LP1 **ASIC Design House Approval - The Evolution of a Standard** 511
C.N.Robinson [1] A.Birks [2]
[1] PRQ, Edinburgh, UK
[2] PCS/DRA, Bromley, UK

SESSION M Reliability Evaluation

M1 **Planning Early Life Reliability Testing Using the 'Hughes' Model** 517
M.Shaw
IBM, Greenock, U.K.

M2 **Life Time Determination Based on Degradation Kinetics** 523
L.Tielemans [1], L.De Schepper [2]
[1] Destin NV, Diepenbeek, Belgium
[2] Limberg University Centre, Diepenbeek, Belgium

M3 **A Physics-of-Failure Approach to Addressing Device Reliability in Accelerated Testing** 527
P.Lall[1], M.G.Pecht[1], M.J.Cushing[2]
[1] CALCE, University of Maryland, MD, USA
[2] US Army Material Systems Analysis Activity, Aberdeen Proving Ground, MD, USA

MP1 **Fault Detection in Analogue Circuits by Dynamic Supply Current Monitoring** 541
M.A.Al-Qutayri
De Montfort University, Leicester, U.K.

MP2 **Testing of Flash EEPROM Structures by means of Co^{60} γ Irradiation and Constant Current Stress** 545
A.Paccagnella[1], G.Crisenza [2], G.Ghidini [2], A.Rizzato [3], A.Scarpa [3], E.Zanoni[3], P.G.Fuochi [4]
[1] Università di Cagliari, Cagliari, Italy
[2] SGS-Thomson Microelectronics, Agrate Brianza, Italy
[3] Università di Padova, Paova, Italy
[4] CNR-FRAE, Bologna, Italy

MP3 **Upper Voltage and Temperature Limitations of Stress Conditions for Relevant Dielectric Breakdown Projections** 551
R.-P.Vollertsen [1], W.W.Abadeer [2]
[1] Siemens Components, Essex Junction, U.S.A.
[2] IBM Microelectronics Division, Essex Junction, U.S.A.

MP4 **Noise Investigation of Metal-GaAs Ohmic Contacts** 557
Zs.Kincses, B.Kovdcs, I.Mojzes
Technical University of Budapest, Budapest, Hungary

MP5 **Identification of Faults in Analog Circuits** 561
Y.Maidon, P.Fouillat, Ch.Bouvet, S.Lesage
IXL, University of Bordeaux, Bordeaux, France

MP6 **A Comparison of Static and Dynamic Techniques for the Detection of Gate Oxide Shorts in Digital MOS Circuits** 565
L.Dong, V.M.Dwyer
IERI, Loughborough Univ. of Tech., Loughborough, UK

MP7 **Estimation of System Reliability Using a Non-Constant Failure Rate Model** 571
J.A.Hayes, J.A.Jones, M.Zahid
IERI, Loughborough Univ. of Tech.,Loughborough, UK

MP8 **BIST & Diagnosis in Safety Critical Micro-Systems Using Reliability Indicators** 575
Th.Olbrich, A.M.D.Richardson, D.A.Bradley
Lancaster University, Lancaster, U.K.

MP9 **Transport and Noise Quality Indicators for PN Junction Diodes** .. 581
P.Hruska, B.Koktavy, P.Vasina, H.Navarova, V.Jurankova, J.Sikula
Technical University of Brno, Brno, Czechoslovakia

MP10 **Application of the Reliability Defect Density Concept for IC Failure Rate Assessment** 585
P. De Pauw, G.Gallopyn, L.Vandenbroucke
Alcatel Mietec, Oudenaarde, Belgium

TUTORIAL 1

RELIABILITY OF ELECTRONIC EQUIPMENT

ASSESSED BY FIELD DATA COLLECTION AND ANALYSIS

A tutorial

by Prof. Jørgen Møltoft

The Danish Engineering Academy

Department of Electronic and Electrical Engineering

Akademivej, Building 451

DK-2800 Lyngby, Denmark

Abstract

Ten years ago the reliability of electronic equipment was in most cases 'determined' by laboratory testing and/or predictions. The basis for the predictions was dominated by results from laboratory testing as well. Very often this type of reliability 'determination' did not correspond very well with the following experience of the field performance.

Today it becomes more and more obvious that field observations and proper analysis of such observations are the route to be followed if the reliability is to be determined realistically and more important is to be improved continuously.

This approach is backed up by engineering methods of analysis which is not based on the assumption of a constant failure rate and which look at the reliability from a systems point of view rather than from a component point of view. Furthermore these methods take into account that after a failure normally systems are repaired and afterwards are neither "as good as new" nor "as bad as old". Finally the results of an analysis point out where to concentrate rectifying efforts most efficiently.

The tutorial goes through a number of newly developed methods for such analysis. The methods presented are at the same time very efficient and simple to understand as well as simple to apply. The presentation is focused on the practical engineering aspects which is supported by a number of industrial cases which the author has worked on succesfully over the last 10 years.

The data requirements for the methods range from crude information (the "I-p"-method) to detailed information (the M(t)-method).

As a final point a strategy for an efficient as well as an economical way to utilise the methods is outlined.

TUTORIAL 2

ADVANCED TECHNIQUES FOR INTEGRATED CIRCUIT FAILURE ANALYSIS

A tutorial by
Karel Van Doorselaer

Alcatel Bell
Francis Wellesplein 1, B-2018 Antwerp, Belgium

Abstract

A general overview will be given of the various techniques that have been developed within the semiconductor industry for integrated circuit failure analysis. This overview will present the principles, the main advantages and difficulties of the different techniques for failure localisation, imaging and sample preparation. After the overview, two techniques that have proven major usefulness in a broad field of application, will be discussed in detail.

Acoustic Microscopy has caused a real breakthrough in understanding failure mechanisms occurring in plastic-packaged ICs. The principles of this technique will be highlighted, as well as its use in package related reliability studies and package cracking (popcorn) evaluations. With this background, hardware independent failure criteria will be discussed.

Focussed Ion Beams are rapidly emerging as the powerful tool for sample preparation and silicon related failure analysis. Its field of application is broadening because of the increasing importance of its capabilities for circuit repair. The different applications will be highlighted, as well as new developments towards increased performance.

ADVANCED FAILURE ANALYSIS OF SEMICONDUCTOR DEVICES

A tutorial by
Massimo Vanzi

University of Cagliari
Istituto di Elettrotecnica, Piazza d'Armi, 09100 Cagliari, Italy

Abstract

Within the general framework of Reliability Evaluation of semiconductor devices, Failure Analysis (FA) should play the role of connecting the detected failure modes with the underlying, and often cryptic, failure mechanisms. To this purpose, many research and technology fields must contribute, ranging from basic semiconductor physics to advanced analytical techniques. Nevertheless, what is specific and crucial for FA is Strategy, which is strongly dependent on the Device Under Test and its history. As many backwards processes, this search for causes of observed effects is not univocally defined, which means that experience is a fundamental component or, what is the same, that FA of advanced devices is largely unsuccessful. In particular, Integrated Circuits beyond a fair level of complexity, compound semiconductor devices and, in this field, the specific area of optoelectronic devices are almost virgin domains for FA. the Integrated Circuit case is just partially relaxed for IC manufacturers, based on their knowledge of project, layout and technology of their own products.

The general situation, on the contrary, is testified by the very low number of actual failure analyses on those fields that are reported on specialised magazines or symposia.

The Tutorial on Advanced Failure Analysis is consequently focussed on on Strategy of some significant case histories, dealing with failures from field applications or life tests of advanced semiconductor devices of the previously indicated categories, and belonging to field applications or life tests. Well established techniques will be called into play, as well as advanced facilities, both in conventional and unusual applications, based on Strategy requirements. Continuous feedback between physical measurements, interpretive hypotheses and specimen manipulations will be presented as the interactive "navigation tool" for successful FA.

WORKSHOP 1

RELIABILITY WITH RESPECT TO SMALL TO MEDIUM ENTERPRISES

CHAIRMAN
M.HUGEN (EUROPEAN COMMISSION - IT R & D PROGRAMME)

9.00	**Introduction to the WOGOR Project** *Herman Maes (IMEC, Belgium)*
9.20	**Reliability Prediction** *Tony Bendell (Director of the Quality Unit, The Trent Nottingham University, UK)*
9.40	**Design for Reliability** *F.Jensen (Finn Jensen Consultancy, Denmark)*
10.10	**Discussion**
10.30	**Coffee**
11.00	**Reliability Testing** *Wolfgang Gerling (Siemens Semiconductors, Germany)*
11.20	**Analysis of Failures** *Ernst Demm (Siemens Semiconductors, Germany)*
11.40	**Relevance of ISO 9000 to Reliability** *Morton Brydon (QaRel Associates, UK)*
12.00	**Discussion**
12.30	**Close**

WORKSHOP 2

Power Devices Workshop
"Reliability of power semiconductors for traction applications"

Introduction	E. Wolfgang	5'
Field reliability results in thyristors and diodes employed in choppers and inverters for Italian railway locomotives	V. Nerozzi	10'
Temperature effects in power electronic devices	P. Merle	10'
Reliability tests on power semiconductors for traction application	P. Jacob	10'
General discussion	Panel	

Panel members:	Yves Danto	University of Bordeaux, France
	Fausto Fantini	University of Parma, Italy
	Robert J. Francey	GEC-Plessey, Lincoln, England
	Piet Jacob	ETH Zürich, Switzerland
	Pierre Merle	University Montpellier, France
	V. Nerozzi	FS, Components Lab., Bologna, Italy
	Pat O'Connor	British Rail Research, Derby, England
	Eckhard Wolfgang	Siemens R&D, Munich, Germany

RELIABILITY: WHICH STRATEGY?

Patrick D. T. O'Connor, Reliability Manager,
British Rail Research

1. INTRODUCTION

The electronic component industry leads the world in the achievement of quality and reliability, whilst at the same time providing products and functionality at steadily reducing prices. The reliability and cost effectiveness of modern electronic systems very largely results from the almost incredible reliability of modern electronic components. In fact, failures of electronic components in modern systems are rare events. People in the industry know how these results have been achieved, and they know that competition demands continuous improvement even on these levels.

In this paper I will discuss the effects of externally-imposed standards on the quality and reliability of electronic components.

2. ISO9000

2.1 Background

The international standard for quality systems, ISO9000, and the British Standard equivalent BS5750, have been developed to provide a framework for assessing the extent to which an organization (a company, business unit, or provider of goods and services) meets criteria related to the quality of the goods or services provided. The concept has been developed from the US Military Standard for quality, MIL-Q-9858, which was introduced in the 1950's as a means of assuring the quality of products built for the US Military services. The UK Ministry of Defence developed a similar standard (Def Stan 05-21), as did NATO (AQAP-1). The BSI, and later the ISO, used these as the basis for creating BS5750 and ISO9000. These two standards are in fact identical, and many other countries have issued national equivalents.

The original aim of supplier registration, as exemplified by standards such as MIL-Q-9858 and Def Stan 05-21, was to provide assurance that the suppliers of equipment operated visible systems, and maintained and complied with written procedures for aspects such as fault detection and correction, calibration, control of subcontractors, and segregation of defective items. They had to maintain a "Quality Manual", to describe the organization and responsibilities for quality. It is relatively easy to appreciate the motivation of large government procurement agencies such as the Ministry of Defence to impose such standards on their suppliers. However, it is widely accepted that the approach has not been very effective, despite the very high costs involved, and in the USA the Department of Defence has largely discarded it in favour of more modern methods, as described later.

2.2 Application of the Standards

The major difference between the standards and their defence-related predecessors is not in their content, but in the way that they are applied. The suppliers of defence equipment were assessed against the standards by their customers. For example, the Ministry of Defence assessed UK companies, and successful assessment was necessary in order for a company to be included on the Defence Contractors' List, and therefore entitled to be considered for MoD contracts. By contrast, the ISO9000/BS5750 approach relies on "third party" assessment: certain organizations such as BSI, Lloyds Register, and several others, are "accredited" by the National Accreditation Council for Certification Bodies (NACCB), which entitles them to assess companies and other organizations, and to issue registration certificates. The justification given for third party assessment is that it removes the need for every customer to perform his own assessment of a supplier. The supplier's registration indicates to all his customers that his quality system complies with the standard, and he is relieved of the burden of being subjected to separate assessments by all of his customers, who might furthermore have varying requirements.

The other main difference is that ISO9000 is applied to every kind of product and service, and by every kind of purchasing organization. Today, schools and colleges, consultancy practices, local government departments, and window cleaners, in addition to large companies in every industrial sector, are being forced by their customers to become registered or are deciding that registration is necessary.

The assessments are performed by trained and approved assessors. The organization desiring to be registered must apply to a certification body, and must pay the fees. Continued certification requires reassessment every 2 years, as well as random "spot check" assessments, typically twice per year.

In order to generate the number of assessors needed, several organizations are authorised to run training courses. Additionally, these organizations, and individual consultants, provide further training and consultancy, for example on management aspects, help with preparing the quality documentation necessary for registration, and preliminary assessments to help prepare for registration. Of course these services must all be paid for.

To an increasing extent, purchasing organizations such as companies, government bodies, and national and local government agencies are demanding that their suppliers must be registered. Many organizations perceive the need to obtain registration in order to comply with these requirements when stipulated by their customers. They also perceive that registration will be helpful in presenting a quality image, and in improving their quality systems.

Of course, the many organizations and

individuals, from the NACCB, the BSI, Lloyds Register, and others, to individual consultants, assessors, and trainers prosetylise and advertise the apparent benefits of the system, since they live by it. It is difficult to derive exact figures, but the overall cost to British industry is many tens of millions of pounds per annum and growing. The taxpayer and other customers also pay a large amount, in support to the BSI and the NACCB, and in the costs of registration by the suppliers of the goods and services they use.

2.3 Does ISO9000 Improve Quality?

ISO9000 does not specifically address the quality of products and services. It describes, in very general and rather vague terms, the "system" that should be in place to assure quality. In principle, there is nothing in the standard to prevent an organization from producing poor quality goods or services, so long as procedures are followed and problems are documented. Obviously an organization with an effective quality system would normally be more likely to take corrective action and improve processes and service, than would one which is disorganised. However, the fact of registration cannot be taken as assurance of quality. It is often stated that registered organizations can, and sometimes do, produce "well-documented rubbish". An alarming number of purchasing and quality managers, in industry and the public sector, seem to be unaware of this fundamental limitation of the standards.

The effort and expense that must be expended to obtain and maintain registration tend to engender the attitude that the optimal standards of quality have been achieved. The publicity that typically goes with initial registration supports this. The objectives of the organization, and particularly of the staff directly involved in registration, are directed at the maintenance of procedures and audits to ensure that people work to them. It becomes more important to work to procedures than to develop better ways of doing things.

2.4 ISO9000 and Total Quality

Total quality is the approach which was pioneered by teachers such as W.E.Deming and K. Ishikawa, and initially applied in Japan in the late 1950's, in which every person in the company becomes committed to a never-ending drive to improve quality. The drive must be led by top management, and must be vigourously supported by intensive training, the application of statistical methods, and motivation for all to contribute. The total quality concept links quality to productivity, and it has been the prime mover behind the Japanese industrial revolution. It is fundamental to the survival of any modern manufacturing business competing in world markets. Such businesses set standards for quality, internally and from their suppliers, far in excess of the requirements of ISO9000. These are aimed at the actual quality levels of the products, and at continuous improvement in these levels.

Much less emphasis is placed on the "system".

Third party assessment is at the heart of the ISO9000 approach, but the total quality philosophy demands close partnership between suppliers and purchasers. A matter as essential as quality cannot be safely left to be assessed by third parties, who are unlikely to have the appropriate specialist knowledge, and who cannot be members of the joint supplier-purchaser team. This principle applies whether the supply is of complex engineering products or of window cleaning services, or anything in between.

Defenders of ISO9000 say that the total quality approach is too severe for most organizations, and that ISO9000 can provide a "foundation" for a total quality effort. However, the foremost teachers of modern quality management all argue against this view. They point out that any organization can adopt the total quality philosophy, and that it will lead to far greater benefits than registration to the standards, and at much lower costs. It is notable that the ISO9000 approach is not used in Japan. David Hutchins, the leading teacher of quality management in the UK, states in his book, "Achieve Total Quality" (Director Books, 1992) that "Eventually, those industries that manage to survive and the governments of the countries whose industries have been taken down this blind alley will live to regret that they did not think all this through before it was too late".

2.5 The Controversy

Since its inception, ISO9000/BS5750 has generated considerable controversy. The journal of the Institute of Quality Assurance has published a large number of letters from quality professionals, arguing that the whole concept is wrong and should be abandoned. Recently, articles and letters have appeared in the national press and in journals such as The Director, mainly relating to the adverse effects of the standard on small businesses.

Small organizations are questioning the value of the exercise, as they do not see how the expensive process of preparing documentation and undergoing registration improves the quality of their products and services, and large organizations are also querying the benefits in relation to the high costs of compliance and questionable effectiveness. The evidence is, however, variable. Some organizations have generated real improvements as a result of registration, and many consultants and certification bodies provide good service in quality improvement.

As remarked above, the leading teachers of quality management all argue against the "systems" approach to quality, and the world's leading companies do not rely on it. So why is the approach so widely used? The answer is partly cultural and partly coercion.

The cultural pressure derives from the tendency to believe that people perform better when told what to do, rather than when they are given freedom, as well as the

necessary skills and motivation, to determine the best ways to perform their work. This belief stems from the concept of "scientific management", developed in the 1920's by the American researcher F.W.Taylor. "Taylorism" became the conventional management doctrine, in which managers perform the functions of thinking, planning and organising, and "workers" perform their tasks as instructed. "Taylorism" was utterly discredited by P.F. Drucker in his classic book "The Practice of Management", published in 1955. Nevertheless, the "scientific" approach to management retains a strong hold on much Western management teaching and practice.

The coercion to apply the standards comes from several directions. For example, the Treasury guidelines to public purchasing bodies states that they should "consider carefully registered suppliers in preference to non-registered ones". In practice, many agencies simply exclude non-registered suppliers, or demand that tenderers must be registered. Several large companies adopt the same policy, and of course all contractors and their subcontractors supplying the Ministry of Defence must be registered, since the MoD decided to drop its own assessments in favour of third the party approach based on ISO9000.

The Department of Trade and Industry claims that the standards are "voluntary", and that they "have been developed by industry to meet their own needs". This is simply untrue. Registration is often as voluntary as a donation to the Mafia. The standards have not been written "by industry", but by people who sit on standards-writing committees, which in practice have unfettered power to "standardise" methods which directly contradict the essential lessons of the modern quality and productivity revolution.

2.6 The Solution

The only rational solution to the situation that has been allowed to develop is to dismantle the structures that have been built around the standards, and to remove all aspects of compulsion, whether stated or implied. The standards should be used only as a guide to what should be included in a minimal quality system. The systems for accreditation and registration should be abandoned. Companies and other suppliers should be encouraged to set up and audit their own quality of goods and services they buy. Of course any organization should be free to seek external advice and auditing if they wish. However, purchasing organizations, particularly in the public sector, must not discriminate against suppliers on the grounds of who audits their quality systems, but only on their quality management and performance.

Finally, the severe limitations of the standards, in relation to modern concepts and practices of total quality, must be recognised and emphasised. Their application has no doubt led to some local improvements, but at great overall cost, and negative overall effect. Therefore,

the major agencies involved, particularly the Institute of Quality Assurance and the Department of Trade and Industry, should proclaim the deficiencies and should support application of the quality methods that are used by the world's leading companies and economies.

3. COMPONENT STANDARDS

Standards for particular component families and types orginated in the USA in the 1950's. For example, MIL-STD-38510 specifies IC's, and there is a range of standards covering every component type. Near-identical standards were issued by UK MoD and BSI (BS9000 series), and later by CENELEC (Europe) and IEC (international). Concurrently, standards for testing were developed, notably MIL-STD-883 for IC's, and this is mirrored in BS9001 and in the equivalent CEN and IEC specifications.

These standards were a good idea at the time. Generally, component quality and reliability were low, and stability of performance, packaging, etc. were seen to be beneficial for systems with long development and use cycles, particularly military systems.

However, the industry's response to the requirements of the industrial and consumer markets has resulted in the availability of ultra-reliable components at costs which are significantly lower than for components which have been produced and tested in compliance with the standards. In fact, commercial grade components, particularly IC's, have for some time been more reliable than the equivalent, expensive, components which comply with the military and other specifications.

Since component failures are so rare, and since replacement of components in modern systems is often inadvisable, the problem of component replacement in service has practically vanished. However, there are managers, particularly in the military and in some other large organisations that purchase electronic systems, who do not appreciate this.

4. RELIABILITY PREDICTION

In Reference 1 I explained why the standardised methods that have been developed for predicting the reliability of electronic systems should no longer be used. Methods such as MIL-HDBK-217 are based upon the totally wrong assumptions that:

1. All electronic components have a propensity to fail during service, and that the rates of failure are known and are constant.

2. All system failures are caused by component failures, and all component failures cause system failures. A corrolary is that failures on component tests (eg. production test) can be correlated to failures in systems.

3. "High specification" components, eg. IC's purchased to MIL-STD-38510/BS9001 etc., are more reliable than commercial grade components. (The comments in the previous section appy).

4. All components are much

more likely to fail if their operating temperatures are increased in the range up to their rated maxima.

There are other assumptions within the MIL-HDBK-217 "models" that are wrong, as described in Reference 1. Other methods also exist, such as BT's HRD4 and the French equivalent. All of these give highly misleading indications of the influence of component reliability on the reliability of systems, and they should therefore not be used, maintained or further developed. It is notable that the world's most successful electronic system developers do not use them.

5. CONCLUSIONS

I have tried to show that the standards that are often imposed with the objective of assuring quality and reliability generally have the opposite effect. They tend to add to costs, they detract from the activities that really are necessary, and they provide misleading and pessimistic indications of what is achievable using the best modern methods of quality control and design. The European electronic components industry must resist these impositions, in order to retain and strenghthen its competitive position.

Reference: 1. Reliability Prediction: Help or Hoax? Solid State Technology, August 1990

BACK SEM INSPECTION OF METALLIZATION: A TECHNIQUE FOR FAILURE ANALYSIS AND RELIABILITY

J. Barbero, E. Cordero, J. Soria

Alcatel, SESA.
Avda. Princesa Juana de Austria Km. 8,7 28021 Madrid, Spain.

1. ABSTRACT

A technique based on plasma etching of the silicon die has been developed in order to study metallization defects from the back side of the component. This technique complements the infrared laser inspection, resolving the problem of the pour resolution. A stressmigration simulation has been carried out in order to determine the possibilities of this method. Back SEM inspection allows to investigate voids and reductions of the metallization lines that are not visible in a top SEM inspection.

2. INTRODUCTION

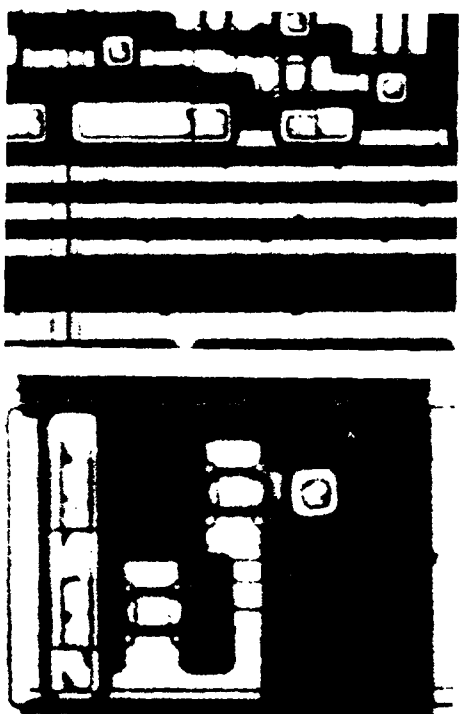
2.1 Background

Most of the voids caused by both stressmigration and electromigration are located in the bottom of the metaliization layer, avoiding the optical microscope detection and making necessary the SEM inspection to be performed at big angles and to study cross sections of several aluminun oxides steps and contact windows (in at least two perpendicular directions). To complement these studies, a technique has been developed in order to perform a SEM inspection of metallization layers from the back side of the chip.

2.1.1 Infrared laser inspection

Infrared laser inspection from the back side of the silicon and gallium arsenide devices has been widely used to study the defects of both, metallization lines (e.g. stressmigration), and contact windows. A problem of this technique is the "pour" resolution due to the size of the spot. The Laser Scanning Microscope (LSM Zeiss) available in our laboratory, has a spot size (using a 50x objetive with a N.A. of 0.50, and a laser of 1152 nm) of 2.81 µm. This implies a resolution that is not enough to perform a deep inspection of most of VLSI devices.

The method to carry out this inspection consists of sanding and polishing the die from the backside, which permits to get a flat surface and, as the silicon is transparent to the infrared laser, to be able to receive the reflective image through the die.



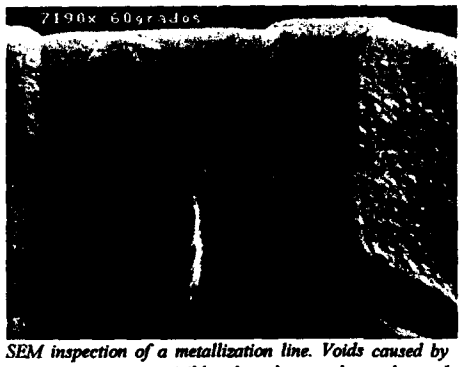
Infrared laser inspection: voids are clearly visible in the metallization lines and in the contact windows. Top SEM inspection does not allow to study these defects. Voids were caused by stressmigration.

2.1.2 Top SEM inspection

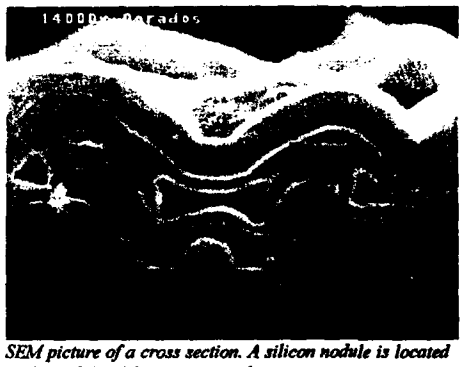
Top SEM inspection of metallization allows the detection of voids in the metallization lines, and the study of oxide steps and contact windows from the top view. If some voids are located in the back side of metallization, and this is the usual situation, it is necessary to observe the samples with an angle, in order to detect those that are near the border. However, it is not possible to study the voids that are hidden, or far from the border.

Cross sections are usually carried out in order to find out reductions at the contact windows caused by silicon nodules or any other mechanisms, such as

stressmigration or electromigration. This is an old and very useful technique, but it only provides information from the contacts that are located in the polishing plane. It is necessary to do several cross sections, at least two, in order to study the oxide steps at different directions. Therefore, a reduction in a contact window that is not located in the plane of the section, cannot be studied, and furthermore, only some contacts per section can be investigated.



SEM inspection of a metallization line. Voids caused by stressmigration are visible when the sample is observed with an angle, but it is impossible to decide how deep these voids are.



SEM picture of a cross section. A silicon nodule is located in the middle of the contact window.

3. BACK SEM INSPECTION

A SEM inspection of the back side of the metallization level can't be performed using standard techniques. Wet etching is a possibility, but it is difficult to assure that the metallization level has not also been slightly etched, and it is very difficult to keep the integrity of the layer. Taking advantage of the very low etching rate of the silicon dioxide passivation, when a CF_4 - O_2 plasma is used, it is possible to etch the silicon die and to study the back of the metallization layer by means of SEM inspection .

3.1 Method

The method is divided in three different steps:

3.1.1 Sanding and polishing

Sanding is done to decrease the thickness of the silicon to be etched. A radiography system is used to control the approaching to the top of the die. This process allows to reduce the thickness to be etched to less than 100 μm.

Radiography of a sample after sanding. The polishing plane is very near to the bonds. This is only possible for plastic packages.

When polishing is done, the thickness of silicon can be measured by focusing the infrared laser at the metallization level and at the silicon surface. This allows to do an infrared laser inspection and to reduce carefully the thickness to be etched.

3.1.2 Etching

The etching process is the common used to eliminate the glassivation, but a longer time is needed. We used a 10:1 plasma of CF_4 - O_2. If the polysilicon level is present, it means that the passivation layer has not been completely removed, and the back SEM inspection of metallization can not be performed yet. The time required to remove a silicon die is 2-5 hours.

3.1.3 Inspection

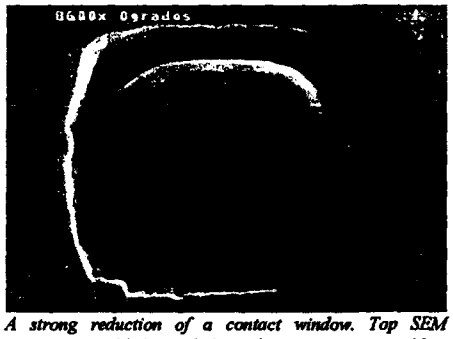
Inspection is made in with SEM. To avoid charge effects that made impossible to perform a suitable inspection, the samples must be sputtered with gold.

4. BACK OPTICAL INSPECTION

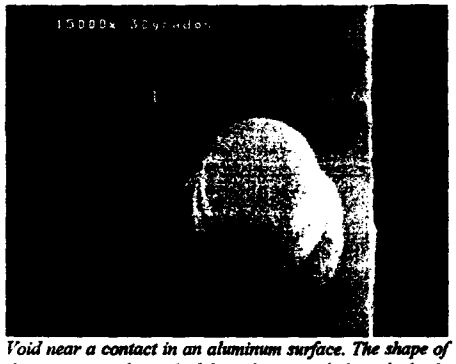
After the plasma etching process, a back inspection of the metallization and the polysilicon level can be carried out using a normal optical microscope. Images are similar to the infrared laser microscope ones, but in full color.

5.3 Examples of different devices.

In order to study the possibilities of the method, some tests have been carried out on different kind of samples. Next pictures show the results:

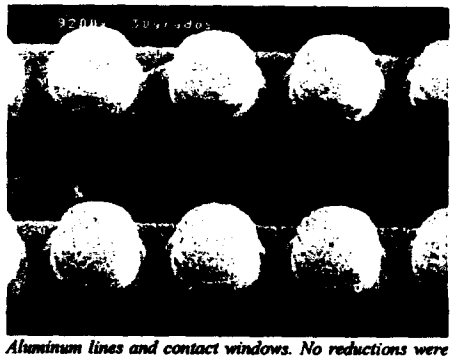


A strong reduction of a contact window. Top SEM inspection could show a hole, and a cross section could or could not show a strong reduction, depending on the polishing plane.

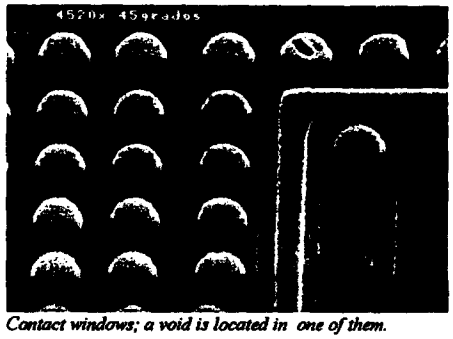


Void near a contact in an aluminum surface. The shape of the contact can be studied from the top and from the back, in order to compare them.

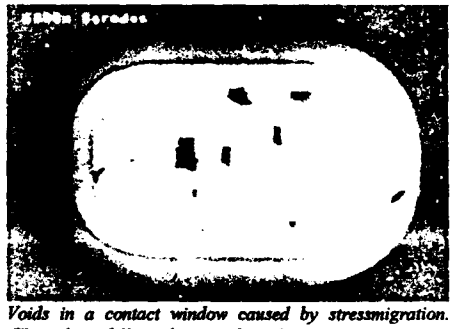
Stressmigration void in a contact window of an Al-polysilicon contact. In this case , the defect would not be detected in a top SEM inspection.



Aluminum lines and contact windows. No reductions were found in this case.



Contact windows; a void is located in one of them.



Voids in a contact window caused by stressmigration. (Their shape follows the grain boundaries).

6. CONCLUSIONS

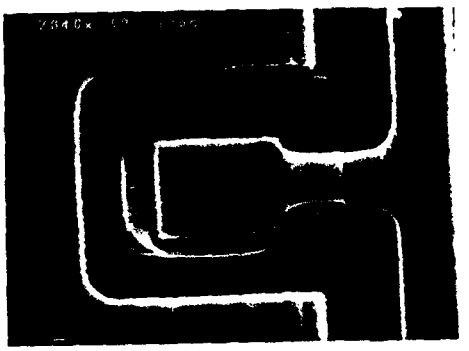
Back SEM inspection has been found to be a very helpful technique to complement the results of the top SEM inspection, providing some information that is difficult to get by any other techniques.

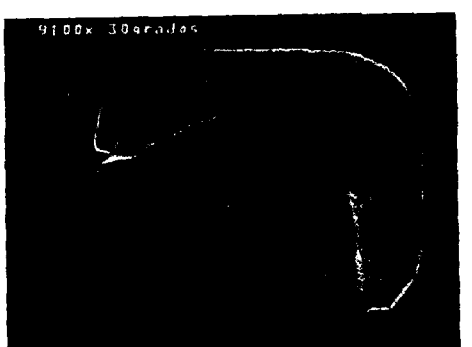
It is quiet easy to do, but care must be taken during the sanding process in order to avoid the damage of the sample, and the plasma process must be well controlled.

5. EXAMPLES

5.1 Comparaison of top and back SEM inspections.

Next pictures compare the contact windows studied from the top and from the back of the metallization lines. A cross section could show a reduction of the metallization or not, depending on the direction of the cross section plane. In the back SEM inspection, the voids, caused by stressmigration, are easily studied in the whole surface of the chip.

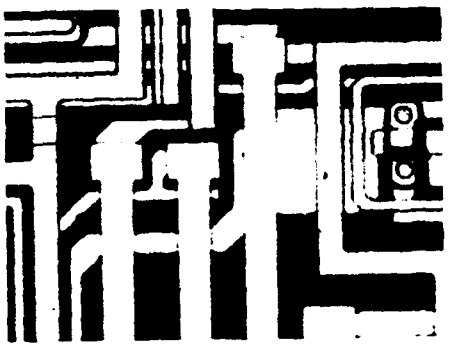


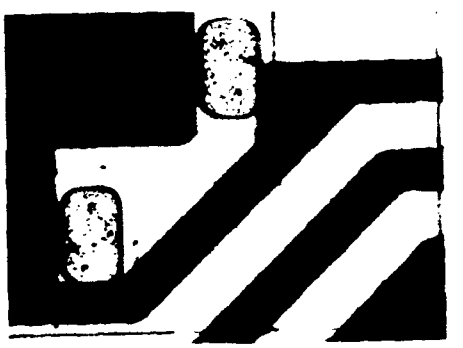


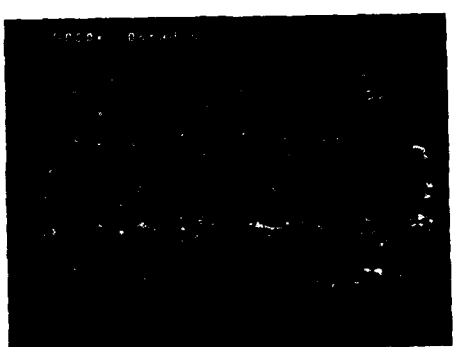
A comparaison of contacts windows of two similar samples. The first picture was got during a top SEM inspection and the other two were obtained during a back SEM inspection. Voids are visible in these cases.

5.2 Back optical and SEM inspections.

After etching the die, an optical inspection can be carried out. If the passivation has not been etched enough, some areas may present the polysilicon level. Next pictures show some Al-polysilicon contacts from the back side, before and after the removal of the polysilicon level. Back SEM inspection gives information that is very difficult to get from both the optical and the Infrared Laser inspection.







Comparaison of back optical and back SEM inspections: In the first picture the polysilicon was not etched. In the second and third pictures, the polysilicon has been etched: voids and nodules are the cause of the bad aspect of the contacts. The damages were caused by an interdiffusion process between aluminum and polysilicon.

This technique can also be helpful in failure analysis when these are related with the presence of voids in the metallization lines or in the contact windows, such as stressmigration, electromigration and interdiffusion process.

A back optical inspection can also be carried out. This allows to detect defects at the contacts and at the polysilicon levels, (before removing this layer). The images are quiet similar to those obtained using infrared laser microscopy, but with a higher resolution and in full color.

7. REFERENCES

1. Williamson P.A. and Miller P.E. "Failure Analysis of Trench Capacitor CMOS DRAMs: The Backside Approach", Proc. ISTFA, 1991, pp 225-229.

2. Miller R.L. "Backside etching techniques", Proc. ISTFA, 1985, pp 147-151.

3. Inuzuka E., Oguri S. and Himura, Y. "Emision analysis of semiconductor devices from backside of the chip", Proc. ESREF 1992, pp 269-272.

4. Diodati R. and Filippi S. "The backside etching: a tool to reveal trench defects", Proc. ESREF 1992, pp 273-276.

5. Librojo R.P., Eng P., Herschbein S. and Tomasi D. "Analysis of oxide defects on 1 megabit DRAMs", Proc. ISTFA 1990, pp. 33-39.

6. Hussey K.P. and Selesky S.L. "Laser scan microscopy applications for microelectronic failure analysis" Proc. ISTFA 1990, pp. 259-265.

7. Wills K.S., DePalois D. et al, "Single Bit Failure Analysis: Once Straight Forward is Now Challengging - A Case Study of an Advanced Microprocesor SRAM Failure", Proc. ISTFA 1992, pp 79-93.

8. Schirle M.J. and Miller P.E., "Application of Multiple Failure Analysis Techniques to Resolve Moisture-Related DRAM Failures", Proc ISTFA 1992, pp 125-134.

VOLTAGE CONTRAST STUDIES ON 0.5 µm INTEGRATED CIRCUITS BY SCANNING FORCE MICROSCOPY

CHRISTOPH BÖHM, JÖRG SPRENGEPIEL and ERICH KUBALEK

Gerhard-Mercator-Universität-GH Duisburg
Fachgebiet Werkstoffe der Elektrotechnik
Bismarckstraße 81, 47048 Duisburg, Germany

1. ABSTRACT

A scanning force microscope (SFM) test system is used for voltage contrast studies on 0.5 µm integrated circuits. Waveform measurements are performed on passivated 0.5 µm conducting lines up to 4 GHz. Additionally two dimensional measurements at 10 MHz demonstrate the potential for a device internal function- and failure analysis in the sub-µm-regime due to direct correlation between voltage contrast and quantitative topography images.

In this paper a SFM test system is used for a contactless device internal function- and failure analysis within silicon ICs. Voltage contrast images of sub-µm-conducting lines are measured and correlated to topography images. Additionally the influence of scan direction is investigated. The achieved spatial resolution is quantified from the voltage contrast images. High frequency measurements up to 4 GHz on passivated 0.5 µm conducting lines are shown and discussed.

2. INTRODUCTION

Device internal function and failure analysis of very large scale integrated circuits (VLSIC) is important for the design and simulation verification. Conventional device external test techniques for the function- and failure analysis of VLSICs must be completed by device internal contactless test techniques which allow a test access to individual devices [1]. The contactless electron beam test technique is well established and suitable for probing within VLSICs in the submicrometer regime, but the temporal resolution is limited due to the electron transit time effect [2]. Also problems in probing passivated ICs occur. A new contactless test technique with simultaneously high spatial and temporal resolution for device internal function- and failure analysis also of passivated ICs is the scanning force microscope test system [3]-[7].

3. ELECTRIC FORCE MICROSCOPY

A SFM test system for device internal voltage measurements uses a sharp conducting tip which is mounted on one end of a conducting cantilever and is scanned across the surface of a device under test (DUT). The tip is biased with a sampling voltage (the mechanical response ability of the cantilever is limited in the kHz regime. Therefore a sampling technique must be used) and interacts with the electric signal on an interconnect line via the Coloumb force. This interaction causes a deflection of the cantilever which is optically detected. By this high frequency measurements up to 40 GHz have been demonstrated [7]. At present the limitation for using higher mixing frequencies is signal coupling to the DUT due to antenna behaviour of the cantilever/tip. Special tip and cantilever geometries preventing this are under development and bandwidths above 100 GHz seem achievible.

4. TEST STRUCTURE

The test structure is build upon a 4.5 mm × 4.5 mm wide p-doped silicon chip with various conducting lines (A-H). The electrodes consist of aluminium, 500 nm thick. The test structure itself is mounted in a standard pin carrier and passivated or unpassivated avaliable. Electrical connections to the chip are made by bondwires.

For the measurements the structure F consisting of three conducting lines (spacing and line width is 500 nm in the test area) is used.

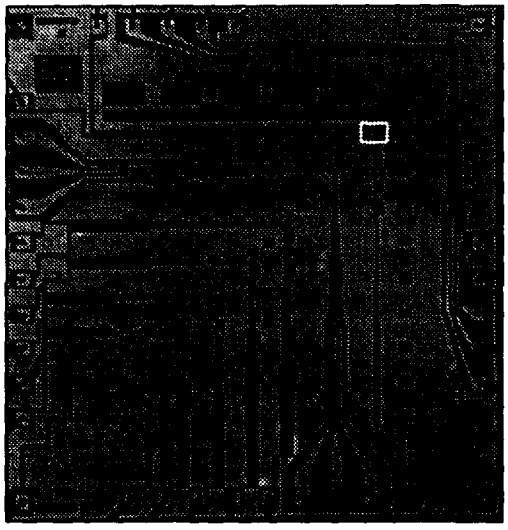


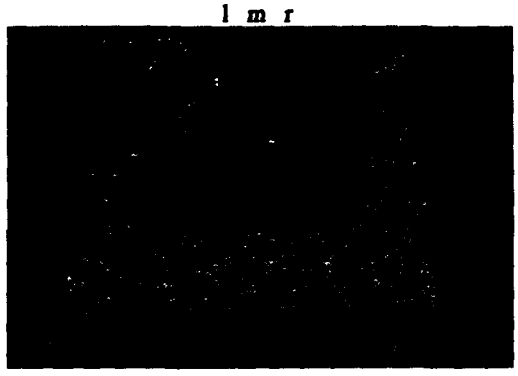
Figure 1: *Layout of the test structure with conducting lines A-H and marked test area (15 µm × 15 µm) on line structure F*

5. EXPERIMENTAL RESULTS AND DISCUSSION

For the experiments a standard SFM [8] with an adapted electrical sampling system is used.

The experimental results should achieve information of the DUT topography with nanometer resolution and also information of the voltage on the conducting lines F. Therefore first the quantitative topography of the test area was measured in contact SFM mode ona an unpassivated IC (see figure 2a). The image shows partly damaged conducting lines with rough surface. Then corresponding voltage contrast images are taken in the same area in non contact SFM mode (see figures 2b-2f). The voltage contrast images are made with applied voltages of left 0V (l), middle 3V (m), right 3V (r) at 10 MHz. An inhomogeneous voltage contrast in correlation to the damaged topography can be seen.

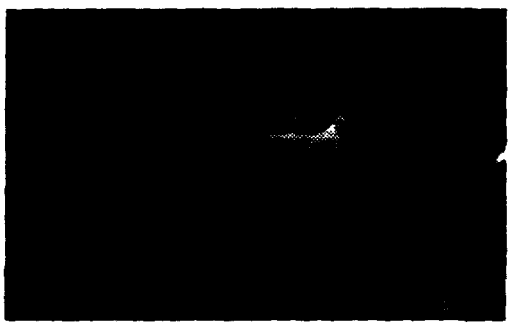
l m r



a) topography image

b) voltage contrast: tip versus DUT position: -90°

c) voltage contrast: tip versus DUT position: -45°



d) voltage contrast: tip versus DUT position: 0°

e) voltage contrast: tip versus DUT position: 45°

f) voltage contrast: tip versus DUT position: 90°

Figure 2: *a) Measured topography of the test area from figure 1 with three conducting lines (l, m, r)*
b-f) *Measured voltage contrast images for various tip versus DUT positions in 45° steps at 10 MHz for applied voltages of 0 V (l) and 3V (m, r)*

To verify that the measured voltage contrast shown is free from scan direction artefacts due to neighbouring lines which excert some additional force on the cantilever/tip several position changes are made. The DUT versus cantilever/tip position was changed in 45° steps. In figure 2b-2f the voltage contrast images of the test area from figure 2a are shown for various DUT versus tip positions. A comparison of significial points in the voltage contrast images from figure 2b-f shows no measurement artefacts due to tip/cantilever versus DUT position.

To quantify the spatial resolution of the voltage contrast we took linescans from image 2a and 2d (line 50 from 200). In figure 3 the linescans with topography information and voltage contrast are shown. To calculate the spatial resolution it is necessary to zoom figure 3. For this purpose the edge of a conducting line was choosen with large signalfall of the voltage contrast signal. In figure 4 the zoomed area from figure 3 is shown. Signal amplitudes at 10% and 90% are marked. A spatial

resolution of about 250 nm can be calculated for a working distance of 50 nm between tip and DUT.

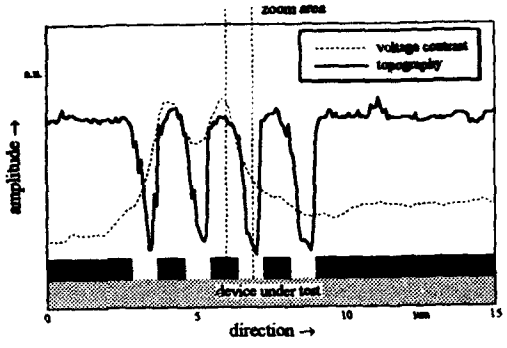


Figure 3: *Linescan from images 2a and 2d (line 50 from 200)*

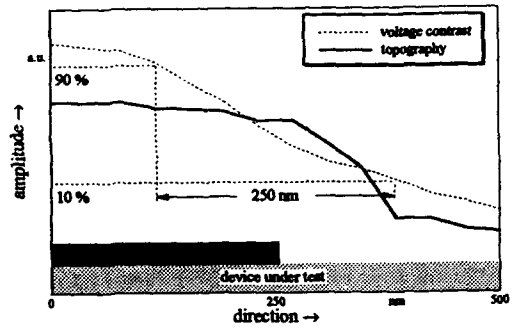
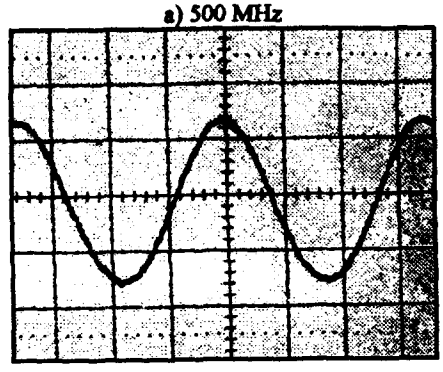


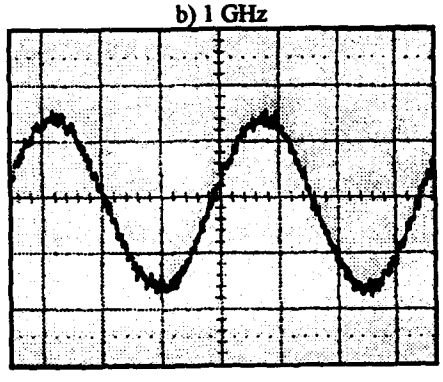
Figure 4: *Zoom area from fig. 3 with signal fall from 90 % to 10 %*

To demonstrate the high measurement bandwidth of the SFM test system simultaneously to the high spatial resolution we applied on a passivated IC an analog high frequency signal (line structure F, line (m) connected, lines (l, r) grounded). Frequencies of 500 MHz, 1 GHz, 2 GHz, and 4 GHz are used. The measured waveforms are shown in fig. 5a-d. The measured signal (see fig. 5a) looks extremly good at 500 MHz, but becomes with rising frequency noisy. This is due to the worse high frequency characteristic of the 0.5 µm conducting line causing large signal damping with raising frequencies into the GHz- regime.

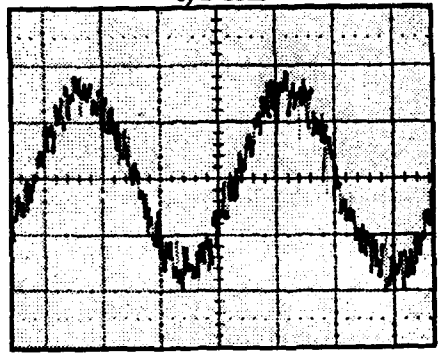
a) 500 MHz



b) 1 GHz



c) 2 GHz



d) 4 GHz

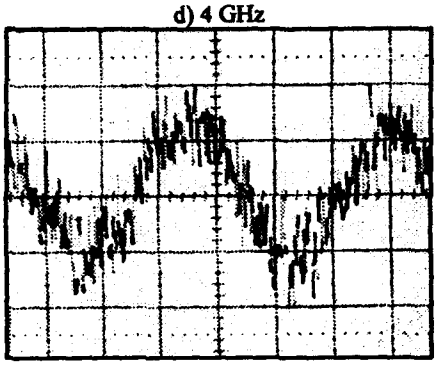


Figure 5: *Measured waveform on passivated 0,5 µm conducting line*
a) 500 MHz applied
b) 1 GHz applied
c) 2 GHz applied
d) 4 GHz applied

6. CONCLUSION

Voltage contrast studies up to 4 GHz are shown on passivated 0.5 µm conducting lines by a SFM test system. A characterization of the test structure was made by correlation between quantitative topography and voltage contrast measurement at 10 MHz. The obtained voltage contrast results are completed by measurements concerning scan direction artefacts showing no correlation between tip/cantilever versus DUT position and measurement results. The spatial resolution of the SFM test system was calculated from the experimental results to 250 nm in voltage contrast mode with still some potential for improvements.

7. REFERENCES

[1] R.B. Marcus, **"Measurement of high-speed signals in solid state devices"**, Semiconductor and Semimetals, Vol. 28, Academic Press, Inc., 1990

[2] E. Menzel, E. Kubalek, **"Fundamentals of electron beam testing of integrated circuits"**, Scanning, Vol. 5, 1983, 103-132

[3] Y. Martin, D.W. Abraham, H.K. Wickramasinghe, **"High-resolution capacitance measurement and potentiometrie by force microscopy"**, Appl. Phys. Lett., 1988, 52, 1103-1105

[4] G.E. Bridges, D.J. Thomson, **"High-frequency circuit characterization using the AFM as a reactive near field probe"**, Proceedings of the 6. Conference on Scanning Tunneling Microscopy, Interlaken, Switzerland, 12-16 August 1991, (Reprint from: Ultramicroscopy 42-44, 1992)

[5] A.S. Hou, F. Ho, D.M. Bloom, **"Picosecond Electrical Sampling Using A Scanning Force Micoscope"**, Electronics Letters, 3. Dec 1992, Vol. 28, No. 25

[6] C. Böhm, F. Saurenbach, P. Taschner, C. Roths, E. Kubalek, **"Voltage contrast in integrated circuits with 100-nm spatial resolution by scanning force microscopy"**, Journal of Physics D: Applied Physics, Vol. 26, No. 10, Oct 93

[7] C. Böhm, C. Roths, E. Kubalek, **"Contactless electrical charcterization of MMICs by device internal electrical sampling scanning force microscopy"**, MTTS - IEEE Digest (3), (1994)

[8] **Topometrix (Explorer, TMX-2000)**, Pallaswiesenstr. 180, 64293 Darmstadt, Germany

EBIC APPLICATIONS FOR COMPLEX DEVICES

A. Jurg

Philips Semiconductors
Gerstweg 2, 6534 AE Nijmegen the Netherlands

ABSTRACT

This paper presents two novel applications of the well established EBIC-technique for multi-junction devices. The first application deals with the localization of gate leakage. The second with localization of the defect die of a hybrid-IC containing 4, 8 or 12 discrete transistors.

1. INTRODUCTION

In general, papers about the EBIC-technique concern a single PN-junction. Moreover these papers are mostly very theoretical. The applications in this paper are on the contrary not only extremely down to earth, but also very powerful, sensitive and fast. This approach could be even more useful for Integrated Circuits.

The gate leakage application adds another technique to localize these kind of failures. Traditionally Liquid Crystal Microscopy, which necessitates dissipation, has been used for these kind of failures. Over the last few years Photo Emission Microscopy has been used widely as well. A big disadvantage of this technique is the fact that a defect covered by metal, which acts as a reflective layer, will not be seen in general.
Both these techniques have a lower resolution when compared with EBIC, which does not have the disadvantages mentioned above. Moreover leakage currents in the nA-range or even below are not difficult to localize.

In order to localize the defective die in a CATV-module (hybrid-IC), it was necessary in the past to remove some of the bonding wires. This was in order to localize the leakage path, which gave rise to a higher power consumption (supply current). It was also necessary to know which voltage should be present at which node. In many cases information from a development engineer was needed. Now it is possible to check the dies first without any destruction. This leaves many possibilities to analyze the device further in case of another defect.

2. PRINCIPLE OF OPERATION

The EBIC (Electron Beam Induced Current)-technique is an option that can be added to a SEM (Scanning Electron Microscope). The electron beam of the SEM is scanned over the device. With a normal detector this results in secondary and backscatter electrons being collected to form a TV like image.
The EBIC-technique however makes use of the electron-hole pairs generated by the e-beam. For each electron-hole pair in silicon 3.6 eV is needed (Ref.1). So for example an electron, accelerated with 18 kV, can result in 500 electron-hole pairs. The micrographs in this paper are made by mixing the normal- and EBIC-signals. No supply voltage was used for the devices.
The electrons and holes are separated by the built-in electrical field of a rectifying-junction. Without an electrical field the electron-hole pairs will recombine. The separation of electrons and holes gives rise to a current flow which is used to form the image. Only a connected junction will yield an EBIC signal.

3. GATE LEAKAGE

The basis of this application is that all junctions which should be present in the device are shorted either externally or by function. A pinhole through the gate oxide can give rise to a Schottky diode. This diode is formed through the gate oxide by alloying of either n-poly silicon with a p-type region or aluminum with an n-type region. The former case is shown in figure 1. This figure shows a schematic cross-section of a damaged VDMOST device.

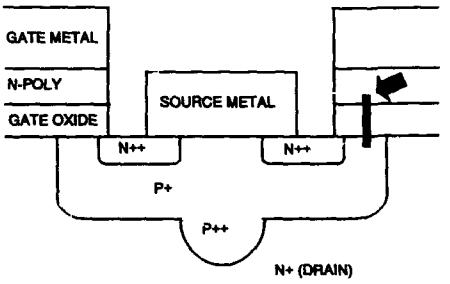


Fig. 1: Cross-section damaged gate oxide.

The EBIC signal will be measured between the gate on the one hand and the drain & source on the other.

The Schottky diode is the only junction which will show up in the EBIC signal in this configuration. Therefore gate oxide defects will appear as bright spots. Figure 2 shows three such locations of defective gate oxide of a VDMOST device.

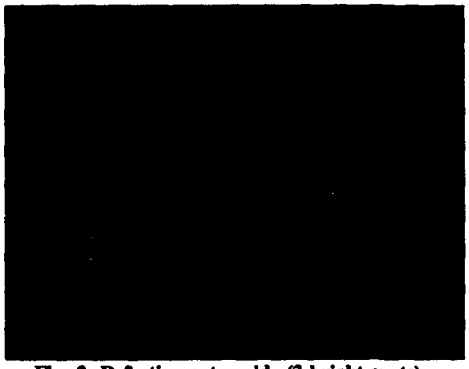


Fig. 2: Defective gate oxide (3 bright spots).

4. CATV-MODULES

In this section the defect die of a CATV-module is localized. The defect is a shorted collector-base junction of one of these dies.
CATV (Cable Antenna TeleVision) - modules contain 4, 8 or 12 similar discrete transistors soldered on analuminum oxide substrate. The module is a push-pull wideband amplifier for the 41 thru 860 MHz range. The push-pull combination is used to diminish second and third order distortion. Figure 3 shows a standard electrical circuit of such a module with four transistors. A power doubler consists of 8 dies. Each die of figure 3 is then doubled. Modules of 12 dies combine a power doubler as a final-amplifier with the circuit of figure 3 as a pre-amplifier.
Often only one of the dies is defective, due to an accidently generated external transient voltage pulse. This type of pulse usually results in a destroyed collector-base junction. These junctions are operated in reverse bias and are more sensitive to damage caused by voltage overstress unlike the forward biased emitter-base junctions.

The EBIC signal will be collected between VCC and ground. This EBIC-signal only is sufficient to show the collector-base junctions of all transistors. Each transistor contains four active base diffusion regions and four emitter ballast resistor diffusion regions. The EBIC signal of the four collector-base junctions and the four emitter resistor regions (in fact collector-emitter parasitic junction regions) of a good transistor are given in figure 4.
The location of the defective junction is clearly indicated by the absence of the collector-base part of the EBIC-signal at the defective transistor as shown in figure 5. So only the emitter resistor regions remain showing up in the EBIC-image.

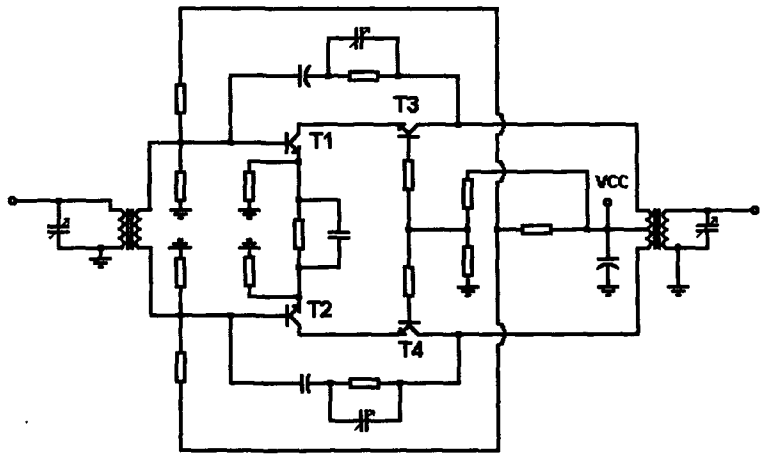


Fig. 3: Electrical circuit CATV-module.

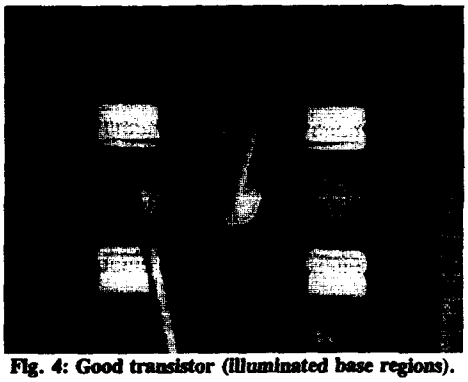


Fig. 4: Good transistor (illuminated base regions).

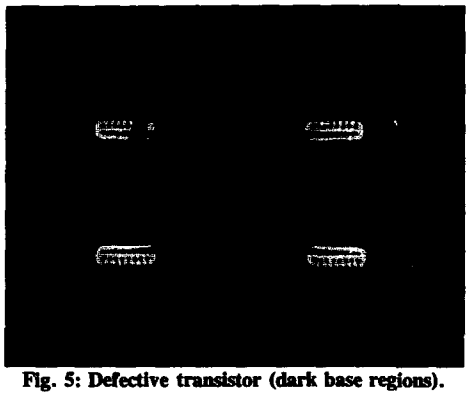


Fig. 5: Defective transistor (dark base regions).

Collection of EBIC-signal from T3 and T4: these transistors are connected to the VCC with their collectors. The bases of these transistors are connected to ground via two resistors.
Collection of EBIC-signal from T1 and T2: the bases of these transistors are connected to ground via one resistor each. The collectors are connected to VCC via T3 and T4.
It is not yet understood how the part of the EBIC-signal that is generated at T1 and T2 can pass T3 and T4 respectively.
This problem will be investigated further in future.

5. CONCLUSION

The applications presented resulted in faster analysis of device failures. This is because of an advantage of the EBIC-technique in general. As this technique combines localization and visual inspection in one single machine.

The gate leakage application is an alternative for Liquid Crystal Microscopy and Photo Emission Microscopy.
The module application made it possible to localize the defect die in a non-destructive way.

ACKNOWLEDGEMENT

The author would like to thank Roger Innes and Rik Jos for their helpful comments on the manuscript and Jan Stoffels and Berry van Druten for their support through the years.

REFERENCES

1. Breese M.B.H., King P.J.C., Grime G.W. and Watt F "Microcircuit imaging using an ion-beam-induced charge", J. Appl. Phys., Vol 72, No.6, September 1992, pp 2097-2104.

CHARACTERISATION AND APPLICATIONS OF FIB/SIMS FOR MICROELECTRONIC MATERIALS AND DEVICES

Alan J. Leslie

Failure Analysis and Reliability Department,
Institute of Microelectronics,
Block 750E Chai Chee Road #07-03/04,
Chai Chee Industrial Park,
Singapore 1646
Tel : (65)-442-0881

ABSTRACT

This paper outlines critical characterisation information such as useful yields, sputter rates and +ve/-ve ion probability of many materials using FIB/SIMS employing a Ga^+ beam. Several examples that demonstrate the application of FIB/SIMS for materials evaluation in the microelectronics industry are presented.

1. INTRODUCTION

FIB (Focused Ion Beam) technology has now been extensively applied to the microelectronics industry and in particular to the semiconductor field. Many applications of FIB including device modification, precision cross-sectioning, TEM sample preparation and even micro-machining and device fabrication have been well documented in previous literature. FIB/SIMS (Secondary Ion Mass Spectrometry utilising a Focused Ion Beam System) on the other hand is a relatively new application of commercially available FIB systems that allows materials characterisation in-situ using basic SIMS capabilities whilst FIB milling or imaging of precision cross-sections.

Several micro-SIMS or ion microprobe SIMS instruments are available or have been developed in research institutes or universities. Levi-Setti et al (Ref. 4-6) have developed an ion microprobe system at the University of Chicago - Hughes Research Laboratories using a 40keV Ga^+ beam and have published extensively on the utility of such a system. Less has been written about FIB/SIMS systems where the system has been designed primarily as an FIB with milling, deposition and enhanced etching capabilities. In this case the system has been optimised with these techniques in mind and the SIMS is an additional optional capability. Therefore the system has not been totally optimised for pure microbeam SIMS analysis and therefore a compromise has to have been made in factors such as transmission efficiency of the ion transfer optics, sputtering rates and vacuum level.

Very little data has been reported using a Ga^+ primary beam and in particular using an FIB/SIMS setup. The objective of this work is twofold. Firstly to report some important data for SIMS analysis using a 25keV Ga^+ beam that will enable more accurate and efficient data collection. Secondly it is to demonstrate some of the many applications of this capability to the microelectronics industry.

Useful Yields of several important materials have been measured and comparisons of these same materials or elements in different matrices have also been made to give some quantitative representation to allow users of such systems more insight into data interpretation. This highlights the systems capabilities in particular its weaknesses and its strengths in material sensitivity. Sputter rates have also been calculated for a range of materials allowing more accurate and widespread calibration of depth profiles.

2. CHARACTERISATION

Benninghoven et al in (Ref. 1) have compiled and presented a large amount of theory and practical data about SIMS in general. They have listed several factors that are commonly used as "Figure's of Merit" for SIMS instruments. Some of these have been used here to characterise the FIB/SIMS instrument utilised.

The following sections outline the formulae used to calculate the data presented in Section 4. These are all standard measures or indicators of ion beam system performance or specifically for SIMS in some cases. Other common indicators such as the instrumental transmission factor f', sample consumption $\dot{V}$ and the material conversion efficiency $\chi^q(A)$ could not be calculated due to the lack of samples with a suitable fractional concentration $c(A)$ of atomic species A. Further work is planned in this area.

2.1 Useful Yield :

The useful yield $\tau_u(A)$, is defined (eqn 2.1) as the number of detected secondary ions of species A per number of sputtered atoms of species A from the same sampling volume. It is therefore is a very useful indication of a system's sensitivity under specific conditions to a specific species relative to the amount of sample that has to be consumed.

$$\tau_u(A) = \frac{N_s^q(A)}{\hat{N}(A)} \qquad (2.1)$$

where $N_s^q(A)$ is the number of detected secondary ions in charge state q of element A and $\hat{N}(A)$ is the total number atoms sputtered.

2.2 Sputter Yield :

The sputter yield S_Y (eqn 2.2), is the number of sputtered atoms per incident primary ions which gives an indication of the sample consumption rate using a specific beam current on a specific material.

$$S_Y = \frac{Y_{tot} e}{I_p t} \qquad (2.2)$$

where Y_{tot} is the total sputter yield of all species irrespective of charge state, I_p is the primary ion beam current, t is the total sputter time and e is the electron charge.

2.3 Target Atom Density (atoms per μm^3) :

The target atom density n_o defined in eqn 2.3, is simply as it suggests, an estimated number of atoms per unit volume in a specific sample. This when multiplied with the volume of the sputter crater will give an estimation of the total number of atoms that have been sputtered during the analysis.

$$n_o \approx \frac{N_A \rho}{\overline{M} \times 10^{12}} \qquad (2.3)$$

where N_A is Avogadro's Number, ρ is the atomic density (g/cm^{-3}) and $\overline{M}$ is the average atomic mass.

2.4 Probability of Charged State q :

The probability of charged state q, α_m^q, of a particular species, m, is an indication of the likelihood of detecting m^{2+} versus m^+, for example.

$$\alpha_m^q = \frac{N_m^q}{\sum_q N_m^q} \qquad (2.4)$$

where N_m^q is the number of particles detected of molecular species m and in charge state q.

3. EXPERIMENTAL PROCEDURE

The system used for all the SIMS data in this paper was a commercially available FIB611 with the "SIMSmap" option from FEI Company. A Ga^+ LMIS (liquid metal ion source) is used with a two lens variable aperture focusing ion optical column. The ions used are at an energy of 25keV and can be focused to a range of spot sizes from 28nm to 280nm delivering a beam current of 6pA to 6nA. The FIB system is configured to perform all standard, and some not-so-standard applications, as well as performing FIB/SIMS.

This is accomplished using an RF quadrupole mass spectrometer mounted to the side of the FIB chamber with ion transfer optics that steer and focus the positive and negative ions generated into the mass spectrometer for analysis.

The transfer optics consist of a spherical sector energy analyser that is positioned over the sample and underneath the end of the ion column. A hole on the top sphere and one in the extraction plate, after alignment, allow the primary ions to pass through and the secondary ions to be collected. A three element

electrostatic lens and deflection plate focus the ions through the quadrupole's acceptance aperture. The quadrupole mass filters the ions to be detected by a channel electron multiplier (CEM) detector. This setup is similar in design and characteristics to that detailed in (Ref. 7).

The addition of the mass spectrometer to the FIB chamber makes use of the high spatial resolution achievable using the focused Ga^+ beam allowing high resolution SIMS maps to be obtained together with highly localised spectral analysis and depth profiling.

A beam current of 1pA and a serpentine scanned analysis area of 7.5x7.5μm was used on a set of standard samples for the characterisation work. The vacuum level of the sample chamber during the FIB/SIMS analysis ranged between 1.2 and $2x10^{-7}$ torr. All other parameters were maintained at a constant from sample to sample to establish some consistency and reproducibility in the results. A variety of system parameters are used for the applications work routinely performed on the system, some of which are described at the end of the paper.

The volume of the sputter craters was measured and calculated using an Atomic Force Microscope (AFM). This system was calibrated using a standard grid at a variety of predefined scan ranges, from 10x10μm to 100x100μm scanned using a tripod piezoelectric scanner, before the sputter crater measurements were made. The AFM used, was a commercially available system, Universal Sample AFM from Topometrix Corporation. The x-y dimension measurements made using the AFM were verified using a Hitachi S4100 field emission Scanning Electron Microscope (SEM) and a Hitachi S2150 tungsten filament SEM with a SEMICAPS image acquisition and analysis system.

4. RESULTS AND DISCUSSION

Figure 1 shows a graph of the sputter rates of various elements measured in $\mu m^3/nC$. This graph indicates the sputter rates of approximately 30 pure elements. It is clear from this diagram that the sputter rate varies quite considerably from element to element. A slight bowing or peak in the distribution can be observed around Se. This slight peak in sputter rate coincides with the atomic mass of Ga, the primary beam. It would be expected that the maximum sputtering rate is achieved when there is maximum momentum transfer between the incident ion and the stationary atoms in the sample. This occurs when both are similar in mass, ie close to the mass of Ga.

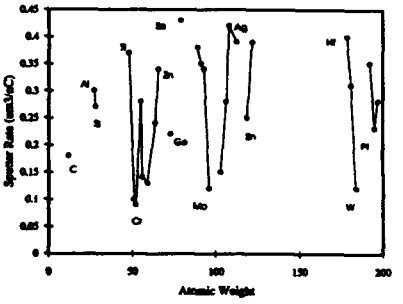


Figure 1 : Graph of Sputter Rate versus Atomic Weight of the elements measured.

Material	Formula Weight	Density (g/cm^{-3})	S_V (atm/ion)	S_r ($\mu m^3/nC$)
MgO	40.32	3.58	0.58	0.15
Al_2O_3	101.96	3.965	0.32	0.08
TiO	63.90	4.88	0.55	0.15
TiN	61.91	5.21	0.57	0.15
FeO	71.85	5.7	1.03	0.28
Fe_2O_3	159.69	5.24	0.91	0.25
Fe_3O_4	231.54	5.1	0.97	0.27
$FeSi_2$	83.95	6.1	0.90	0.24
FeS_2	119.98	5.0	1.19	0.33
Cu_2O	143.08	6.0	1.15	0.44
CuS	159.14	5.6	1.11	0.35
$CuFeS_2$	-	-	-	0.28
GaP	100.69	-	-	0.54
GaAs	144.64	5.31	2.78	0.61
SrF_2	125.63	4.24	1.67	0.44
BaF_2	175.34	4.89	1.47	0.39
LaB_6	203.78	4.61	0.28	0.07
$CeAl_2$	-	-	-	0.37

Table 1 : Sputter rates of various compound materials.

Table 1 shows the sputter rates measured for some compounds used during the characterisation. Some are very relevant for the microelectronics industry, others are there due to novel applications or for completeness. Again, it is clear in this table that the maximum sputter rates are when there is Ga in the sample.

Figure 2 shows a graph of the calculated $\tau_u(A)$ versus the atomic number of the species. All data plotted is for positive ions with q=1. The solid dots on the graph represent yields measured from pure elements and the hollow dots represent yields calculated of atomic species from a compound matrix (as listed in Table 1). Note the approximate 1/M relationship as encountered when using quadrupole spectrometers.

The sputter rate and useful yield data is critical to obtain precise data interpretation from SIMS results and to be able to plan an experiment effectively. The sputter rates allow more accurate depth profile calibration. Normally a depth profile will display with a linear x-axis representing time. However if the sample is multilayered (nearly all applications in microelectronics are like this) then the different sputter rates will cause a non-linear time scale. Using the sputter rates presented in Figure 1 and Table 1 the time scale can be converted into an approximate depth scale. As an example, analysing a material with a sputter rate of 0.25μm^3/nC using a 100pA beam in a 5x5μm area is equivalent to removing approximately 4nm/min.

$\tau_u(A)$ values are important to consider since they indicate if the instrument is even capable of analysing the problem at hand. They also give a measure of how much material has to be present or has to be consumed before any results will be obtained. This is very critical when deciding how to analyse sub-micron particles on an IC wafer for example. If the suspected particle has a very low useful yield on the graph in Figure 2 then you would have to think very seriously about using FIB/SIMS as a tool to determine its composition.

The yields displayed in Figure 2 may seem lower than others that have been published previously (Ref 3) but it is important to consider the instrument on which the data was measured and its capabilities.

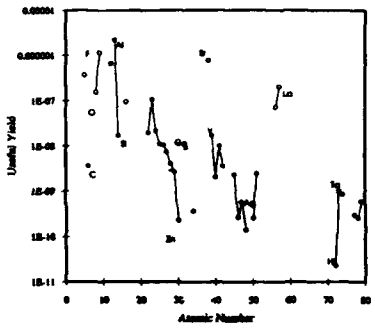


Figure 2 : Graph of Useful Yield $\tau_u(A)$ versus Atomic Number.

Here the system has a high sputter rate due to the high beam current densities and the transfer optics and quadrupole have been quoted to have a transmission of approximately 0.5% (Ref 7). These factors taken into account and the fact that the beam that is being used is Ga^+, which is not as efficient at generating secondary ions as Cs^+ or O^- beams, account for the lower yields.

Another important factor to consider before embarking on an analysis is whether the species of interest is likely to be detected as a +ve or a -ve ion. Figure 3 shows a graph of some elements and the ratios of +ve ion yield to -ve ion yield. This graph therefore indicates the likelihood of a material being either +ve of -ve. Anything above the line is likely to be seen as +ve and the further above the line the more likely it is to be +ve. Similarly for points below the line being more likely to be detected as -ve ions. This knowledge is very fundamental to the use of a SIMS instrument and should be clear before trying to detect a particular species.

The matrix effect is another complication to SIMS data interpretation and has to be considered when interpreting depth profiles as well as area maps and spectra.

Table 2 gives some examples of the matrix effect. Here it can be seen that the presence of oxygen

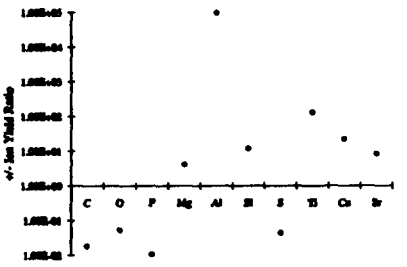


Figure 3 : Graph of +ve/-ve ion yield ratio indicating the relative likelihood of a species being detectable as either a +ve or a -ve ion.

in the metal oxides has enhanced the signals of Al^+, Cu^+, Ti^+ and Fe^+ from that of the pure metal. Likewise the enhancement has been different for different species. Al^+ and Cu^+ have been enhanced by 100x whereas Fe^+ and Ti^+ have only been enhanced 10x by oxygen. The enhancement is not just due to the oxide, it may also be due to other compound/species. For example, the copper signal has been enhanced in the sulphide as well as the oxide, however not by the same magnitude. In addition the matrix may not be enhancing but actually suppressing signal. The Fe^+ yield from iron silicide is less than that from pure Fe.

This effect is commonly seen in depth profiles for example where there may be a thin layer of oxide present on top of a metal layer between two insulators. When the ion beam reaches the metal oxide the signal of the metal ion peaks perhaps 10-100x higher than the actual bulk metal itself. Once the oxide has been eroded the peak in signal will stabilise at an equilibrium level relative to the useful yield of that metal. The same phenomenon can be seen in area maps sometimes as a bright ring around a feature or particle that would have been expected to give a solid area of signal.

Table 3 presents the data obtained for α_M^q. Any material that is not represented in Table 3 was found to have an α_M^q equal to 100% for the singly charged positive ion state.

All of the above measurements will obviously have some error due to the errors in the instrumentation and those inherent with the measurement technique.

Ion	Matrix	$\tau_M(A)$
Al^+	Al	1.18×10^{-6}
	Al_2O_3	2.95×10^{-4}
Ti^+	Ti	1.93×10^{-8}
	TiO	3.06×10^{-7}
	TiN	4.44×10^{-8}
Cu^+	Cu	2.61×10^{-9}
	Cu_2O	3.63×10^{-7}
	Cu_2S	1.79×10^{-7}
Fe^+	Fe	1.04×10^{-8}
	FeO	3.17×10^{-7}
	Fe_2O_3	2.78×10^{-7}
	Fe_3O_4	2.03×10^{-7}
	$FeSi_2$	4.72×10^{-9}
	FeS_2	6.33×10^{-7}
O^+	Al_2O_3	1.24×10^{-7}
	MgO	3.43×10^{-8}
	TiO	1.50×10^{-7}
	Cu_2O	8.44×10^{-8}
	FeO	2.21×10^{-8}

Table 2 : Examples of the Matrix Effect

Species	Ion	α_M^q
Si	Si^{2+}	0.40
	Si^+	0.55
	Si^-	0.05
Al	Al^{2+}	0.01
	Al^+	0.99
Ti	Ti^+	0.995
	Ti^-	0.005
C	C^+	0.01
	C^-	0.99

Table 3 : Probability of Charged States for those materials that were tested and did not have 100% for q=+.

One such error is introduced by using AFM to measure the sputter crater depth. Since the tip has a finite shape or profile it will not follow the steep sidewalls of the sputter crater necessarily accurately. There will therefore be an error introduced in the volume measurement. It is however felt that this error is fairly negligible in comparison to the other errors introduced

by the sputter crater roughness and non-cuboid shape.

A large source of error or more accurately, variation, in the results arises from the dependence of sputter rate upon the crystal orientation of the sample. Therefore a crystalline sample sputtered from different faces or a polycrystalline sample sputtered in random locations will give different results. Although some effort has been made to reduce the variation due to this effect, it has by no means been eliminated. Therefore there may be a range of values of which the above results fall within.

5. APPLICATIONS

There are in fact many applications of FIB/SIMS use within the microelectronics industry. Only a few very brief accounts can be given here as a sampler of what is actually possible. The two main areas that the FIB/SIMS technique has over other SIMS techniques or common laboratory capabilities (such as SEM/EDX) is its spatial resolution for high resolution mapping and highly localised depth profiling. These are of particular importance in the wafer fabrication / IC assembly and testing areas.

The system at the Institute of Microelectronics has been targeted at three main areas within the microelectronics industry : disk drives, wafer fabrication and IC packaging. The examples given in Figures 4-9 can be divided among these areas.

The first example, shown in Figure 4, is of a high resolution map of the air bearing surface (ABS) and the pole tip region on a disk drive read/write head. The top-left image is a secondary electron image showing the topographical information. The manufacturer had found particles on the head after initial testing and wanted to establish their origin. The subsequent images are of Cr^+ (bottom-left), F^- (top-right) and Cl^- (bottom-right) maps. These clearly show that one of the particles is mainly Cr (from magnetic layer on media) which indicates that the head has crashed into the media while under test. The fluorine and chlorine are found to be fairly wide spread over the ABS and pole tips. These are thought to have been due

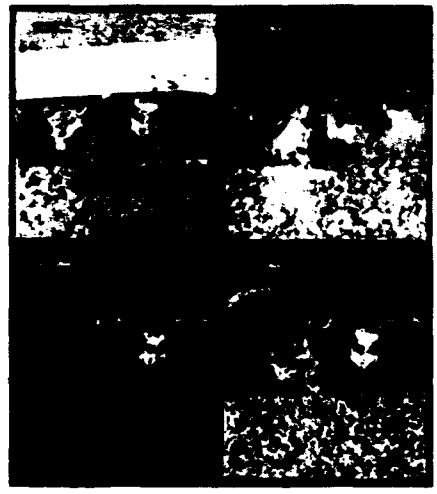


Figure 4 : SIMS Map of Head after disk crash.

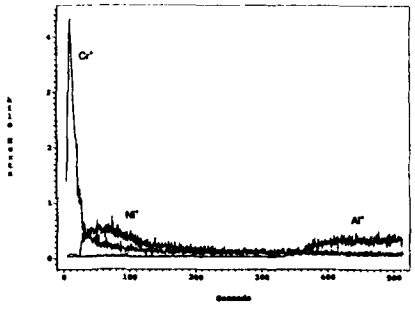


Figure 5 : Depth Profile of Magnetic Media from Hard Disk Drive.

to an insufficient clean using CFC's or similar solvents causing stiction problems in the drive.

The second example from the disk drive industry is a depth profile of a hard disk media itself and is shown in Figure 5. This displays three clear regions in the media's structure. Cr^+ from the magnetic layer near the surface (possibly Co/Cr), Ni^+ from an underlying NiP layer used as a smooth coating of the Al substrate seen as the third layer.

Figures 6 and 7 show examples from the wafer fabrication industry. The first being a mass spectrum from a sub-micron particle that was causing a preferential polysilicon growth during deposition by

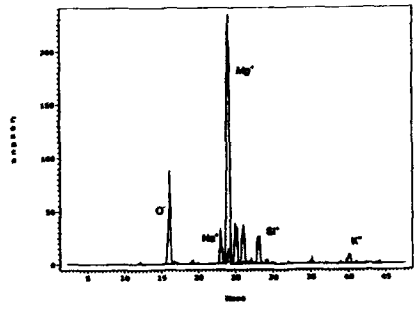


Figure 6 : Spectrum from sub-micron particle on polysilicon layer.

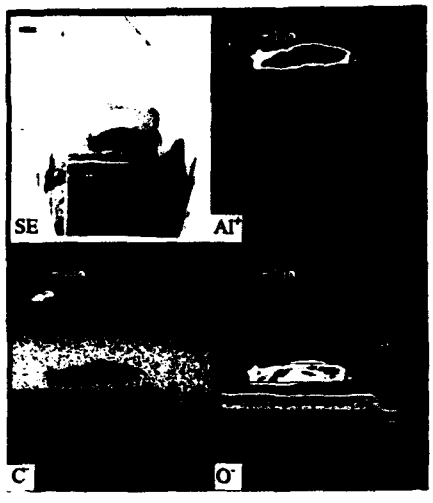


Figure 7 : SIMS map of FIB cross-section of 10μm aluminium oxide particle in polyimide layer.

acting as a seed or nucleation site. It was found that this Mg particle was one of several species (including Al and Fe) generated by the asher whilst ashing the photoresist from the previous step. Figure 7 shows a SIMS map of an FIB cross-section of a large Al_2O_3 particle embedded in the polyimide protective overcoat. Since the particle was seen to be totally embedded in the polyimide it was deemed not to be a problem by the manufacturer.

Finally Figures 8 and 9 show examples of the systems application in the IC assembly industry. Figure 8 displays two depth profiles from bond pads on two different IC's. The top profile is from a device that had

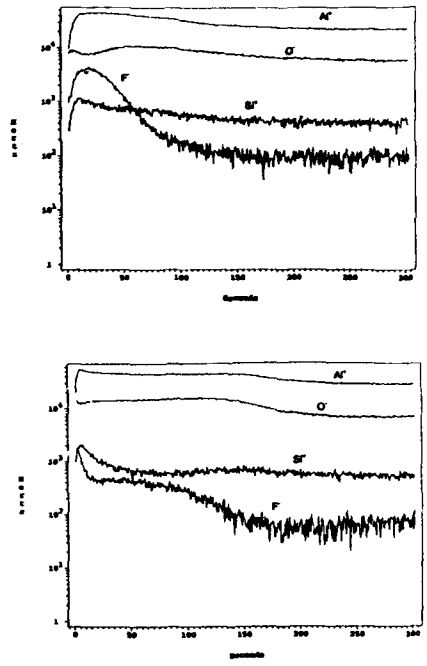


Figure 8 : Depth profiles from Al pad with bondability problem (top) and from device with no bondability problem (bottom).

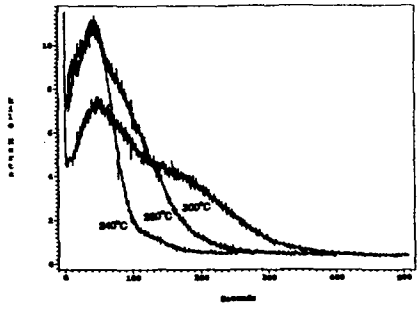


Figure 9 : Depth profiles to measure the copper oxide thickness on the backside of the die pad from copper leadframes after wire bonding at three different temperatures.

a bondability problem and the bottom one is from a good device. It can be clearly seen that the device with poor bondability has an aluminium fluoride layer on the surface of the pad. This layer could not be detected using SEM/EDX. A lot of studies have already been performed that demonstrate that the formation of aluminium fluoride after passivation window opening by HF of fluorinated plasma results in poor bondability. The final example in Figure 9 shows part of an experiment to characterise the copper oxide thickness on the backside of the die pad on copper leadframes after thermosonic gold wire-bonding at three different temperatures : 300°C, 280°C and 240°C. The increase in heater block temperature clearly shows an increase in the oxide thickness.

6. SUMMARY

It has been demonstrated very briefly that the FIB/SIMS instrument has many capabilities within microelectronics and that to allow better data interpretation and more effectively planned experimentation some fundamental data about the system is required. Some of this data has been presented. The instrument has been confirmed to have high sputter rates which have now been measured and presented for a large variety of materials. Similarly the useful yields of these materials have been calculated and presented together with some brief results on probability of charged states of a sample of species, probability of some species being detected as either a positive or negative ion and a sample example of the matrix effect and how that can affect analysis.

7. ACKNOWLEDGMENTS

I am grateful to Albert Chekanov of the Magnetics Technology Centre of the National University of Singapore for performing the AFM measurements without whom this work would not have been possible. Also, a special thanks go to my colleagues Pey Kin Leong and Sim Kian Sin for their useful discussions and support during the preparation of this paper.

8. REFERENCES

1. Benninghoven A., Rudenauer F.G. and Werner H.W. "Secondary Ion Mass Spectrometry : Basic Concepts, Instrumental Aspects Applications and Trends", Vol. 86 of Chemical Analysis Series, John Wiley and Sons.

2. Dean J.A. "Lange's Handbook of Chemistry", 13th Edition, Mcgraw Hill

3. Schumacher M, Migeon H.N, and Rasser B. "Comparative Useful Yield Measurements Under Oxygen , Gallium and Cesium Bombardment", SIMS VIII, pp49-52

4. Levi-Setti R, Chabala J, and Wang Y.L, "Micro-Secondary Ion Mass Spectroscopy : Physical and Instrumental Factors Affecting Image Resolution", Scanning Microscopy Supplement 1, 1987, pp13-22

5. Levi-Setti R, Chabala J.M., Hallegot P, Wang Y.L, "Chemical Characterisation of Electronic Microstructures with sub-100nm Lateral Resolution", Microelectronic Engineering 9 (1989), pp391-399

6. Levi-Setti R, Chabala J.M, Li J, Gavrilov K.L, Mogilevsky R and Soni K.K, "Imaging-SIMS (Secondary Ion Mass Spectroscopy) studies of advanced materials", Scanning Microscopy, Vol 7, No 4, 1993, pp1161-1172

7. Crow G.A, "Endpoint Detection and Microanalysis with Ga FIB SIMS", Proc. ISTFA'91, pp401-407

"Express" failure analysis of electronic components

C P Strudwick

ERA Technology Ltd
Cleeve Road
Leatherhead
Surrey KT22 7SA
England

Abstract

"Express" failure analysis concentrates on the important question for users of whether a failed component was badly made or whether it was misused. The methods of achieving a rapid analysis are described with a statistical summary on hundreds of devices.

1 CUSTOMER REQUIREMENTS

The failure analysis of components was once thought of as being too expensive, long-winded and inconclusive to be valuable for users of electronic components, but a procedure has been developed and optimised over several years at ERA Technology Ltd which removes these objections.

When failures occur in electronic systems, the user needs to find out several important features as quickly as possible. Once the fault has been located down to component level, further analysis is required to identify the cause of failure to ensure appropriate corrective action without wasting time on inconclusive and unnecessary investigations.

The user has a different attitude to failures from a component manufacturer. The user wants to know:

- Whether the component was well made
- If not, whether the fault is an isolated case or batch related
- Whether the component has been correctly assembled into the product
- Whether the circuit design keeps the component within its rated use
- Whether the component has been electrically overloaded
- If so, was it over-current, over-voltage, or turned on too rapidly
- Whether the environment was too severe, such as too high a temperature or humidity, or excessive thermal shock.

The cost of the components may be trivial but the results of these questions can involve considerable expense. For example redesign or replacement of parts under guarantee all over the world may be required. The answers need to be found rapidly to ensure that the systems continue to operate safely and reliably. Corrections must be made before reputations are adversely affected.

2 METHODS USED

An Express Service has been running for several years at ERA Technology to answer these needs within two days of receipt of the devices and relevant data.

The procedure consists of the following steps:

- Examine the outside of the package for damage
- Make electrical measurements to confirm or locate defect region
- Remove lid of hermetic package or open plastic case
- Inspect internal parts to 1000 times magnification
- Locate visible defects and photograph them
- Identify mechanism from evidence as far as possible
- Suggest remedial action or further analysis if required
- Supply report with photographs

3 SUMMARY OF ANALYSES

A large number of devices examined under this scheme have been semiconductors, but the range of items has included printed circuit boards, quartz crystals, electro-mechanical components and opto-electronic devices. The distribution is shown in Fig.1, where "discretes" includes diodes, transistors and passive components. The range represents the type of components that actually fail in use and not necessarily those of the latest technology.

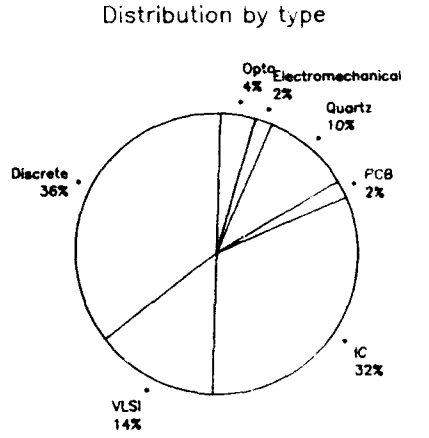


Fig.1. Distribution of components analysed under the Express Service.

An analysis of the distribution of the causes of failure is shown in Fig.2. Defects in manufacture and electrical over-load account for the largest proportions of the failures. The assembly faults were related to the insertion into printed circuit boards or to solder joints.

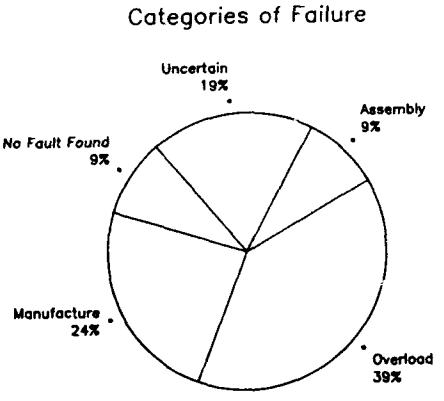


Fig.2. The proportion of components within each failure category.

The cause of failure was identified in the majority of cases within the two day period, but some components required additional further analysis.

4 EXAMPLES

From the hundreds of components analysed four examples have been selected from recent work where the failure mechanism can be briefly described by a single photograph. Several photographs might be used in reports. Normally optical photographs are used but the selection here includes some from the SEM. The examples cover manufacturing faults, assembly faults and misuse.

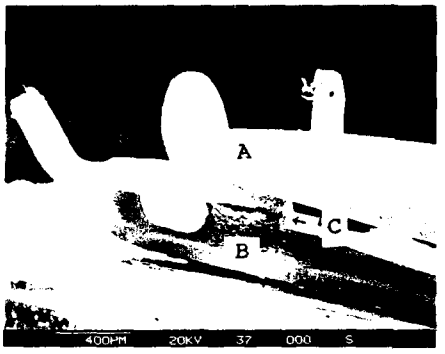


Fig.3. Component manufacturing fault in a crystal oscillator.
The SEM image shows the end of the crystal at A where it is attached to the mounting frame by silver loaded epoxy resin. Excessive adhesive had been applied, which had dripped through to the mounting frame below at B. After curing, the excess had then cracked at C, resulting in unusual damping properties and an incorrect operating frequency.

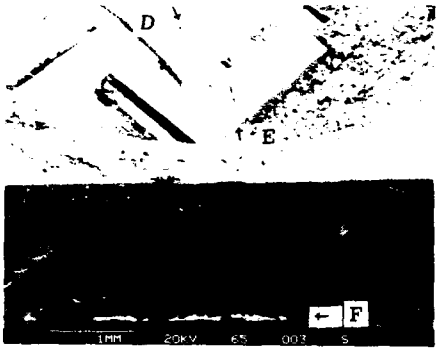


Fig.4. Fault during assembly of a thyristor into a system.
The SEM image shows the device at an angle after the plastic immediately above the die has been removed to reveal cracks in the silicon at D and E. These had been caused by the user breaking off the heatsink tab at F on the outside of the package to fit the component into a small space. The broken tab is partly obscured by a conformal coating.

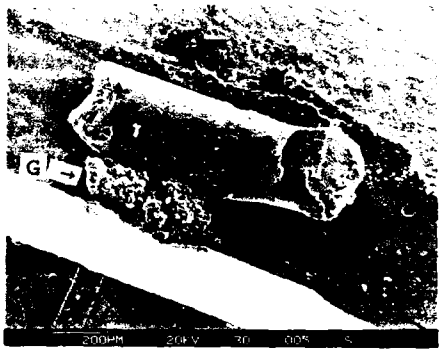


Fig.5. Fault due to misuse in a 1200 Amp "hockey puk" SCR.
The SEM image shows the wire bond to the gate contact at the side of the 25 mm diameter die. The wire was removed

while opening the package. The fault is a hole in the die at G filled with an eruption of silicon caused by the anode current rising too rapidly. As the gate is turned on, the area available for conduction spreads across the die at about 100 μm/μs. Here the rate of rise of the anode current (dI/dt) was too large, so the current was concentrated close to the gate contact, causing excessive temperature rise and failure.

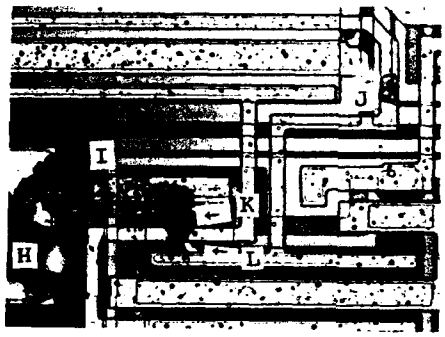


Fig.6. Fault due to overload in an integrated circuit.

The optical image shows part of an IC with severe damage in the region next to a bond pad at H. The track leading away from the pad has fused at I and another track at J. Short circuits have formed above the die at K and another beneath the oxide at L. Similar faults had occurred on other devices at the same external pin number.

5 CONCLUSIONS

The Express Service has successfully answered the main questions required by users in a short time scale on the range of components that actually fail in service.

The well established procedure, coupled with decades of experience, provides the answers required by the user without levels of detailed technical analysis of component manufacturing faults, which may not be immediately necessary.

A METHOD FOR THE CALCULATION OF THE SOFT-ERROR RATE OF SUB-μm DYNAMIC LOGIC CMOS CIRCUITS

T. Juhnke, M.-P. Bringmann, H. Klar

Technische Universität Berlin, Institut für Mikroelektronik, Sekr. J13
Jebensstraße 1, 10623 Berlin, Germany

ABSTRACT

As a prerequisite for predicting the soft-error rate (SER) of CMOS circuits with dynamic registers a method to calculate the SER is presented which takes into account charge collection by drift and diffusion. It has been found that besides collection due to drift, the noise charge collected by diffusion has to be considered to accurately predict the SER of dynamic CMOS circuits. Calculated results are compared to device simulations and SER measurements.

1. INTRODUCTION

In 1978 May and Woods (Ref. 1) identified α-particles as source of soft errors in DRAMs. α-particles originate from radioactive decay of uranium and thorium impurities in chip and packaging materials. When penetrating silicon, the α-particles generate electron-hole pairs along their track which then may be collected by pn-junctions via drift and diffusion mechanisms, thus leading to current pulses which can disturb normal operation of the circuit, resulting in a soft error.

To obtain high throughput rates with small chip area and low power dissipation dynamic pipeline registers (Fig. 4) are widely used in CMOS circuits for digital signal processing. In the pipeline registers information is stored as charge on floating nodes. Noise charge caused by an α–particle hit cannot be compensated by a current from a driving gate as in the case of static CMOS logic. Thus dynamic registers are especially sensitive to soft errors. Furthermore, due to minimisation the charge needed to upset a logic CMOS circuit has decreased, so that soft errors may become a problem in submicron logic circuits. For these reasons it is necessary to examine the soft-error rate of CMOS logic with dynamic registers.

This paper describes a procedure for calculating the SER of such CMOS circuits. Results of this calculation are compared to device simulations and accelerated SER measurements. In a similar way this procedure may be applied to other logic CMOS circuits with floating nodes, e. g. dynamic CMOS logic.

2. CALCULATION OF THE SER

In order to calculate the SER the charge needed to upset the circuit — the so called *critical charge* — has to be evaluated and the frequency of α-particle incidents causing collection of an amount of charge exceeding the critical charge has to be determined. The critical charge may easily be obtained by analog circuit simulations.

The frequency of α-particles causing charge transfers greater than the critical charge can be calculated using a modified version of the program "ALF" (Ref. 2). In this program the α-particle source is modelled by a thin layer of radiating material (Fig. 1). This is a good approximation to the uranium and thorium contaminated aluminum wires on the chip.

The program uses a monte-carlo method for predicting the SER: a source location in the source layer and a location on the silicon surface are randomly chosen, defining the position of an α-particle track in space. The charge collected due to this α-particle incident and the probability of its occurrence are calculated. After simulating a sufficient number of α-particle hits in this manner, the program will add up the probabilities of all hits which led to a charge transfer exceeding the critical charge. An estimate of the SER of the circuit can then be calculated.

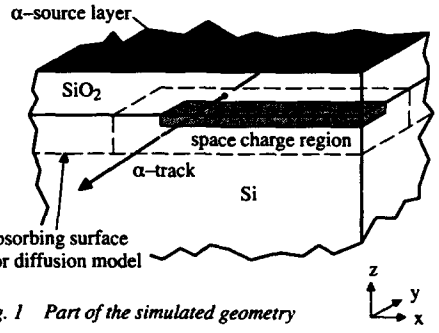


Fig. 1 Part of the simulated geometry

Due to the large number of α–particle incidents to be evaluated the collected charge has to be computed in short CPU–time. Therefore, analytical formulae are to be used rather than numerical calculations. These formulae are henceforward referred to as "n.odel", as opposed to numerical device "simulations".

In calculating the collected charge for a given α-particle track both charge collection mechanisms — drift and diffusion — have to be considered. In sub-μm circuits with rather high doping concentration the collection length for funnelling will be small, so that a relatively large amount of charge will be collected by diffusion rather than by drift (Table 2). Hence, a diffusion model has been added to "ALF".

Note that the time constant of charge collection by diffusion is much larger — in the order of 10 ns to 100 ns — than the time constant of the drift mechanism — in the order of ps to 1 ns — so that the diffusion current is much smaller than the drift current. Therefore, the charge collected by diffusion must be considered for floating nodes in dynamic circuits. In static circuits the diffusion current is easily compensated by a current from the driving gate, so that in this case the charge collected by diffusion does not result in a significant voltage change at the disturbed node.

2.1 Modelling charge collection by drift

When a pn-junction is directly hit by an α-particle, the high density of electrons and holes along the α-particle track leads to a deformation of the space charge region. The electric field extends into the substrate, thus enhancing the collection of minority carriers ("funnelling effect"). According to Hu (Ref. 4) this effect may be described using an effective collection length

$$l_C = \left(1 + \frac{\mu_n}{\mu_p}\right) l_{SCR} = \frac{w_{SCR}}{\cos\theta}\left(1 + \frac{\mu_n}{\mu_p}\right) \qquad (1)$$

for n^+p-junctions, where l_{SCR} denotes the length of that part of the α-particle track traversing the undisturbed space charge region (Fig. 2). All charge generated along the track down to l_C is collected at the junction. For p^+n-junctions μ_n and μ_p have to be reversed. The implementation of this funnelling model has been adopted from the original "ALF"–program (Ref. 2).

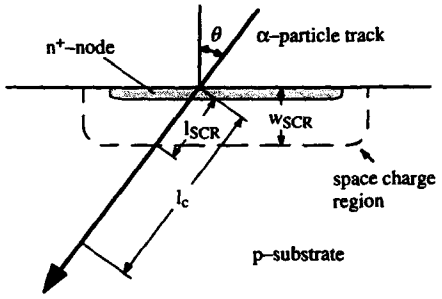


Fig. 2 Diffusion model

2.2 Modelling charge collection by diffusion

Since the time constant for the funnelling effect is much smaller than the time constant for diffusion, both effects may be modelled separately as if they occurred sequentially. After the funnelling has ceased, part of the remaining electron-hole pairs diffuse to the space charge region where they are collected. This is governed by the diffusion equation which Kirkpatrick (Ref. 3) has solved analytically for a point source below a silicon surface with infinite recombination rate (absorbing surface). The charge collected from the point-source Q_0 at $(x_0; y_0; z_0)$ by a small area dA located at $(x; y; z \equiv 0)$ is given by (Ref. 3)

$$q(x,y) = \frac{Q_0 z_0}{2\pi\left((x - x_0)^2 + (y - y_0)^2 + z_0^2\right)^{\frac{3}{2}}} dA \,. \qquad (2)$$

The total charge collected by a rectangular area with corners $(x_1; y_1)$, $(x_2; y_2)$ is obtained by integrating $q(x, y)$:

$$Q(x_1, y_1, x_2, y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} q(x, y)\, dy\, dx = \qquad (3)$$

$$= \frac{Q_0}{2\pi}(f(x_2, y_2) - f(x_2, y_1) - f(x_1, y_2) + f(x_1, y_1))$$

where

$$f(x,y) = \arctan\frac{(x - x_0)(y - y_0)}{(-z_0)\sqrt{(x - x_0)^2 + (y - y_0)^2 + z_0^2}} \,.$$

Eq. 3 is then integrated numerically over the length of the α-particle track by the modified "ALF"–program to give the total charge collected by diffusion.

Eq. 2 was derived for an infinitely thin space charge region surrounded by a silicon surface with infinite recombination velocity (absorbing surface, Ref. 3). In reality the silicon surface below the field oxide has a recombination velocity near zero, i.e. the regions surrounding the collecting node may include absorbing (other pn-junctions) as well as non–absorbing surfaces (field regions). This may be accounted for by multiplying eq. 3 by an empirically determined correction factor. The largest possible value for this factor has been found by comparison with device simulations as a maximum of 2.5 (Table 1 and Ref. 5) for a collecting node totally surrounded by field oxide. Thus the correction factor depends on the distribution of absorbing and non-absorbing regions in the vicinity of the pn-junction and varies from 1 to 2.5. To match simulation results with experiments the correction factor was set to 1 in this work.

To apply eq. 3 to a space charge region with finite thickness the absorbing surface was moved to the depth of the space charge region (Fig. 1 and Ref. 6).

Furthermore, to account for well structures where electron-hole pairs diffuse to the pn–junction as well as to the well–junction, the maximum depth for charge collection may be entered as a parameter into the program. This could better be achieved by introducing a second absorbing surface at the location of the well junction, leading to an infinite sum of terms such as eq. 3 (Ref. 6). Due to the increase in program run time such a model has not been used in this work.

3. VERIFICATION OF THE MODEL

The models for drift and diffusion developed in sections 2.1 and 2.2 respectively were verified by transient device simulations.

The applicability of the method developed to calculate the SER was then confirmed by accelerated SER measurements.

3.1 Device simulations

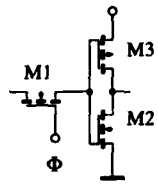
The models developed above for charge collection by diffusion and drift have been compared to the results of 2D transient device simulations in cylinder coordinates as well as to 3D transient device simulations (Table 1, 2). In order to separate the charge due to drift from that due to diffusion, an exponential time dependance of the drift current has been assumed. A piecewise defined function consisting of an exp–term and a constant was fitted to the current vs. time relationship and integrated to obtain the charge collected by drift. The fit was done by approximating a straight line to a logarithmic plot of the current (Fig. 6).

surface condition		centre hit normal to the surface		centre hit at an angle of incidence $\theta = 60°$	
		abs.	non–abs.	abs.	non–abs.
3D device simulation	drift charge Q_{drift} / fC	3.3	3.7	6.3	6.1
	diffusion charge Q_{diff} / fC	6.3	13.6	3.0	12.8
	total charge / fC	9.6	17.3	9.3	18.9
model	drift charge Q_{drift} / fC	2.9	2.9	5.8	5.8
	diffusion charge Q_{diff} / fC	6.7	6.7	5.5	5.5
	total charge / fC	9.6	—	11.3	—
Quotient $Q_{diff,simulation}$ / $Q_{diff,model}$		(0.9)	2.0	(0.6)	2.3

Table 1: 3D transient device simulation and model results of charge collected for absorbing and non–absorbing surface conditions; node size 2μm × 2μm, α–particle energy 5 MeV, applied voltage 0V, substrate doping 5 10^{16} cm^{-3}.
In the case of an absorbing surface, total charges obtained by the models and by device simulation match well, whereas their components (drift charge and diffusion charge) don't match. The reason may be that the method used to separate drift and diffusion charge in the simulation results yields imperfect results (fig.6).

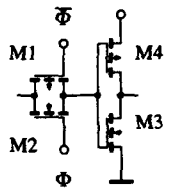
3.2 Measurements

The soft error rates of several dynamic shift registers fabricated in a 0,6 μm n-well CMOS technology have been measured at various supply voltages (Fig. 3, 4). Each shift register has 250 stages. In each stage information is stored as charge on the input capacitance of an inverter. α-particle hits in the drain/source areas of the pass transistor(s) connected to the floating node may charge/discharge this node causing destruction of stored information.



$$\left(\frac{W}{L}\right)_{M2} = \left(\frac{W}{L}\right)_{M3} = \frac{3}{0.6}$$

Fig. 3 Shift-register stage with pass-transistor



$$\left(\frac{W}{L}\right)_{M2} = \left(\frac{W}{L}\right)_{M3} = \left(\frac{W}{L}\right)_{M4} = \frac{3}{0.6}$$

Fig. 4 Shift-register stage with transmission-gate

For accelerated SER measurements an α-particle source consisting of a steel disc coated with a thin layer of ^{241}Am approx. 7 mm in diameter and having an activity of 6.9 kBq was placed approx. 1.2 mm above the chip (Fig. 5). A series of alternating 1's and 0's was then written into the shift register while the output was examined to detect and count the errors caused by α-particles.

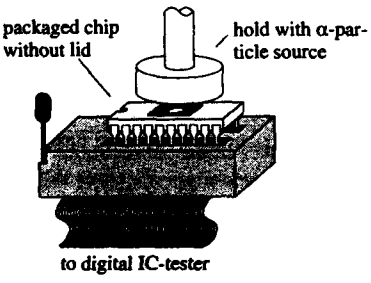


Fig. 5 Measurement setup

The results of the accelerated SER measurements were compared to results calculated with the modified version of "ALF". In the case of the test circuits involving transmission gates with pMOSFETs, the maximum depth of charge collection in the calculation was limited since the pMOSFETs were located in an n-well. The maximum collection depth was chosen to fit the calculated results to the measurements, as charge is collected both by the node hit and by the well junction and predicting the ratio between these two effects is difficult. The agreement of the calculated results with measurements is fairly good as can be seen in figures 8 and 9.

4. CONCLUSION

A method for predicting the SER of logic CMOS circuits with dynamic registers has been presented and verified. It has been shown that inclusion of charge collected by diffusion is essential to accurately predict soft-error rates of such circuits.
The method presented may also be applied to every CMOS logic where information is stored as charge on floating nodes, e. g. dynamic CMOS logic. Based on this work, prediction of soft-error rates for dynamic CMOS circuits built with future sub-µm technologies is possible.

5. ACKNOWLEDGEMENT

The authors wish to thank Dr. S. Griep and Dr. D. Schmitt–Landsiedel for many helpful discussions and for providing the test chips and the "ALF" source code.
This work is supported by SIEMENS AG, Munich.

6. REFERENCES

1. May, T.C. and Woods, M.H.: "A New Physical Mechanism for Soft Errors in Dynamic Memories", Proc. IEEE Int. Rel. Physics Symp., 1978, pp 33–40

2. Griep, S. and Künemund, Th.: "Determination of the Charge Transfer Caused by Alpha–Particles in Integrated Circuits", Proc. of the ESREF, 1992, pp 43–46

3. Kirkpatrick, Scott: "Modeling Diffusion and Collection of Charge from Ionizing Radiation in Silicon Devices", IEEE ED–26, No 11, Nov. 1979, pp 1742–1753

4. Hu, C.: "Alpha–Particle–Induced Field and Enhanced Collection of Carriers", IEEE EDL–3, No 2, Feb. 1982, pp 31–34

5. Terrill, K.; Hu, C. and Neureuther, A.: "Computer Analysis on the Collection of Alpha-generated Charge for Reflecting and Absorbing Surface Conditions Around the Collector", Solid-State Electronics 27, No 1, 1984, pp 45–52

6. Toyabe, T. et. al.: "A Soft Error Rate Model for MOS Dynamic RAM's", IEEE ED–29, No 4, Apr. 1982, pp 732–737

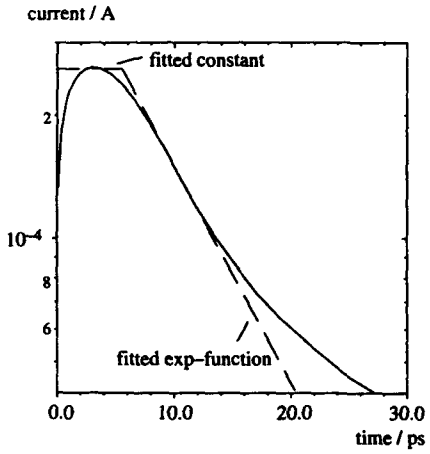


Fig. 6 Determination of charge collected by drift from 3D transient device simulation. Simulated α–particle strike: energy 5 MeV, node size 2µm × 2µm, substrate doping 5 10^{16} cm^{-3}, node voltage 0V, absorbing surface.

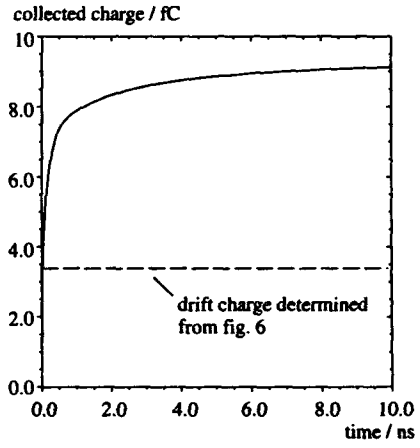


Fig. 7 Collected charge vs. time (3D transient device simulation). Simulated α–particle strike: energy 5 MeV, node size 2µm × 2µm, substrate doping 5 10^{16} cm^{-3}, node voltage 0V, absorbing surface.

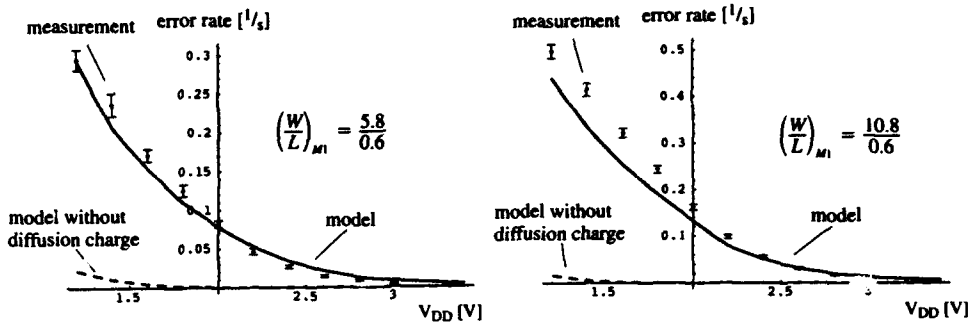


Fig. 8 Comparison measurement – model, shift registers with pass transistors (width varied, cf. fig. 3)

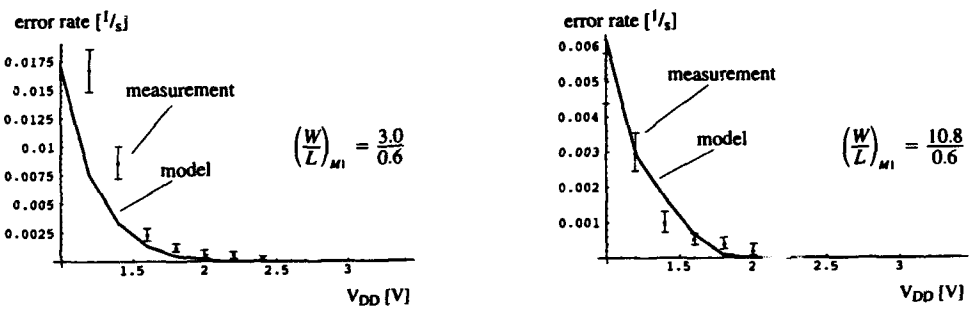


Fig. 9 Comparison measurement – model, shift registers with transmission gates (width of pMOSFET varied, cf. fig. 4). The model without diffusion charge yields a SER of 0.
The large difference of the SER of this figure to fig. 8 may be attributed to a large difference in the critical charges of the circuits in fig. 3 and fig. 4)

technology		2µm	0,6µm	0,3µm	0,12µm
node voltage / V		5.0	3.3	2.5	1.5
node diameter / µm		6.5	2.2	1.1	0.5
substrate doping / cm^{-3}		$5\ 10^{15}$	$8\ 10^{16}$	$3\ 10^{17}$	$1\ 10^{18}$
2D device simulation	drift charge / fC	20.3	4.8	2.5	1.4
	diffusion charge / fC	18.2	6.7	4.1	1.95
	total charge / fC	38.5	11.5	6.6	3.35
model	drift charge / fC	21.7	4.6	2.2	1.0
	diffusion charge / fC	24.3	8.0	4.4	1.9
	total charge / fC	46.0	12.6	6.6	2.9

Table 2: Comparison model – device simulation: Simulation was done using a 2D transient device simulation in cylinder coordinates simulating a vertical α–particle hit with 5 MeV energy. An absorbing surface boundary condition was assumed.

SIMULATING SINGLE EVENT UPSET RATE IN LARGE DIGITAL CIRCUITS

Paolo Pavan* Eric R. Minami Robert H. Tu Ping K. Ko
Chenming Hu

Department of Electrical Engineering and Computer Science,
University of California, Berkeley, CA 94720, U.S.A.

Abstract

Using this new module of Berkeley Reliability Tools (BERT) (Ref.1), user can predict the error rate due to single event upset (SEU) in large circuits. The error rate model described here uses a well established methodology, but for the first time a different choice is made on picking up the sensitive nodes, enabling a quick prediction even for very large circuits.

1 Introduction

As electronic components have grown smaller in size and power and have increased in complexity, their enhanced sensitivity to the radiation environment has become a major source of concern. Circuits can be randomly bombarded by ionizing particles of very high energy (cosmic rays and/or α-particles), the so-called single events. Pulses of ionizing radiation are known to be effective in corrupting the information ICs store. The mechanism is a single event upset and it occurs when an energetic particle passing through a microelectronic cell generates and deposits enough minority carriers so that the cell changes state. The probability of this occurring depends upon the particle environment, the possible path-length in the cell and the critical charge necessary to cause the change of state. IC designers can use radiation-hardened process technologies and radiation-hard circuit design techniques, but an important role can be played by predictive modeling to optimize the design for single-event tolerance and prediction of failure level; finally ground base testing should be used to verify radiation tolerance.

The circuit-level modeling of single event effects is an area of on-going research. The vulnerability analysis of important subsystems, such as analog subcircuits, CCD imagers, nonvolatile RAMs, and sensors, is becoming more and more important. This module is part of the BErkeley Reliability Tools - BERT, an IC reliability simulator. This new module can predict the SEU error rate for complex digital circuits. The user needs to provide the circuit and the device model parameters (as for circuit simulation), and to choose the environment in which the circuit is going to operate. Moreover, if data from on-earth characterization are available, they can be used to have a more accurate prediction of the failure rate according to the model proposed in (Ref.2).

*Paolo Pavan is now with the Dipartimento di Scienze dell'Ingegneria, Universitá di Modena, Via Campi 213/B, 41100 Modena, Italy.

2 Overview of the Simulator

Previous work did not specifically address large VLSI circuits (Ref.2). This module addresses this need. The probability of a SEU event occurring depends upon the charged particle environment, the possible path length in the cell and the critical charge necessary to cause the change of state. Error rate calculations are made according to well established methodologies developed mainly for CMOS SRAM cells. They are based on an exact path-length distribution function (Refs.3,4), a Linear Energy Transfer (LET) spectrum of the cosmic ray environment (Ref.5), and a parallelepiped-shaped sensitive volume. The sensitive volume is evaluated from the process parameters and funneling contributions, as calculated in (Ref.6), from the process parameters and from the geometries in a SPICE-deck-like input.

Before the error rate for a particular node can be calculated, the minimum charge disturbance at that node required to cause an error must be deterrmined. The amount of charge present at the node is determined by the total capacitance of the node and the voltage at the time of a single event interaction. Node capacitance is determined by interconnect capacitance, gate capacitance, junction capacitance, and stray capacitance due to overlap-regions, etc. The voltage is the noise margin of the circuit ΔV_{NM}. Critical charge, Q_C, is thus defined as $Q_C = C_N \cdot \Delta V_{NM}$. ΔV_{NM} can be either determined from other simulations, or set at $V_{dd}/2$. The module allows the user to choose whether Q_C is to be calculated according to one of these options, or assigned by user, presumably from other kinds of simulations.

If the dimensions of the parallelepiped sensitive volume are l, w, and h, the error rate can be calculated as:

$$ER = \bar{A}_p \int_{S_{min}}^{S_{max}} \Phi(s) f(s) ds \qquad (1)$$

where $\bar{A}_p = 1/2\,(lw + wh + hl)$ is the average projected area of the parallelepiped, $S_{max} = \sqrt{l^2 + w^2 + h^2}$ is the maximum path-length through the sensitive region, and $S_{min} = Q_c/(\Delta Q/\Delta S)_{max}$ is the minimum path-length for which a particle can deposit the required minimum charge Q_c. $(\Delta Q/\Delta S)_{max} = 0.28\,pC/\mu m$ is the maximum stopping power for any particle in space environment, and $\Phi(s)$ is the flux of particles which can deposit a charge greater or equal the critical charge Q_c. Finally, $f(s)$ is the distribution of pathlengths through a parallelepiped. This formulation is the basis for the calculation of an error rate for different kind of circuits. The approach is coupled with a technique modified from the prediction of degradation by timing simulator (Ref.7), which enables to choose which are the transistors, and thus the junctions, susceptible to cosmic-ray-induced errors in digital circuits. We assume that only SEU at nodes that can be pulled up or down to the rail voltage values (strong nodes) have any effect on the rest of the circuit.

As an example, we can consider the pull-down section of a NAND circuit, Figure 1. Node O, the output node, is a strong node, node A is a weak one. If N1 is OFF any dis-

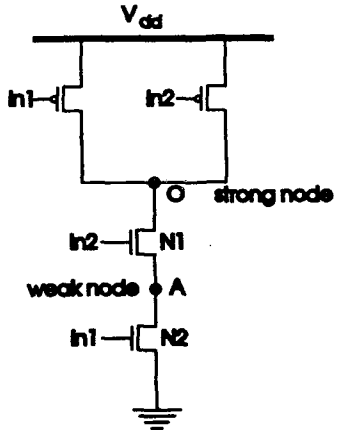


Figure 1: Nand Gate

turbance at A will not affect the output O. So, too, is the case if N2 is ON. If N1 is ON and N2 is OFF, the charge disturbance at A, ΔQ_A, is shifted to O in the worst case option. In the more accurate option, ΔQ_O is the smaller of $\Delta Q_A C_O/(C_O+C_A)$ and $(V_{dd}-V_t)C_A C_O/(C_A+C_O)$. The third and default option is that, because the internal capacitance at node A is smaller than the output capacitance at node O, the glitch produced at the output, limited to $(V_{dd}-V_t)C_A/(C_A+C_O)$ will not be large enough to propagate to any other gate in the circuit.

The BERT preprocessor determines the signal strengths at each node. Summing the SEU rate for each strong node will give the rate for the circuit.

3 Simulation Examples

The first set of simulations was carried out on CMOS SRAMs to validate the procedure. Table 1 shows the geometry parameters taken from literature (Ref.8) for devices HM6508 and CDP1821, together with process parameters and an estimate of the critical charge Q_C. For devices HM6508, the substrate doping is $1.3 \times 10^{15}\, cm^{-3}$, and the p-well doping is $1.4 \times 10^{16}\, cm^{-3}$. For a bias of $V_{dd} = 5V$, the depletion widths for the n- and p-channel transistors determine the value of H reported in Table 1. The hardened versions HS6508 and HS6508RH are assumed to have the same sensitive region geometries for the unhardened HM6508. The critical charge Q_C was estimated in $0.26\, pC$ for the unhardened version, in $0.82\, pC$ for the hardened ones. Devices CDP1821 are Silicon-on-Sapphire (SOS) devices. Ionization charges can be collected only from the epitaxial silicon region between the source and the drain. In this case, the sensitive region is limited to the dimensions of the region under the gate of a sensitive transistor. The epitaxial thickness is estimated as $0.5\, \mu m \pm 10\%$ (see values of H in Table 1). In this case, the critical charge is estimated as $1.1\, pC$. Results are in good agreement with those in (Ref.8) (Table 2).

	HM6508			CDP1821		
MOS	L	W	H	L	W	H
N1	28.4	10.4	0.72	5	15	0.5
N2	28.4	10.4	0.72	5	5	0.5
P1	22.2	19.2	2.10	5	25	0.5
P2	22.2	17.2	2.10	10	5	0.5
A1	28.4	10.4	0.72	5	64	0.5
A2	28.4	10.4	0.72	5	64	0.5

Table 1: MOS geometric parameter as from (Ref.8); L, W, H are in μm.

Error Rate	This work	Ref. [8]
HM6508	$2.2 \cdot 10^{-5}$	$2 \cdot 10^{-6} \div 7 \cdot 10^{-5}$
HS6508 HS6508RH	$3.2 \cdot 10^{-6}$	$2 \cdot 10^{-7} \div 5 \cdot 10^{-6}$
CDP1821	$2.5 \cdot 10^{-8}$	$2 \cdot 10^{-9} \div 7 \cdot 10^{-8}$

Table 2: Error Rate results for SRAMs in $Errors/bit \cdot day$, compared with those reported in (Ref.8).

Simulations were also performed on larger circuits, whose benckmark and netlist format are taken from ISCAS'85 (Ref.9) for combinatorial networks, and from ISCAS'89 (Ref.10) for digital sequential circuits. Each simulated circuit is characterized in Table 3. Simulation of digital sequential circuits was carried out either taking into account the contribution of each strong node, or considering only the nodes belonging to the D-type flip-flop. SEU error rates are reported in Table 3. The error rate of the same large circuits was simulated also for different values of V_{dd}, varying from 5 to 3 V. As shown in Figure 2, the error rate increases as V_{dd} decreases, as can be expected, since lower V_{dd} implies lower critical charge, thus a higher sensitivity to perturbations, leading to a higher error rate.

4 Conclusions

A new module has been added to BERT. It can determine the error rates for as a result of SEU in very large digital circuits. By using a timing

Circuit Name	C432	C1355	S208	S838
Circuit Function	Priority Decoder	ECAT	DFM	DFM
Total Gates	160	546	96	390
Input Lines	36	41	11	35
Output Lines	7	32	2	2
Number of D-FF	-	-	8	32
Error Rate	3.9	15.4	4.2	17.9
D-FF E-Rate	-	-	0.9	3.7

Table 3: ISCAS Benchmark Circuit Characteristics (Refs.9,10), and SEU Error Rates in $10^{-5} Errors/day$.

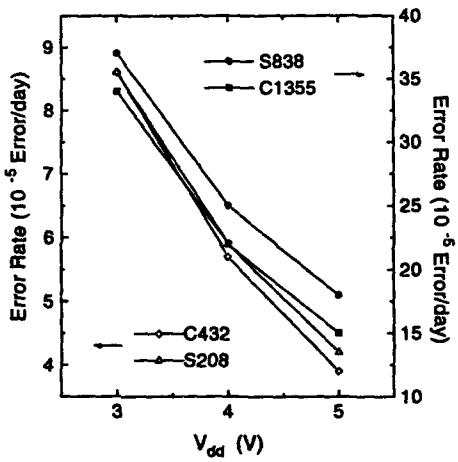


Figure 2: SEU error rate of the ISCAS benchmark circuits for different V_{dd} values

simulator and a new choice of sensitive nodes, a fast simulation of very large circuits is enabled, thus providing a useful tool for circuit designers, who can use BERT repetitively during the design to understand how specific choices can affect reliability.

Acknowledgements

This research is sponsored by SRC, ISDO/STIO administred through ONR under Contract N0019-85-K-0603, Signetics, Rockwell International, and Texas Instruments under the California MICRO program.

References

[1] R. H. Tu, E. Rosenbaum, W. Y. Chan, C. C. Li, E. Minami. K. Quader, P. K. Ko, C. Hu, *Berkeley Reliability Tools - BERT*, IEEE Trans. on CAD, Vol. 12 (10), Oct. 1993, pp. 1524-1534.

[2] E. L. Petersen, J. C. Pickel, J. H. Adams, E. C. Smith, *Rate Prediction for Single Effects - A Critique*, IEEE Trans. on Nuclear Science, NS-39 (6), 1992, pp. 1577-1599.

[3] J. N. Bradford, *A Distribution Function for Ion Track Lengths in Rectangular Volumes*, Journal of Applied Physics, vol. 50, no. 6, 1979, pp. 1570-1572.

[4] J. C. Pickel and J. T. Blandford, *Cosmic-Ray-Induced Errors in MOS Devices*, IEEE Trans. on Nuclear Science, NS-27 (6), 1980, pp. 1006-1015.

[5] W. Heinrich, Radiation Effects, 1977, 34, pp. 143-148.

[6] C. Hu, *Alpha-Particle-Induced Field and Enhanced Collection of Carriers*, IEEE Electron Dev. Letters, EDL-3 (2), 1982, pp. 31-34.

[7] E. R. Minami, K. N. Quader, P. K. Ko, C. Hu, *Prediction of Hot-Carrier Degradation in Digital CMOS VLSI by Timing Simulation*, IEDM Tech. Digest, Dec. 1992, pp. 539-542.

[8] J. C. Pickel and J. T. Blandford, *CMOS RAM Cosmic-Ray-Induced-Error-Rate Analysis*, IEEE Trans. on Nuclear Science, NS-28 (6), Dec. 1981, pp. 3962-3967.

[9] D. Bryan, *The ISCAS'85 Benchmark Circuits and Netlist Format.*

[10] F. Brglez, D. Bryan, K. Kozminski, *Combinational Profiles of Sequential Benchmark Circuits*, Proc. IEEE Int. Symposium on Circuits and Systems, pp. 1929-1934, May 1989.

ELEMENTARY TEST STRUCTURES FOR RELIABILITY PREDICTION

V. RIVIERE, A. TOUBOUL
IXL, URA 846-CNRS
Université Bordeaux I
33405 Talence Cedex, France

S. BEN AMOR, G. GREGORIS
ALCATEL ESPACE
BP 1187, 31037 Toulouse, France

J.L. STEVENSON, P.S. YEUNG
INTELSAT
3400 International Dr., N.W., Washington D.C. 20008, U.S.A

SUMMARY

A methodology assessed for wafer level reliability is described. An accelerated lifetest performed on specific test structures designed on purpose allowed us to correlate reliability with elementary process yields.

Standard High Reliability assurance procedures currently rely on component qualification and, more generally, on end-of-line testing. They can be very cost effective and add significant delays. An alternate methodology would be to monitor the quality during part manufacturing. This approach led us to design and manufacture a test chip which includes test structures to monitor elementary yields and defect densities. After electrical tests and analyses on two wafer lots, 160 devices were packaged and submitted to an accelerated lifetest. Afterwards, we studied the reliability of various features of the process and investigated possible correlations between wafer level yields and reliability.

1. TEST CHIP DESIGN

The work reported here required a mature, qualified manufacturing process and was hence performed in partnership with a space qualified Rad Hard CMOS ASIC manufacturer. The firt step was the test chip design, based on a structural decomposition of the process (Refs. 1, 2): each structure had to emphasize the impact of an elementary defect related to a process step and minimize all others. This was done within the design rules of the technology to keep a maximal commonality with "functional" circuits.

The final test chip included 34 substructures (Ref. 3) devoted to investigate the critical features of the technology.

2. WAFER LEVEL ELECTRICAL TESTS

In a second phase, two wafer batches were manufactured and electrically tested using a Keithley tester. Parameters were stored into a VAX computer and analysed using BBN's commercial RS1 statistical analysis software. Tables including test structure parameters and die x/y location on the wafer were defined. This x/y traceability was kept throughout all the phases of the project.

Most of the time, parametric data were found to be log-normally distributed: reject criteria were hence defined from cumulative distribution function plots. Yields and defect densities were obtained for the various structures. From these wafer level analyses, two different yield behaviours were found, depending on the die location on the wafer. Defects from the inner area of the wafer appeared to be randomly distributed, following a Poisson yield model, while those from peripheral regions tended to cluster, revealing more systematic defects occurrence. An additional parameter was therefore defined in order to identify the geographical origin of the dice, leading for each test structure to two values of yield and defect density (center/peripheral).

3. LIFETEST CONDITIONS

After these first analyses, 160 devices extracted from the second batch of wafers were packaged and submitted to an accelerated lifetest.

These dice were randomly selected from ten wafers preliminary extracted from the batch. Dice from center (75%) and peripheral areas (25%) of the wafer were selected and tracked in order to cover as wide a distribution of defect densities as possible.

These dice were split into different test boards in order to apply various stress conditions to each test structure. The lifetest was then performed on all chips and test structures simultaneously. The stress conditions included two temperatures (125°C and 150°C) and different bias voltages depending on the test structures. This lifetest was meant to be easily implemented in a production environment:

- thin oxides were submitted to electrical fields of 4 and 6 MV/cm
- on metal serpentines, contacts and vias chains, two voltages (8V and 16V) were applied in order to induce current densities from a few 10^6 A/cm^2 to 10^7 A/cm^2
- intermetallic oxides were stressed with an electrical field around 1 MV/cm.

All the devices were tested at given instants (100h, 200h, 400h, 840h, 1600h, 2100h).

4. LIFETEST RESULTS

For each test structure and each stress duration, we extracted the defect rates associated with the stress conditions. These results were then correlated with those obtained after wafer level tests. We have found useful to define a youth defect rate D_y (at 100h) and a wear defect rate D_w ($D_w = D_y - D_t$ at t = 1600h or 2100h) for each stress condition. In the present work, we will focus on the metal and thin oxide test structures which presented the higher defect rates.

4.1. Metal Structures

Metal serpentines showed an evolutive defect rate during the lifetest which suggested voltage and temperature related effects. Those were different according to the underlying topology of the serpentines. Three topologies were under test:

- metal serpentine over epitaxial silicon (epi) lines (serpentine 1 - see fig. 1a -)
- metal serpentine over parallele epi lines (serpentine 2 - see fig. 1b -)
- metal serpentine over an epi grid (serpentine 3 - see fig. 1c -)

Each serpentine was interdigited in two metal combs to investigate shorts but none appeared during the lifetest.

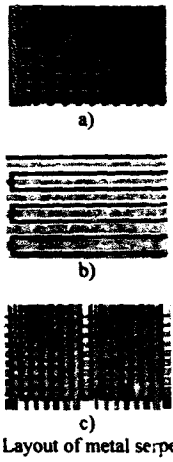


Fig. 1: Layout of metal serpentines.

Defects on serpentines 2 and 3 appeared dependent on both temperature and voltage while those on serpentine 1 were mainly voltage dependent (see fig. 2).

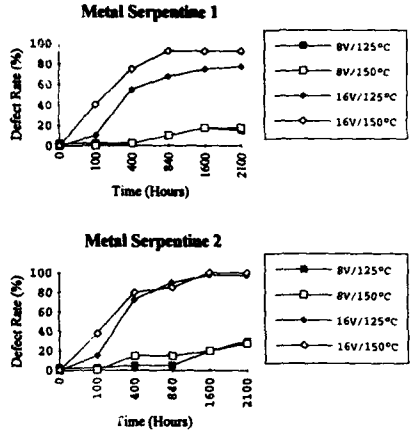


Fig. 2: Metal defect rate evolution for different test conditions.

Under the same stress conditions, the first serpentine had the higher defect rate. We hence concluded that, in the first serpentine, the effective applied current density was higher than the theoretical value because of metal thinnings: they were attributed electromigration mechanism leading to opens. SEM photographs were performed in order to support that point as showed in photo 1. Opens were more often localized above polysilicon steps as expected because of step coverage.

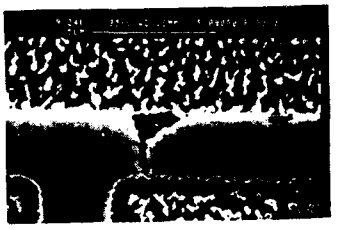


Photo 1: Sem view of an opened metal serpentine.

We then compared the youth defect rate with the wear defect rate as shown in fig. 3. This graph shows a clear correlation between both of them leading us to an important conclusion: the more youth defect are observed, the less reliable structures are.

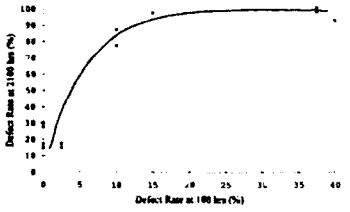


Fig. 3: Defect rate evolution between 100h and 2100h for metal structures.

4 2. Thin Oxide Structures

With regard to the thin gate oxide, three structures were studied: the first one was dedicated to investigate the area feature while the others were devoted to edge effect investigations (Ref. 3).

All of them showed a clear size dependence: the bigger test structure, the higher defect rate.

Edge-dedicated structures were found to behave in slightly different ways though they both could investigate edge effects. Those devoted to study oxide overetching showed a lower defect rate than the one dedicated to reveal oxide thinnings along epitaxial islands. The latter rather behave as the area structures which had the higher defect rate. This confirms the correct design of the test structures, showing that edge effect-induced oxide failures here originated more from thinnings above silicon island than from oxide overetching.

Our analysis consisted first in performing correlations between lifetest results and wafer level test data.

We compared Dy with Dw for each stress condition and we observed that, mainly for the biggest size, the more youth defects, the less reliable the structures.

Then, we extracted youth defect rates for each wafer and we compared them with the corresponding defect rates obtained after wafer level tests. We found a correlation between these parameters which indicated that we tend to have more youth defects when we detected more defects on wafers.

In the same way, we compared these youth wafer defect rates with the wafer breakdown voltage measured on Process Control Modules (PCM). This led us to point out a correlation showing that the lower the breakdown voltage, the higher the defect rate (see fig. 4). This correlation was even enhanced by normalizing the PCM breakdown voltage by the stress voltage: we hence observed the direct link with wafer level characteristics without including stress condition effects.

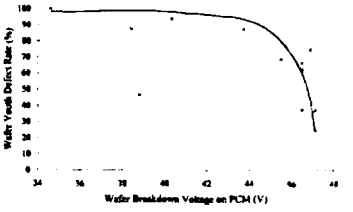


Fig. 4: Youth defect rate vs breakdown voltage on PCM.

All these correlations were then performed again by separating the dice geographically (center/peripheral). We found similar links as previously but these links were even stronger for the chips coming from the inner region while they appeared weaker for the devices from the peripheral region. This can be understood as the latter appeared more prone to systematic defects (poor Poisson dependencies at wafer level). They were therefore proportionnally less affected by the "reliability" failure mechanism that caused inner chips to fail.

Finally, we investigated the evolutive chip distributions according to the various stress conditions, using cumulative distribution function plots. The latter exhibited distributions which were divided into three parts, the first representing the good devices and the others characterizing the failures (see fig. 5). We hence found two types of defects revealing two simultaneous failure mechanisms.

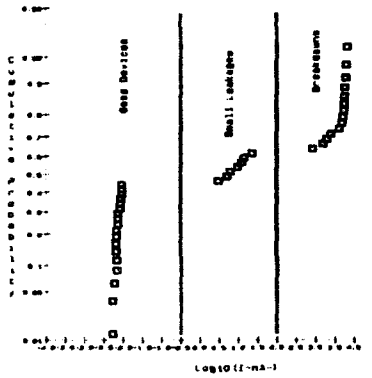


Fig. 5: Cumulative distribution function plot for oxide leakage current.

We therefore precised our failure criteria in or ler to separate these mechanisms and we studied them according to the stress conditions.

The first mechanism, corresponding to the scattered distribution, was called "small oxide leakages" as the leakage current was low and belonged to a large range of values. The second one rather corresponded to instantaneous oxide breakdowns. It appeared mainly voltage-dependent while oxide leakages were more or less temperature-dependent. In fact, oxide leakage behaviour showed linked temperature and voltage dependencies (see fig. 6):

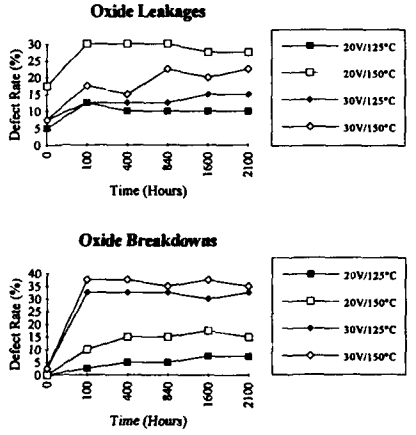


Fig. 6: Oxide failure mechanisms evolution with time and stress conditions.

- under a low stress field, the defect rate was greatly enhanced by a temperature increase
- under a higher stress field, the temperature effect vanished and the defect rate remained low. Sometimes, we could even observe a decrease of the defect rate with a temperature increase: this last phenomenom must be linked with the second failure mechanism which occured simultaneously. Under the higher stress field, breakdowns were more prone to occur than oxide leakages, which was slightly enhanced by a temperature increase.

First assumptions were made according to the origin of these failures.

Breakdowns occurrence was attributed to latent defects in the oxide layers since they appeared under rather low stress conditions.

Oxide leakage behaviour was compared to the one observed by Olivo et al. (Ref. 4) and it appeared to be rather similar. The mechanism proposed by Olivo et al. relies on the presence of weak spots in thermal oxides; high field stress destroys the oxide integrity at these spots, leading to a current enhancement. Although our oxide was thicker than the one described in (4), we can assume that the leakages we observed came from oxide imperfections which influence was enhanced by the stress. This mechanism predominated in the area structures and was weaker in the edge structures: this can support the fact that oxide leakages are more linked with oxide integrity (weak spots due to particle contamination for example) than with oxide thinnings which are predominant in edge structures. More analyses are required to accurately identify this mechanism.

5. CONCLUSION

We have proposed a wafer level reliability prediction methodology, based on the analysis of specific test structures. A test chip was designed, manufactured and electrically tested before submitting it to an accelerated lifetest. Close correlations were found on metal structures between the youth defect rate and the structure reliability. We also observed the underlying topology influence on failure occurrence.

Results on thin oxide structures allowed us to validate the design of the structures and to point out the link between reliability and elementary yields. We also observed two simultaneous failure mechanisms, one characterized by oxide leakages, the other by oxide breakdowns. More analyses are needed to accurately describe and model these mechanisms.

A last phase will be to manufacture test structures and functional circuits on the same chip. We will hence validate the results presented here and assess the suitability of such a methodology for a High Reliability procurement strategy.

References:

1. W. Lukaszek et al. "CMOS Test Chip for Process Debugging and Yield Predictions Experiments", Solid State Technology, March 1986, pp 87.

2. W. Lukaszek et al. "Test Chip Based Approach to Automated Diagnosis of CMOS Yield Problems", IEEE Trans. on Elec. Dev., Vol. 3, n°1, Feb. 1990, pp 18.

3. S. Ben Amor et al. "Test Structure for Reliability Assurance in Spacecraft Microelectronics",Proc. of ESREF93 (Bordeaux, FRANCE), 1993, pp 71.

4. P. Olivo et al. "High-Field-Induced Degradation in Ultra-thin SiO_2 Films", IEEE Trans. on Elec. Dev., Vol 35, n°12, Dec. 1988, pp 2259.

CHARACTERISATION OF Al(Cu)/TaAl THIN FILMS USED FOR THERMAL INK JET DEVICES USING ATOMIC FORCE MICROSCOPY AND TRANSMISSION ELECTRON MICROSCOPY

KL Pey, Y Chen[#], WK Chim[*] and W Qin

Institute of Microelectronics
Block 750E, Chai Chee Road #07-03/04
Chai Chee Industrial Park
Singapore 1646

[#]National Microanalysis Center for Microelectronic Materials and Devices
Institute of Materials Science, Fudan University
220 Handan Road, Shanghai 200433
People's Republic of China

[*]Centre for Integrated Circuits Failure Analysis and Reliability
Faculty of Engineering
National University of Singapore
Kent Ridge Crescent
Singapore 0511

Abstract

The microstructure of Al(4%Cu) films deposited on Ta_2Al resistor films used for thermal ink jet printheads was studied using atomic force microscopy (AFM), energy dispersive X-ray (EDX) spectroscopy and transmission electron microscopy (TEM). Characterisation of the films revealed that the Al(4%Cu) films contain Al-Ta and Al-Cu-Ta phases and precipitates possibly related to Guinier Preston zones and θ particles.

1. INTRODUCTION

Thermal ink jet printheads (Refs. 1, 2) utilise fast pulse heating of thin film heaters to locally form a bubble in the ink over the heater surface. The printhead resistor structure is normally fabricated on a silicon substrate using standard IC technique. A barrier film of silicon dioxide is first deposited. The resistor film is tantalum aluminum and is magnetron sputtering deposited. Aluminum doped with a small percentage of copper is deposited next by magnetron sputtering to form the conductor film.

Efficient performance of the thermal ink jet printer requires that the heater excitation lasts for microseconds and raises the heater surface to several hundred degrees Celsius at frequencies of a few kiloHertz. During the operation and life of the printhead, the conductor and resistor experience severe electrical, mechanical, thermal and chemical stresses which are detrimental to the performance and life of the devices. For example, electromigration may occur under the electrical stress and affects the reliability of the Al electrode considerably. Hence, it is important to characterise the metallurgical properties of the Al(4%Cu) films.

We have investigated the microstructure of Al(4%Cu) thin films deposited on Ta_2Al. A 100nm Ta_2Al resistor film was sputtered onto a 1.6μm SiO_2 barrier film. The conductor film was formed by sputtering a 500nm Al doped with 4% of Cu on the resistor layer. The entire structure was fabricated on an unpatterned p+ Si substrate. The structure and composition of the Al(Cu) films were analysed by scanning electron microscopy (SEM), atomic force microscopy, transmission electron microscopy and energy dispersive x-ray spectroscopy.

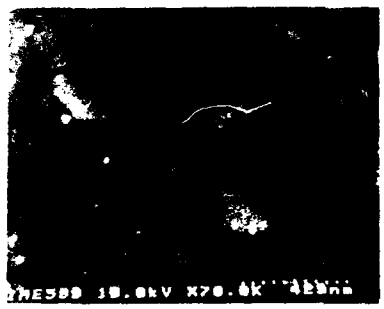
Two sets of Al(4%Cu) films, hereafter referred to as Samples #1 and #2 respectively, fabricated separately by two identical systems with slight difference in the operating conditions were studied. The Al films in Sample #1 were magnetron sputtered by a system which was run continuously overnight whereas Sample #2 was fabricated by a system during a qualifying test. In this paper, Sample #1 is used as a reference for comparative study, and only the microstructure in Sample #2 will be analysed in detail.

2. RESULTS

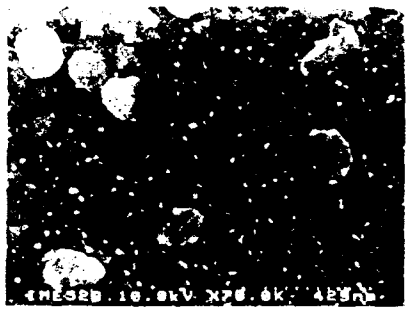
2.1 SEM/EDX and AFM

The high resolution scanning electron micrograph shown in Fig. 1 indicates that the surface of the Al conductor films consist of three distinct regions:

1) A fairly smooth and uniform region which appeared as dark contrast. This background region showed strong Al signals, with a small amount of Cu, in the SEM/EDX spectrum;
2) Small spots appeared as bright contrasts of about 20nm and
3) "Grains" of approximately 200nm - 400nm.



(a)

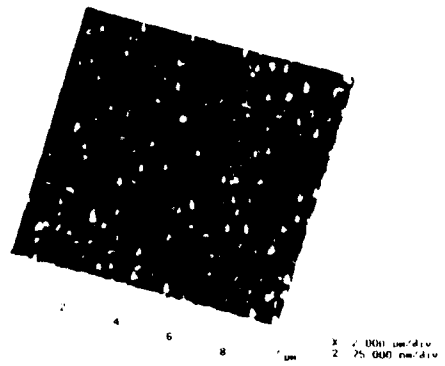


(b)

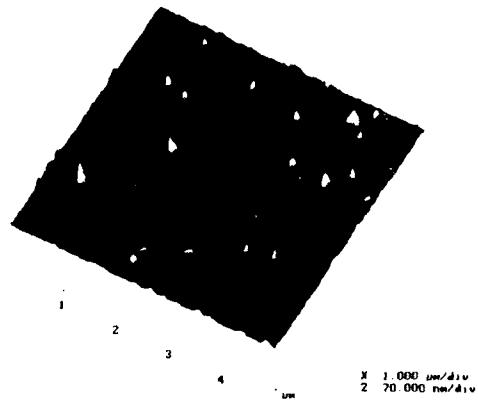
Fig. 1 *High resolution SE micrographs of a region of the Al(Cu) films. The SEM/EDX results of the background (i.e. darker contrast) showed the presence of Al and Si only. Due to the spatial limitation of the EDX in SEM, no useful elemental information could be obtained from the small bright spots and "grains". (a) Sample #1, (b) Sample #2.*

It can be clearly observed that Sample #2 contains a much higher amount of small bright spots than Sample #1.

When the surface structures seen in regions 2) and 3) in Fig. 1 were analysed by atomic force microscopy, they were shown to consist of hillocks of different sizes protruding from the Al surface. Fig. 2 shows a typical AFM image of the hillocks, the height of which varied from a few nanometres to a few tens of nanometres.



(a)



(b)

Fig. 2 *AFM images of the Al(Cu) films showing hillocks protruding from the film surface. (a) Sample #1, (b) Sample #2. Note that the scan range in (a) is 2 times larger than that of (b).*

Both the SEM and AFM images correlate very well. The "grains" seen in SEM are actually "big" hillocks with height varying from 20nm - 60nm. Whereas the small bright spots correspond to "small" hillocks of height less than 10nm.

The SEM and AFM results reveal that the Al(Cu) film surface in Sample #2 is uniformly occupied by small hillocks, with a few larger ones randomly scattered. On the other hand, only big hillocks are generally visible in Sample #1.

2.2 TEM/EDX

In the TEM analysis, only Sample #2 will be discussed in detail. Figs. 3 and 4 show a transmission electron microscope bright field image of a region of the Al(Cu) thin films in Sample #2 and the associated TEM/EDX results. The planar samples were prepared by polishing the samples from the back side mechanically and ion milling them to a thickness of approximately 0.1μm. Basically, the films contain three types of microstructure:

1) A large number of the regions (marked 1 in Fig. 3) appeared as dark contrast of about 150nm - 400nm in diameter, showing Al, Cu and Ta in the EDX results (Fig. 4(a)) collected by the TEM. These precipitates have the highest amount of Cu signals as compared to other areas. They maintained as dark contrast even at other sample tilt angles;

2) A second group of the regions (marked 2 in Fig. 3) appeared as dark circular precipitates of about 10nm - 50nm in diameter. Although the contrast was found to be a function of the sample tilt angle, they did not disappear from the image completely. Again Al, Cu and Ta were detected (Figs. 4(b) and 4(c)). However the relative composition of Cu and Ta could vary even though they appeared to have the same contrast and geometrical properties in the TEM image; and

3) The third group (marked D in Fig. 3) comprised Al grains of about 500nm in diameter, showing very strong Al signals and insignificant small amounts of Ta and Cu (Fig. 4(d)). Similar to the small dark regions seen in 2), the contrast of the Al grains varied with the sample tilt angles. But at certain tilt angle, the grain boundaries might disappear completely, forming a large Al crystal. Fig. 4(e) shows the background signals for comparison.

To ascertain that the Ta signals collected from the planar sample were not due to the effect of the Ta in the underlying Ta_2Al layer, TEM analysis on the cross section of the Al(Cu)/Ta_2Al structure was performed. Fig. 5 shows a TEM image of the cross-section of the resistor structure. Results similar to those in Fig. 3 were

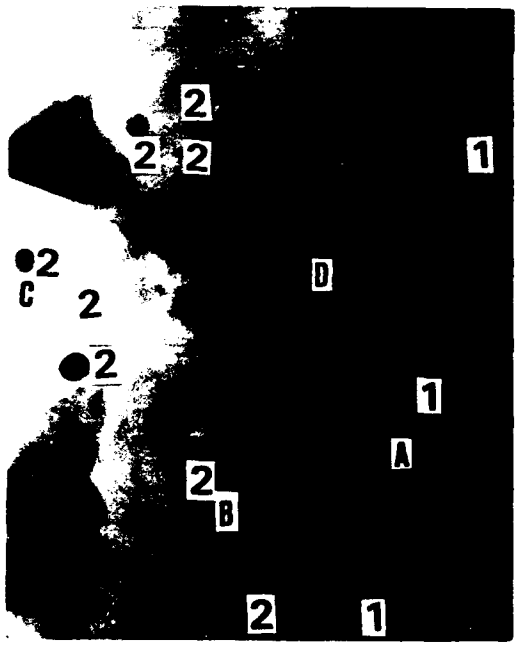


Fig. 3 *TEM bright field image of a planar region of the Al(Cu) film in Sample #2. Magnification = 100,000x. Regions indicated by markers 1 and 2 are about 150nm - 400nm and 10nm - 50nm respectively. These regions show Al, Cu and Ta signals.*

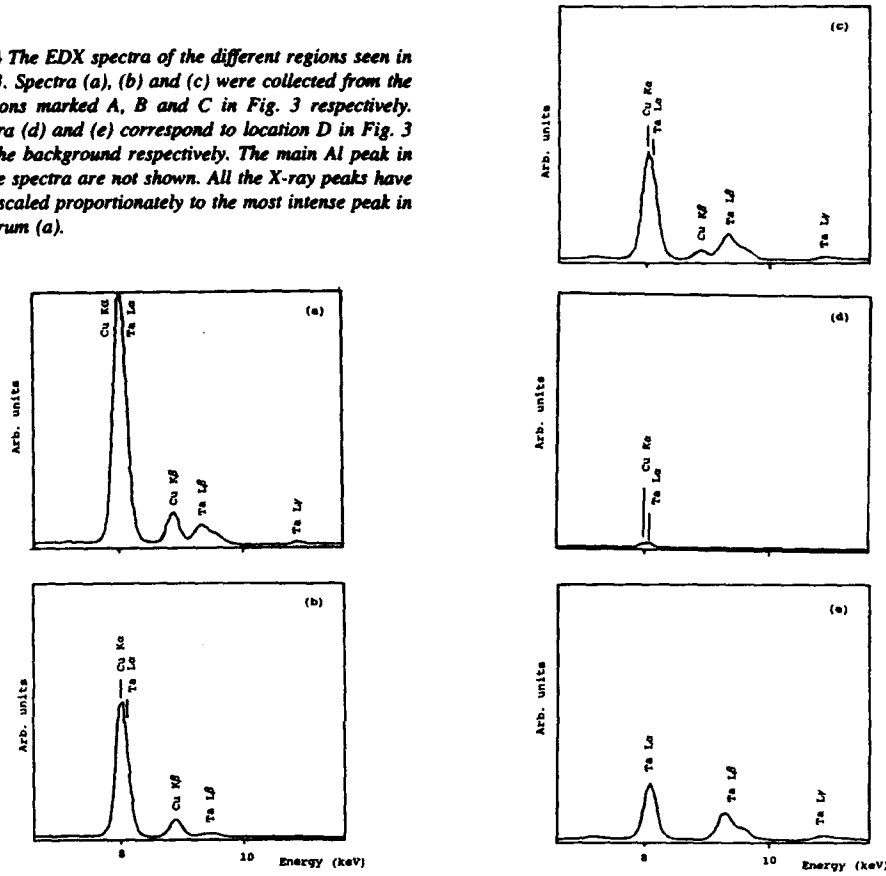


Fig. 4 *The EDX spectra of the different regions seen in Fig. 3. Spectra (a), (b) and (c) were collected from the locations marked A, B and C in Fig. 3 respectively. Spectra (d) and (e) correspond to location D in Fig. 3 and the background respectively. The main Al peak in all the spectra are not shown. All the X-ray peaks have been scaled proportionately to the most intense peak in Spectrum (a).*

obtained. Besides the Al-Cu-Ta phase found in the planar sample, an extra Al-Ta phase (Fig. 6(a)), which always appeared as dark contrast for different sample tilt angles, was observed in the Al(Cu) films. Precipitates showing Al-Cu-Ta phase, similar to those indicated by 1 and 2 in Fig. 3, were observed not only in the "bulk" of the thin film (arrows 2 and 3 in Fig. 5 and Figs. 6(b) and 6(c)) but also at places very near to the top surface (arrow 4 in Fig. 5 and Fig. 6(d)). All the microanalysis results have good reproducibility even at other regions.

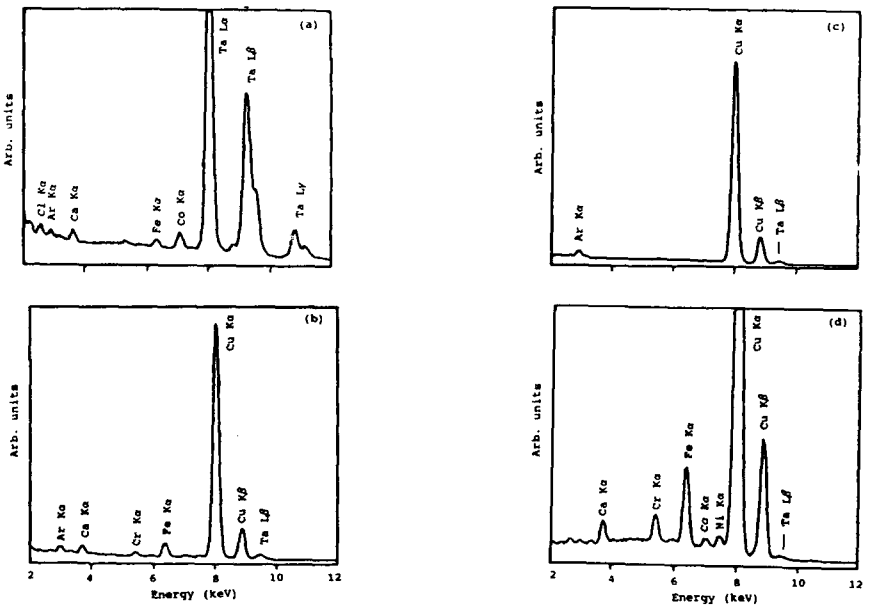
3. DISCUSSION

Based on the TEM analysis, the precipitates of size of 10nm - 50nm are suspected to be related to the formation of Guinier Preston (GP) zones before the equilibrium θ phase appears (Refs. 3,4,5). For example, GP(I) zones less than 100nm normally occur at 160°C aging temperature whereas GP(II) zones with size less than 50nm appear between 100°C - 200°C. Those particles of size of 150nm - 400nm are likely formed during the θ phase and are commonly found along the α-Al grain boundary as shown in Fig. 3. The preliminary Kikuchi line investigation and electron diffraction study of these θ precipitates from which Al, Cu and Ta were detected show that their structure is very different from the fcc structure of the α-Al.

Certainly, the presence of the precipitates

Fig. 5 *(right-hand side) TEM bright field image of a cross-section of the Al(Cu) film in Sample #2. Magnification = 42,500x. TEM/EDX was performed at those regions indicated by arrows 1 to 6. Arrow 7 marks the location of the Ta_2Al resistor film.*

Fig. 6 *(below) TEM/EDX spectra of the cross-section of Sample #2. These results have to be studied together with Fig. 5. Spectra (a) to (d) were the results of the regions marked by arrows 1 to 4 in Fig. 5 respectively. Regions 5 and 6 in Fig. 5 basically show Al signals. Also, the Al peak in all the spectra are not shown. All the X-ray peaks have been scaled proportionately to the most intense peak in Spectrum (a).*



related to the GP zones is due to the difference in the aging parameters such as the aging time or temperature. In the present study, it also appears that the hillocks revealed by SEM and AFM are likely to be Al-Cu and Al-Cu-Ta precipitates formed during different transition phases. Preliminary indications are that the large and small hillocks (precipitates) on Samples #1 and #2 surface are somewhat related to the θ phase and GP zones respectively.

From these results, it is clear that the Ta has diffused from the underlying Ta_2Al resistor film into the Al(Cu) films, giving rise to the Al-Cu-Ta and Al-Ta phases. The distribution of Ta in the Al(Cu) films was found to be random. However, the reaction between Ta and Al-Cu and Al is not clear yet. Further work is necessary to determine the mechanism of forming the Al-Cu-Ta and Al-Ta precipitates in the Al(Cu) films.

Conclusions

Al(4%Cu) films deposited on Ta_2Al resistor layer were studied. Al-Ta and Al-Cu-Ta particles were observed in the Al films. In addition, hillocks possibly related to Guinier Preston zones and θ precipitates were found on the surface of the films. Since the quality of the Al(Cu) film is dependent on the precipitating "particles", which in turn depends on the aging process, it is important to study the cause of the existence of the Al-Cu-Ta and Al-Ta precipitates. Further work in this area will be carried out. This piece of information will be very useful in determining the optimum properties of the alloy.

Acknowledgements

The authors are grateful to Mr. Marcus Chai for generating the AFM images. KL Pey would also like to thank Institute of Microelectronics for providing the opportunity to work with Fudan University on this project.

Reference

1. Davids D.A., Chan K.H. and Bertom K.L. "Temporal and spatial response in thermal transfer printing", Proc. SID, Vol. 28/4, 1987, pp 465-470.

2. Chang L.S. "Mechanisms of failure of thermal, ink-jet thin-film devices under stressed conditions", Proc. SID, Vol. 30/1, 1989, pp. 57-63.

3. Guy A.G., In *Introduction to Materials Science*, McGraw-Hill, Singapore, 1972.

4. Brooks C.R., In *Heat treatment, structure and properties of nonferrous alloys*, Metals Parks, Ohio, American Society for Metals, 1982.

5. Lancker M. Van, In *Metallurgy of aluminium alloys*, Chapman and Hall Ltd, London, 1967.

EVALUATION OF LOW-VOLTAGE ZnO VARISTORS QUALITY BY LOW-FREQUENCY NOISE CHARACTERISATION

F. Verdier, S. Gouverneur, C. Lucat, A. Touboul

IXL (URA 846-CNRS), Université Bordeaux I
351 Cours de la Libération, 33405 TALENCE CEDEX (FRANCE)
Fax : (33) 56 37 15 45, Tel. : (33) 56 84 65 40/41

Abstract : ZnO low voltage varistors are of great interest for the protection of electronics systems against transient surge overvoltage. In the opposite of classical destructive reliability tests, reliability indicators allow the determination of device reliability with non destructive stress conditions. Low-frequency noise power, controled by the conduction mechanisms in Schottky barriers of grain boundaries between ZnO grains, has been related to the main device parameters, i.e. the non-linearity coefficient and the leakage current. Measurements have been performed on samples of different technologies, before and after high energy pulse tests in order to correlate excess noise and degradation of the devices.

1. INTRODUCTION

Varistors are electronics devices which resistance decreases strongly when increasing the applied voltage. Such a property presents a great interest to protect electronics systems against surge overvoltages. Varistor effect is characterised by a fast current density change from 10^{-2} to about 10^{4} A.m^{-2}. In this domain, I(V) dependence can be described by the empirical equation (Ref.1) :

$$I = (V/C)^{\alpha} \qquad [1]$$

where α is called the non-linearity coefficient and is a criterium of efficiency of the device.

Several theories have been developped to explain conduction mechanisms and model ZnO varistors behaviour. They point out the fundamental role of a microstucture constituted of several semiconductor grains and intergranular insulating boundaries. The main non-linearity effect is attributed to the boundaries where threshold voltage is about 3 V. In the application we are involved, low threshold voltage values must be achieved (less than 20 V), so that thin ZnO layers are required (20 or 30 µm for a mean grain size of 3 to 4 µm). Screen-printing thechnology fits well this requirement and presents many advantages :

- The desired threshold voltage is controlled by the layer thickness.
- Several protection elements can be deposited on a single substrate, allowing an individual protection of several I/O (like connectors).
- Hybrid components, with various electronics functions, as well as discrete components can be implemented.

Usual methods to determine varistors reliability consist in statistical destructive tests under rated single pulses of high energy (lightening waves, 0.1x6.4 µs, 8x20 µs...) defined by their shape and duration (Ref.2) and representative of risks in operating conditions. An alternative solution is the use of reliability indicators, such as C(V) characterization as a function of the frequency, which should be measured under lower stress conditions.

Electrical properties related to quality and reliability of low-voltage varistors have been investigated here on the basis of excess low-frequency noise measurements. Related to the conduction mechanisms in Schottky barriers between ZnO grains, the spectral intensity of current noise is found to be linked to the principal parameter of quality of a nonlinear clamping device : the non-linearity coefficient α.

2. TEST SAMPLES

Most of measurements have been performed on new technology varistors, built at IXL Laboratory, consisting in thick ZnO screen-printed layers on an alumina substrate, with small addition of other metals. Varistor ink is composed of ZnO powder, organic binder and a few percents of doping impurities ($CoCO_3$, Cr_2O_3, Bi_2O_3, $MnCO_3$, Sb_2O_3). A Platinium common electrode is first deposited and annealed at a temperature of 950°C. Varistor ink is then printed in successive layers, dryed during 15 min at 120°C (Figure 1).

Screen-printed layers	V_{th} (V)
1	3 - 10
2	10 - 20
3	20 - 35

Figure 1 : Threshold voltage *vs* number of screen-printed ZnO layers

A mechanical pressure (either uniaxial or isostatic, with a maximum of 6.2×10^{8} Pa) is included in the fabrication process to control the clamping voltage by densifying the screen-printed films without the use of

inorganic binder. Before the deposition of six Silver electrodes, samples are put in an oven at 1150°C during 15 min under controlled atmosphere. Two final annealing processes (850°C and 650°C) stabilise the active materials (Ref.3). Test vehicles are encapsulated in a 7-pins SIL package as shown in figure 2.

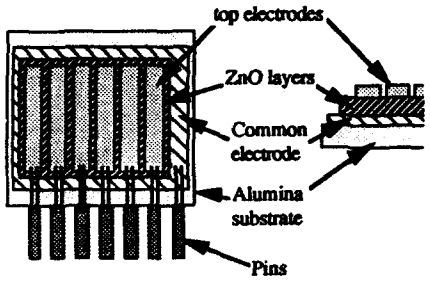


Figure 2 : 7 pins SIL Test stucture

Commercial ceramic varistors from Harris (bulk), Draloric (bulk), AVX (multilayer), in the same range of threshold voltage (18-25 V) have been compared to our samples (Ref.4) on the basis of their electrical and noise properties.

3. ELECTRICAL CHARACTERISATION

3.1 I(V) characteristics

I(V) characteristics of a ZnO varistor can generally be divided in three or four zones according to the range of voltage applied to the devices (Figure 3).

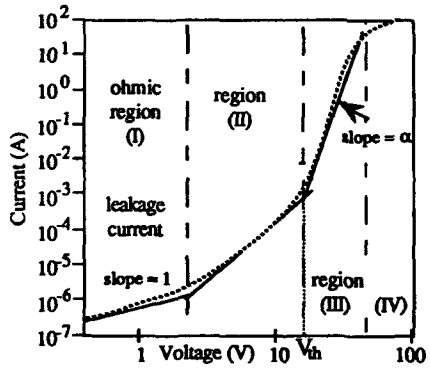


Figure 3 : Wide range current I(V) characteristics of a screen-printed varistor

A peculiar attention must be paid to the region I (ohmic region) where the leakage current reflects microstructure quality (Figure 4), and the region III where the non-linearity coefficient α modelling the I(V) dependence [1] describes the efficiency of the protective element.

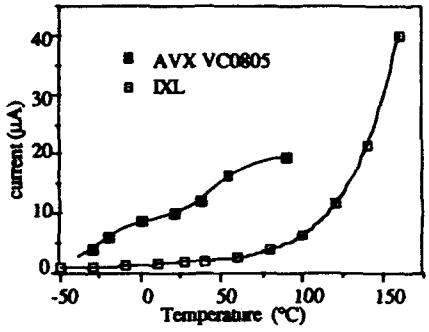
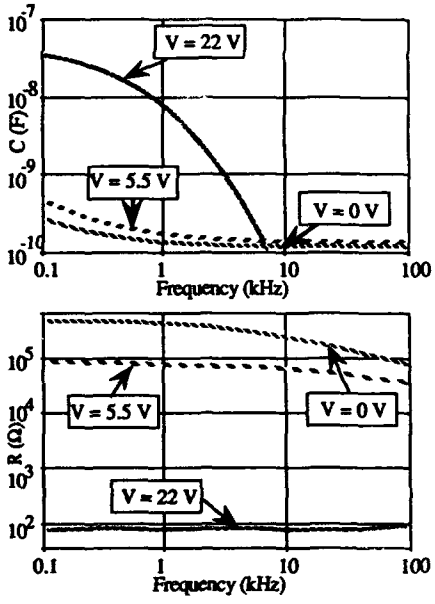


Figure 4 : Leakage current of a screen-printed varistor compared to a ceramic one, *vs* temperature (V = 5.5 V)

3.2 Frequency response

Varistor impedance can be modelled at low frequencies (below a few 10 kHz) by an equivalent parallel RC circuit where R and C depend on frequency (Ref.5). This impedance is measured in our case to correct the inherent filtering effect when studying background current noise between 10 Hz and 10 kHz, by introducing this equivalent circuit in the calculation of noise measurement set-up frequency response. A typical result is presented on Figures 5a and 5b.



Figures 5 : Equivalent circuit parameters for frequency response, *vs* frequency and bias conditions
Figure 5.a : Equivalent parallel capacitance
Figure 5.b : Equivalent parallel resistance

A strong dependence on bias condition is observed, for dynamical resistance R of course, but also for parallel capacitance C.

3 3 Conduction mechanisms and excess noise sources in varistors (Ref.6)

Conduction mechanisms in complex structures like ZnO varistors is identified as a conduction of Schottky type through the barriers of thin amorphous zone between the ZnO grains. It can be defined an equivalent number of junctions in series : in our case of low voltage drop, the mean number of ZnO grains is about eight. Considered as a junction noise, current noise spectral intensity has been modelled in the ohmic region (Ref.7) according to the empirical Hooge relation :

$$\frac{S_I}{I^2} = \frac{K R_0 C_0}{A_{eff} N_d} \frac{1}{f} = \frac{U}{f} \qquad [2]$$

where R_0 et C_0 are the equivalent parameters of the ohmic region in the frequency range, A_{eff} is the equivalent conduction section, N_d is the donors number in doped ZnO, K is a factor including microscopic electrical parameters.

Under higher currents (10^{-4}, 10^{-3} A), i.e. in region II and at the beginning of region III, excess current noise spectra exhibit one or more cut-off frequencies which depend on the bias conditions (Figure 6).

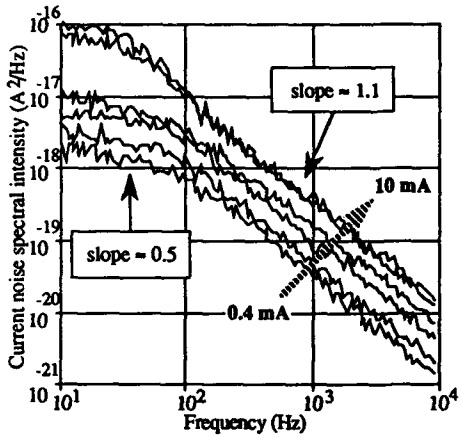


Figure 6 : Noise spectra *vs* bias current (from 0.4 to 10 mA, T ≈ 300 K)

Noise slope is about 0.5 below the knee, nearly pure 1/f noise is found for higher frequencies. Furthermore, quadratic dependence on current is not verified for current corresponding to the most non-linear region, around threshold voltage, where $S_I(f)$ is rather proportional to I (Figure 7). A correlation has been found between the non-linearity coefficient and the normalised excess noise level U (evaluated before or after the cut-off frequency) : the higher the non-linearity coefficient, the lower the noise level (Figure 8).

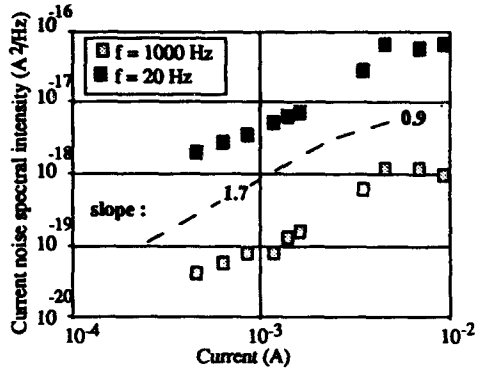


Figure 7 : Dependence of noise level on bias current (at f = 20 Hz and f = 1000 Hz)

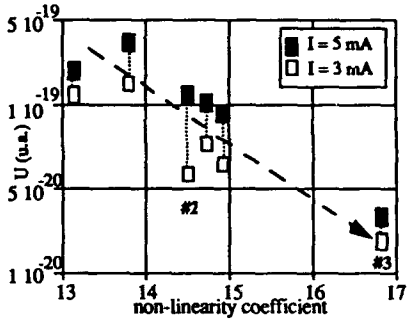


Figure 8 : Normalized noise level (K) for 6 screen-printed varistors at two bias currents *vs* non-linearity coefficient

Noise measurements performed on varistors of other technologies (bulk, multilayers) showed that both noise level and spectrum shape (number of cut-off frequencies, slopes...) cannot be compared from one sample to another. But the relationship between low-frequency noise level (below 100 Hz) and the non-linearity coefficient does not significantly vary.

4. VARISTORS DEGRADATION

4.1 Burn-in tests

Screen-printed varistors have been submitted to standard burn-in pulses, described by their duration, maximal ratings (voltage/current). The calibrated 0.1 µs/6.4 µs lightening wave, representative of short overload risks for electronics and power systems (Ref. 8-9) has been applied 5 times to the samples

(5 varistors in a SIL package, sample n°5 is kept for reference), with a period of one minute between pulses.

A typical response for pulse current and clamped voltage of the varistor is presented on figure 9.

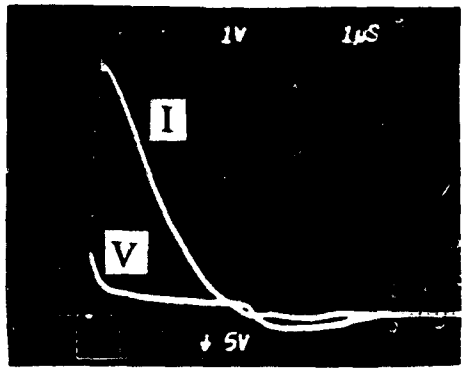


Figure 9 : I(t) and V(t) response to 0.1 µs/6.4 µs lightening waveform

4.2 Damage characterisation

Failures or degradation have been characterised. Damaged samples have to be divided in two groups : the first one includes totally failed varistors (n°1, 4, 6) which present open-circuit. An optical observation shows a partial destruction of the upper or lower electrode of the device due to a high current spot.

The other varistors (n° 2 and 3) do not present letal failure, but degradation of their electrical parameters (Ref. 10-11) are measured : increase of the leakage current (x 10^3), decrease of the nonlinearity coefficient α (Figure 10).

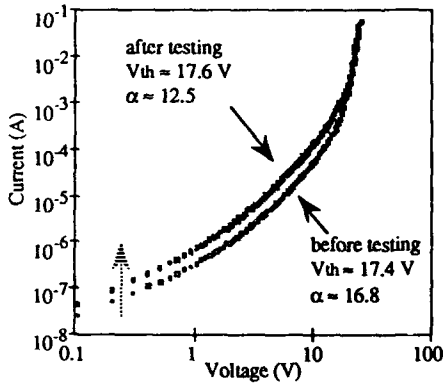


Figure 10 :I(V) characteristics of a varistor before and after burn-in test (sample n°3)

On the other hand, threshold voltage presents a rather low variation of less than +1.5%. Electrical parameters are reported on figure 11 for both unfailed samples.

parameter	threshold voltage		non-linearity coefficient	
sample n°	2	3	2	3
before test	17	17.4	14.4	16.8
after test	17.1	17.6	12	12.5

Figure 11 : Comparison of electrical parameters before and after testing

4.3 Noise measurements after burn-in test

Noise measurements performed on the damaged samples reflect the degradation of α (Figure 12).

A variation of -25% for α of sample n°3 results in an increase of a decade and half in noise ; for sample n°2, noise increase is about half a decade for a change of -16% for the non-linearity coefficient.

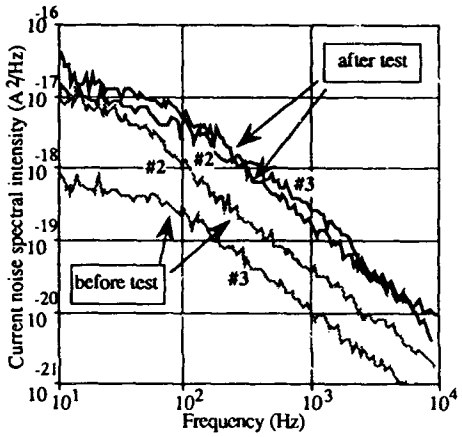


Figure 12 : Noise spectra before and after test (samples n°2 and 3, I = 3 mA)

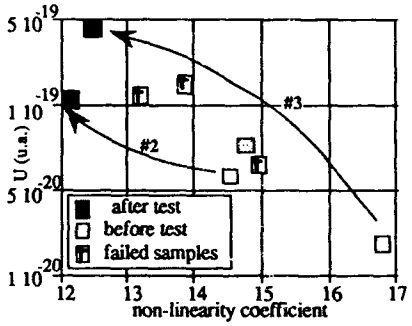


Figure 13 : Normalized noise level (U) *vs* non-linearity coefficient (before and after burn-in test, I = 3 mA)

If we superimpose noise data : normalised noise U *vs* non-linearity coefficient α, to the previous data (Figure 8), we notice that the points after test are

located according to the same trend (Figure 13). So, we can suppose that the variation of noise is strongly correlated to the variation of α, i.e. to the micromechanisms of conduction in the varistor, either at the electrode contact (where failures occur for most of samples), either in the bulk (grains of ZnO).

5. CONCLUSION

The most important characteristics of a ZnO varistor is its non-linearity coefficient which can be measured only in the breakdown region, that means under rather high currents. However, although such devices are able to stand high energy in surge conditions, during very short duration, the small volume of screen-printed varistors does not allow a high mean power dissipation. An indicator of the value of α (as excess noise) which is significant at very low current, i.e. without destructive Joule heating, can be used to characterise devices quality. Excess noise measurement include indeed both noise sources behaviour (microstructural effects) and macroscopic evolution of the component (frequency response of the varistance, equivalent resistance and capacitance). The electronics functionality of the ZnO varistor should be evaluated when taking into account the noise level and hence the consequences on the reliability.

6. REFERENCES

1. M. Matsuoka, Jap. J. Applied Physics, 10(6), 1971,pp 736

2. "Transient Voltage Suppression Devices", HARRIS Semiconductor, Data and Application Book, 1992

3. S. Gouverneur, "Mise au point d'un nouveau procédé de densification des couches épaisses sérigraphiée. Application à la fabrication de composants passifs", Thèse de l'Université Bordeaux I, 1994

4. S. Gouverneur, C. Lucat, F. Menil, F. Verdier, A. Touboul, J.L. Aucouturier, J. Pinel "Electrical Response of Screen-printed varistors to Transient Overvoltages", Proc. of Lightening and Mountains, Chamonix (France) 1994

5. G. E. Pike,"Electronics Properties of ZnO Varistors : a New Model", Grains Boundaries in Semiconductors, Nov. 1981, pp 369-379

6. L. K. J. Vandamme, "Conduction Mechanisms in ZnO Varistors", J. Appl. Physics, April 1980, pp 4240-4244

7. A. Kusy, "1/f Noise in ZnO Varistors", Proc. of Noise in Phys. Systems and 1/f Noise, Montpellier (France), 1983, pp 213-216

8. J.M Wu, J.J. Shyu, "Electrical properties and degradation improvement of ZnO varistors doped with $PbO-B_2O_3$ based glasses", J. Mat. Sc, Vol. 24, 1989, pp 1881-1888

9. A. Vicaud, D. Eguizabal, "Caractérisation électrique des céramiques ZnO pour parafoudres", Proc. of Journ. d'études SEE-Club Etudes Gén., March 1986, pp 85-87

10. L. J. Bowen, F. J. Avella, "Microstructure, Electrical Properties, and Failure Prediction in Low Clamping Voltage Zinc Oxide Varistors", J. Appl. Physics, 54(5), May 1983, pp 2764-2772

11. P. Manuel, "Mécanismes de conduction et de degradation dans les varistances ZnO : Revues des Différents Modèles", Proc. of Journ. d'études SEE-Club Etudes Gén., March 1986, pp 25-33

SEMICONDUCTOR STRUCTURE CHARACTERIZATION BY FOCUS ELECTRON AND OPTICAL BEAMS

V.Aristov, E.Rau, E.Yakimov.

Institute of Microelectronics Technology and High Purity Materials of Russian Academy of Sciences
Chernogolovka, Moscow distr., 142432 Russia

1. INTRODUCTION

Methods of scanning electron (SEM) and optical (SOM) microscopy are very promising for the nondestructive reconstruction of three-dimensional internal structure of objects under investigation with spatial resolution in micron and submicron range. Such possibility is of great importance for semiconductor structure inspection, for the characterization of technology processes and for failure analysis. The investigations carried out during last years allow to develop the background for computer electron-beam and laser tomography. Some special methods of "apparatus" tomography in which the reconstruction of physical properties distribution or of internal structure of the object under study was achieved by means of specially designed setup were proposed. In the present paper the possibility of electron-beam and optical tomography have been discussed.

2. SOM TECHNIQUES

The setup for the transmitted, scattered and polarized scanning infrared microscopy designed in the Institute of Microelectronics Technology has been described (Ref.1). The reflected or scattered radiation is collected by the elliptic mirror optics. Such confocal mode of IR tomography allows to achieve resolution about 1 micron. The detection of Rayley and small-angle scattering allows to reconstruct the spatial distribution of microprecipitates. IR images in the transmission mode gives an information about the nonhomogeneities in dopant distribution, nonhomogeneous distribution of thermal properties and metallization topology.

3. SEM TECHNIQUES

3.1. Electron-beam tomography in the backscattering electron mode

Electron-beam tomography in the backscattering electron (BSE) mode is very promising for the inspection of multilayer structures (Ref. 2). It has been shown that the measurements of (BSE) coefficient dependence on electron beam energy or of BSE energy spectrum allow to obtain thicknesses in the range from 0.1 to some microns. Compact high-resolving energy filtration of BSE and separate the signals formed in subsurface layers at different depth with submicron resolution.

3.2. EBIC with depletion region width modulation

This new in-situ differential technique allows (Ref. 3) :
- to map diffusion length and depletion region width two-dimensional distributions simultaneously,
- to reconstruct the diffusion length profile with spatial resolution in depth better then 0.1 micron,
- to reconstruct the dopant distribution with depth resolution compared with that of the C-V method and lateral resolution about 1 micron,
- to measure diffusion length profile in thin layers with thickness smaller than diffusion length.

3.3. CL with energy and intensity modulation

This modulation cathodoluminescence technique has submicron depth resolution and allows (Ref. 4) :
- to separate the spectra from any layer of two- or three-layer structure,
- to measure the layer thicknesses in such structures,
- to reconstruct the radiative recombination center depth distribution.

3.4. Induced surface EMF

This technique bases on contactless measurements (Ref. 5) of the electric potential induced by the focused electon beam and allows to reveal the inhomogeneities in doping level, recombination rate and so on without contact or other special sample preparation.

4. MULTI-FUNCTIONAL SCANNING PROBE

Scanning probe with dimensions in the micrometer range with metallic needle or some microsensor, e.g. thermocouple, has been designed (Ref. 6). This probe allows to map the spreading resistance, topographic and potential relief, to study the distribution of temperature fields, doping level, lifetime, dielectric constant, surface charge and so on. The results demonstrating the possibilities of the techniques for the semiconductor structure characterization an failure analysis are presented.

REFERENCES

1. Aksenov L., Aristov V., Frolov K., Rau E. "Infrared Introscopy and Microtomography of Semiconductor Structures", Proc. SPIE, 1991, v. 1843, pp. 288-300.

2. Dremova N.N., Drokin A.P., Zaitsev S.I., Rau E.I., Yakimov E.B."Characterisation of multilayer microstructures and surface relief in bakscattered elrctrons in the scanning electron microscope", Izv. RAN, ser. Fiz.,1993, v.57, No 8, pp.9-14 (in russ.).

3. Kononchuk O.V., Yakimov E.B. "Mapping of diffusion length and depletion region width in Shottky diodes", Semiconductor Sci.Tecnol., 1992, v.7, pp. A171-A174.

4. Kireev V.A., Razgonov I.I., Yakimov E.B. "Possibilities of Modulated Cathodolumonescence for Multilayer Structure Characterisation", Scanning, 1993, v.15, pp. 31-36.

5. Aristov V.V., Kononchuk O.V., Rau E.I., Yakimov E.B. "SEM investigation of semiconductors by the capacitance trchniques", Microelectronic Engineer., 1990, v. 12, pp. 179-185.

6. Aristov V.V., Andrianov M.V., Rau E.I. "Possibilities and future trends of scanning sensor probe multiscan", Izv. RAN, ser. Fiz.,1993, v. 57, No 8, pp. 2-8 (in russ.).

AN EMBEDED CHARACTERIZATION STRUCTURE FOR ASIC FAILURE ANALYSIS AND PERFORMANCE ASSESSMENT [1]

C. NOUET - G. ROLLAND* - M. BARRE

MATRA DEFENSE - 20/22 rue Grange Dame Rose - 78141 VELIZY - FRANCE
Tél : (1) 34.88.19.77 / fax : (1) 34.88.19.88

*CNES - 18 Avenue Edouard Belin - 31055 TOULOUSE - FRANCE
Tél : (33) 61.27.36.64 / fax : (33) 61.27.47.32

ABSTRACT

To support ASIC validation, debug and failure analysis, a characterization structure called MICROMOS was developed to be integrated on ASIC at the design stage. MICROMOS with use of SPICE modelling parameter extraction allows to extract SPICE parameter values at the die level by probing electrical characteristics of characterization patterns. These SPICE parameter values being correlated to the process control of the run and to the die position on the wafer, comparison with the SPICE parameter specification of the manufacturer, on different circuits and different runs, allows to perform process parametric conformity analysis of the circuits evaluating position and spread of the process versus functional and parametric margins of the circuit.

MICROMOS and SPICE parameter extraction procedures has been developed and validated on a test vehicle with a 1.2 µm process. They are already used in MATRA DEFENSE's ASIC designs to built-in analysis. In the same way, the French national space agency CNES uses MICROMOS in ASIC for research and development activities and plans to extend use of MICROMOS in operational circuits.

Future works will complete this approach by developing electrical failure simulation of the ASIC or a part of ASIC (the faulty localised structure), with use of extracted SPICE parameter values.

INTRODUCTION

Without considering intrinsic failure mechanisms, CMOS ASIC performance and "a fortiori" CMOS ASIC failure are dependent on design specification conformity, fab defect density and process specification conformity. For an ASIC designer and user, well controlled techniques exist like automatic testing, electron beam testing and physical analysis to validate design specification conformity and fab defect density, but finally no single one can be used to validate process specification conformity, or to diagnose failure when no hard defect is visible on the die but where process drift or critical design is the cause [1].
It can be achieved by allowing SPICE parameters extraction on each circuit, from prototype to circuit utilisation, then by performing conformity analysis of the "real SPICE data base" with the SPICE parameters specification of the manufacturer to diagnose process parametric conformity defect.
To reach this goal, a test structure called MICROMOS and a SPICE 2G.6 level 3 parameters extraction-optimisation procedure was developed, based on UTMOST III transistor modelling software. The mission of MICROMOS is to be integrated in MATRA'S CMOS ASIC designs to be the memory of the wafer process, read at any time of the circuit life, for failure analysis, for process validation at the prototype level and for process quality indicator at the production level through SPICE parameter specification conformity analysis.

SPICE PARAMETER SPECIFICATION CONFORMITY

SPICE parameter specification of a manufacturer defines the typical and worst case values for SPICE

1 This work was supported by the French National Space Agency CNES under contract 846/CNES/92/1778/00.

parameters which are representative of the process variation limits [2]. These typical and worst case values are related to typical and worst case timing characteristics of transistors in terms of switching time. In fact, a SPICE parameters specification is made of four databases representative of the limits of the process variation, in term of transistor switching performance (tab 1).

NMOS	PMOS	SPECIFICATION
FAST	FAST	FF
FAST	SLOW	FS
SLOW	FAST	SF
SLOW	SLOW	SS

Tab.1 : SPICE parameter specification

To extract PMOS and NMOS SPICE parameters values on a circuit, is to be able to measure process position between the process variation limits (fig. 1) and so to evaluate process specification conformity on a circuit.

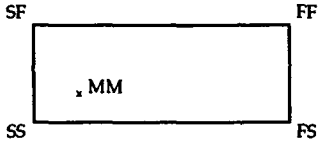


Fig.1 : Process Position. MM (Measured-Measured)

MICROMOS DESCRIPTION

The study has identified an optimum of 22 SPICE parameters in a specification to provide accurate DC and transient analysis with SPICE electrical simulation [3]. A characterization structure, MICROMOS, has been specified to measure the value of these parameters using MOS transistor modelling parameters extraction. MICROMOS was designed with a minimum area in order to be integrated in any CMOS ASIC design, and to be probed at the die level. It is divided in two parts to provide extraction of PMOS parameters and NMOS parameters, these two parts MICROMOSN, and MICROMOSP having the same geometry but different substrate polarity. Here, a MICROMOS (N) description is presented for a 1.2 µm process (fig. 2).

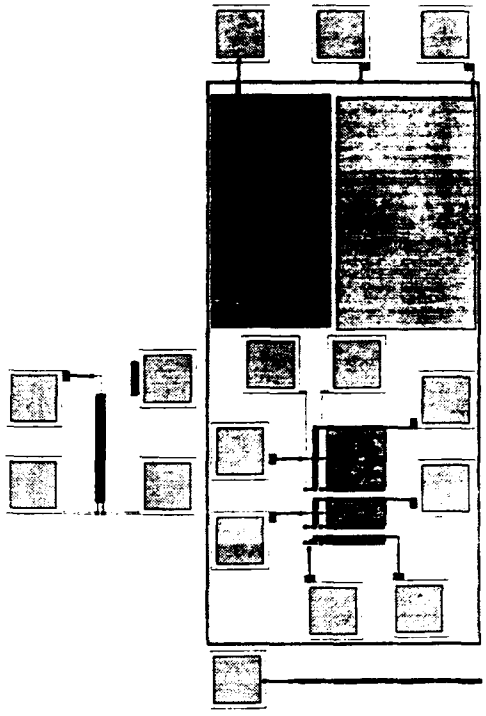


Fig.2 : MICROMOS (N) lay-out

MICROMOS(N) is made of five patterns which allow to extract the 22 SPICE parameters :

- CAPAMOS is a MOS thin oxide capacitance pattern wich allows to extract TOX (and other parameters like flat band voltage vfb, flat band capacitance cfb,...) with C(V) measurements. Its dimension is (200 x 100) μm^2.

- CAPA is a two junctions pattern. One junction bottom capacitance which allows to extract CJ, MJ, PB, FC with C(V) measurements. Its dimension is (200 x 120) μm^2. And one junction sidewall capacitance which allows to extract CJSW, MJSW with C(V) measurements. Its dimension is (1500 x 1.2) μm^2.

- OVER is a MOS thin oxide capacitance pattern wich allows to extract CGBO (with C(V) measurements. Its dimension is (160 x 1.2) μm^2.

- TRANSLAP is a MOS transistor pattern which allows to extract CGSO, CGDO with C(V) measurements. Its dimension is W = 90 µm, L = 1.2 µm.

- TRANS is a MOS transistor matrix pattern with six transistors which have different geometries. They allow to extract VTO, NSUB, UO, Rs, Rd, LD, THETA, WD, DELTA, ETA, VMAX, NFS with I(V) measurements depending on the different geometries (tab. 2).

50/50	50/20	50/1,2
3/50	3/20	3/1,2

Tab.2 : transistor dimensions in µm (W/L)

One advantage to design the six transistors in a matrix is to reduce the pad quantity from 18 pads for elementary transistors to 8 pads.

The patterns of MICROMOS are connected to probing pads with small dimensions (50 x 50) μm^2 to minimize occupied surface. All the patterns make a MICROMOS total area of 0.385 mm^2 for a 1.2 µm process. It is negligible, considering MATRA DEFENSE's ASIC which are essentially pad limited with an average area of about 1 cm^2. Moreover, the test structure pads can be reduced to (15 x 15) μm^2 without problems for probing.

SPICE PARAMETER EXTRACTION - OPTIMIZATION PROCEDURES

To extract SPICE parameters from MICROMOS implanted in a circuit, extraction-optimization procedures were devellopped on the basis of a testing bench based on UTMOST III transistor modelling software [4], a probing station, a C(V) analyser and a I(V) parameter analyser (fig. 3).

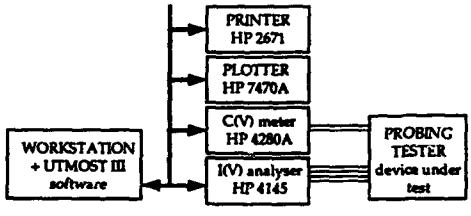


Fig.3 : test bench

The procedures were developed in order to keep a physical meaning of the measured parameters while allowing accurate electrical simulation of the transistors' characteristics [5].

The extraction routines of the UTMOST III software are based on the decomposition property of the SPICE 2G.6 Level 3 model for particular geometrical and bias effects [6]. Each of these routines needs a specific pattern and particular bias conditions to measure appropriate electrical characteristics on which a simplification of the SPICE Model can be applied. So using this simplified model, a parameter or a set of parameters representative of the geometrical or bias effect can be extracted (fig.4).

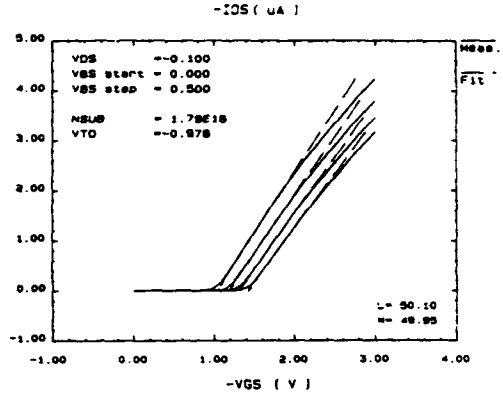


Fig.4 : NSUB, VTO parameter extraction

The developed extraction procedure defines the measurement process using these routines, and for each routine, it defines the optimal bias conditions to extract parameters from adequate electrical characteristics and patterns. The application of the extraction procedure results is a set of initial values for the 22 parameters, providing a first accurate simulation of the transistors' characteristics.

Optimization procedure development is based on a SPICE parameters influence analysis for each electrical characteristic. It was performed with SPICE model analysis and electrical simulations in order to define for each characteristic and each geometry, the parameters' influence and so, the best set of parameters to optimize. The result is a parameters' optimization procedure, using optimization software of UTMOST III [4], which allows to minimize errors between measured characteristics of the different geometries of transistors, and the simulated characteristics (fig.5).

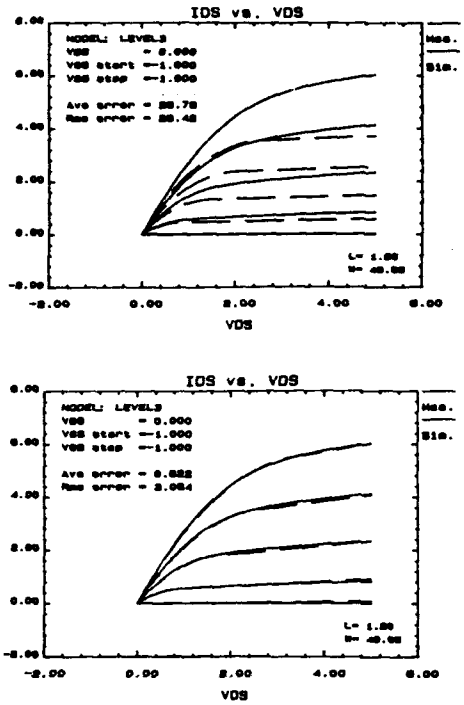


Fig. 5 : Optimisation result on 50/1.2 transistor

Finally, extraction-optimization procedures applied on MICROMOS, allow to access SPICE parameter values representative of the process position (tab. 3) for the circuit under test.

TOX (m)	2.73E-08	KAPPA	3.318
RS,RD (Ω)	29.4	ETA	0.1026
VTO (v)	0.9594	CJ (F/m^2)	9.90E-05
NSUB (cm^{-3})	1.86 + 16	PB (v)	0.886
UO (cm^2/v.s)	197.51	MJ	0.51
THETA (v^{-1})	0.0999	FC	0.5
LD (m)	1.27E-07	CJSW (F/m)	6.70E-10
XJ (m)	3.50E-08	MJSW	0.434
WD (m)	1.5E-07	CGSO (F/m)	2.00E-09
DELTA	0.8683	CGDO (F/m)	1.95E-09
VMAX (m/s)	2.56E + 05	CGBO (F/m)	4.10E-07
NFS (cm^{-2})	4.8E + 11		

Tab. 3 : PMOS SPICE parameter values (1.2 µm process)

This real SPICE database provides accurate electrical simulation for the different geometries with an average root mean square error between measured and simulated characteristics of about 6 % (fig.6).

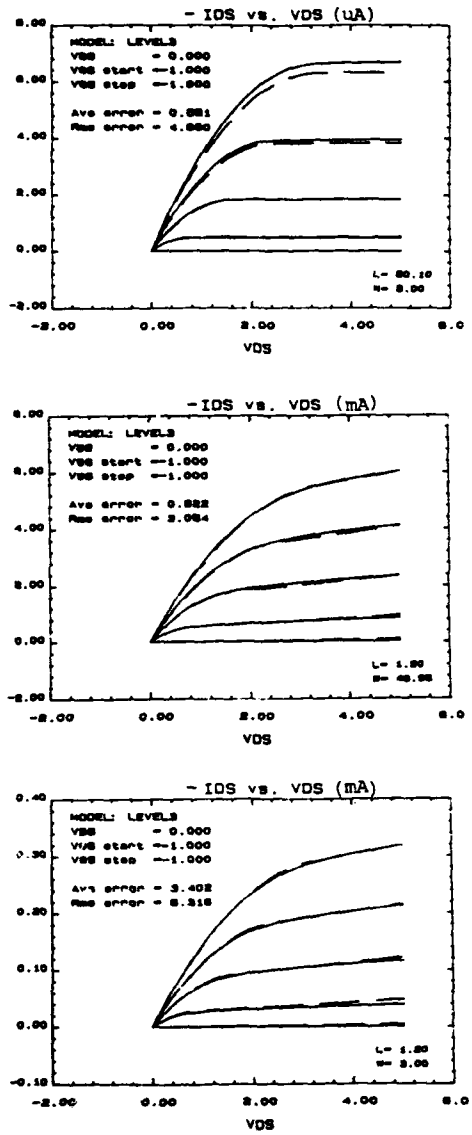


Fig.6 : Electrical Simulation with measured SPICE parameters
--- simulated characteristics
measured characteristics

APPLICATION ON THE ASIC LIFE CYCLE

To give the most confidence in MICROMOS integration and the ASIC manufacturing, discussion with the manufacturer has to begin at the earliest stage of the ASIC project to present the mission of MICROMOS and to define the best way for the two parties to benefit from the results.

MICROMOS is designed in accordance with the manufacturer design rules, and it is validated before integration in the ASIC design by performing design rules checks (DRC). As MICROMOS is not an active cell, it is completely independent of the ASIC functionality and structure. So integration of MICROMOS at the ASIC design stage will be performed before the last verifications (DRC, ERC) of the ASIC design, considering that a free area has been provided for.

After ASIC manufacturing, MICROMOS and SPICE parameter extraction procedures are used as a tool for prototype performance assessment using process specification conformity validation, for incoming quality control in the production phase using process quality indicators, and for failure analysis to identify process specification conformity defects as a possible cause for failure of the circuit.

Prototype performance assessment is achieved through ASIC design margins validation. This validation is performed with two complementary tasks :

- performance margin characterization at circuit level and margin validation versus the simulated functional and parametric characteristics (data sheet)
- process specification conformity validation.

These two tasks must give confidence in ASIC performance capability, considering the process will stay under control.

For a manufacturing process identified as critical in terms of maturity, incoming quality control is achieved during ASIC production phase monitoring on samples some SPICE parameters representative of the process, they are the process quality indicators (TOX, VTO,...). It gives confidence in process variability and it allows to identify process drift before ASIC performance characteristics get out of the data sheet.

Failure analysis can be performed over all the ASIC life cycle to identify a process specification conformity defect when it is the cause of a functional or parametric failure of a circuit. The nature of the parameter, or set of parameters, out of specification gives major information to identify process steps where potential drift occured.

Future work will complete this approach developing electrical failure simulation of ASIC or a part of ASIC (the faulty localised structure), with use of extracted SPICE parameter values. Another field investigated is to use process specification conformity to support library's cells validation [7].

CONCLUSION

A characterization structure called MICROMOS has been developed for ASIC performance assessment and failure analysis using SPICE modelling parameter extraction to provide measurement of the "real SPICE parameters database" of a circuit. It is achieve with integration of MICROMOS at the ASIC design stage, then by the application of extraction-optimization procedures on the manufactured circuit. The extracted SPICE parameters database is used to evaluate process specification conformity as a first input for ASIC design margin validation, for in-coming process quality indicators, and for failure analysis.

As a result of this study, CNES is using MICROMOS for ASIC in research and development activities and plans to extend its use to operational circuits.

For MATRA DEFENSE, MICROMOS is used to support validation of the operational ASIC prototype, and for failure analysis. The first ASIC integrating MICROMOS was a 25,000 gates in 0.8 µm CMOS technology. The next step is to implement process quality indicators for the production phase, and to develop electrical failure simulation concepts.

REFERENCES

[1] "ASIC session"
- BARRE - NOUET / MATRA DEFENSE - TRIAS / ICEu - FORGERIT / ES2
- SEE WORKSHOP ON FAILURE ANALYSIS 1990-1992

[2] "Statistical worstcase MOS parameter extraction"
- BOLT - TRIP - VERHAGEN - PHILIPS RESEARCH LABORATORIES
- IEEE/ICMTS - 1989 -

[3] "Simulation of MOS integrated circuits using SPICE 2"
- VLADIMIRESCU, LIU
- BERKLEY - 1980 -

[4] "UTMOST : Modelling and User Manual"
- SILVACO INTERNATIONAL
- SANTA CLARA - 1992 -

[5] "Contribution à la caractérisation et à la modélisation des technogies MOS"
- RINAUDO
- Thèse - Laboratoire IXL - UNIVERSITE DE BORDEAUX - 1993 -

[6] "Subtleties of SPICE Mosfet Parameter Extraction"
- BENDIX-SEMICONDUCTOR OPTIMIZATION AND SIMULATION INC.
- IEEE/ICMTS - 1989 -

[7] "Timing characterization of library's logical operators with E-Beam techniques"
- NOUET, BARRE/MATRA DEFENSE - BRIET/MATRA MHS
- ESREF 91

Ion Implant Charging and Machine Fingerprinting Methodology

C.Messick [1], S.Reno [1], J.Shideler [1], W.Lukaszek[2], W.Dixon [2], M.Vella [3], J.Reedholm Snr [4]

[1] National Semiconductor, West Jordan, USA
[2] Stanford University, Stanford, USA
[3] Lawrence Berkeley Labs., Berkeley, USA
[4] Reedholm Instruments, Freemont, USA

No text available

THE USE OF DAVINCI FOR THE MEASUREMENT OF THE TRAP ENERGY LEVEL AND SURFACE STATE DENSITY: A NEW APPROACH FOR QUALITY EVALUATION OF IC's

I. Hachicha, P. Fouillat, J.P. Dom, H. Levi

IXL (C.N.R.S. URA 846), Université Bordeaux 1
351 Cours de la libération, 33405 TALENCE CEDEX (FRANCE)
Fax : (33)56371545, Tel.: (33)56846541

Abstract: A method is described that provides a good quality process evaluation of a bipolar transistor by the measurement of its trap energy level in the base and its surface state density. It is performed by focusing on the experimental Gummel plots and current gain of a bipolar transistor compared to computed data from the physical device simulator DAVINCI. A test vehicle has been designed with the SUBILO N technological process from PHILIPS with LOCOS isolation technique. A BICMOS technological process from THOMSON with a P^+ wall isolation technique has also been tested.

1. INTRODUCTION

The measurement of the trap energy level and of the surface state density is necessary to evaluate the technological process quality when fabricating bipolar transistors. In a classical way, the trap energy level may be characterised by the following method: low frequency generation-recombination excess noise spectroscopy, mutual conductance g_M-frequency dispersion spectroscopy, low frequency oscillations and conductance-deep level transient spectroscopy (C-DLTS) [ref.1]. Trap energy levels (Etrap) appear when impurities are introduced into the base of the device. Surface state densities (Nst) are generated during the oxydation steps. Assuming the fact that the SRH recombination process modeled by formula (1) is a performant way to illustrate the influence of these parameters (Etrap, Nst), then we can propose a new method for their evaluation which combines the use of a three dimensional physical simulator such as DAVINCI, with the use of a classical measurement equipment. The main problem is to choose the experiments which highlight these parameters effects. Then, fitting experimental and simulated curves by the use of standard optimisation techniques will lead to correct values of the SRH model parameters. In order to reach this aim, a study of the SRH model parameters contribution to the Gummel plots and current gain is necessary.

2. CONTRIBUTION OF THE SRH MODEL PARAMETERS TO THE GUMMEL PLOTS

The SRH recombination process is described by the well known formula reported by the authors[ref.2].

$$U_{SRH}=\frac{np-n_i^2}{\tau_n\left(n+n_i e^{\frac{Ei-Et}{kT}}\right)+\tau_p\left(p+n_i e^{\frac{Et-Ei}{kT}}\right)} \qquad (1)$$

where n (/p) is the electron (/ hole) carrier density, n_i is the intrinsic density, Et is the trap energy level, Ei is the intrinsic energy level, τ_n and τ_p are the minority carriers lifetimes for electrons and holes. Inside the volume these lifetimes depend essentially on the doping profile as reported by the authors[ref.3]:

$$\tau_n=\frac{TAUn0}{1+\frac{N(x,y,z)}{NSRHn}} \qquad (2) \qquad \tau_p=\frac{TAUp0}{1+\frac{N(x,y,z)}{NSRHp}} \qquad (3)$$

where TAUn0 is the SRH carrier lifetime for electron, TAUp0 is the SRH carrier lifetime for hole, N(x,y,z) is the local total impurity concentration, NSRHn is the SRH concentration parameter for electron and NSRHp is the SRH concentration parameter for hole. Close to the surface a reduction of the lifetimes occurs due to additional recombination phenomena which are represented by the surface recombination velocities SN and SP[ref.4]. All the difficulty is to distinguish the respective effects of the different parameters and to determine the appropriate experiments.

The influence of all the parameters is illustrated with a bipolar transistor issued from the SUBILON technological process from PHILIPS, from which doping profiles and dimensions are known. Experimental results will be given also with a component realised with the BICMOS technological process from THOMSON. All the curves in the figures (Gummel plots or current gains) are given versus the base potential which is referred to the emitter potential in the forward active region and the collector potential in the inverse active region.

3. THEORETICAL RESULTS

Figure 1 presents the influence of Et on the current gain of the Bipolar transistor. When operating in

the forward-active region (figure 1a) , no significant effects appear except for values of Et far from the intrinsic energy level (Et-Ei=-0.45eV). This is due to the fact that the recombination center is less efficient when its energy level is far from the middle of the bandgap. This induces a higher current gain because of a lower recombination current in the base. But, when operating in the inverse active region (figure 1b), the curves diverge in the low applied voltages area (Vbc<0.7V), even for low values of Et-Ei. This is due to the fact that Et acts principally on the recombination current located in the depletion region of the forward biased junction. In the case of a component forward biased, it is the emitter-base depletion region which imposes its contribution. But this one is not important because of its narrow dimensions due to high doping concentrations in the vicinity of this junction.

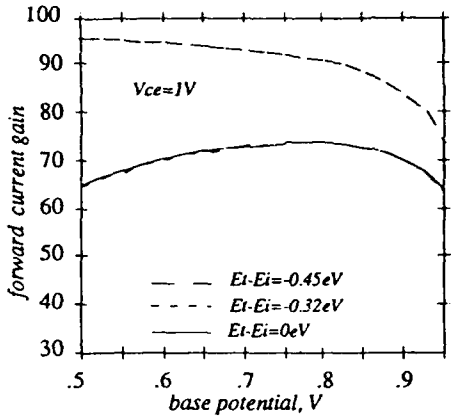


Figure 1a: Influence of Et-Ei on forward current gain

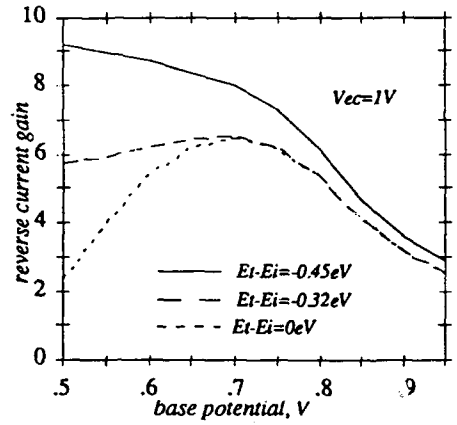


Figure 1b: Influence of Et-Ei on reverse current gain

Inversely, when the component is reverse biased, it is the collector-base depletion region which imposes more clearly its contribution because of its large dimension due to a low doping concentration near the junction. In the two previous cases (figure 1a et 1b), we have neglected the influence of the reverse biased junction because of its low contribution due to a high electrical field implying a low probability of recombination in these very depleted area[refs.5,8].

Figure 2 gives the influence of the surface recombination phenomena modeled by different values of SN and SP. In the forward active region (figure 2a) the curves are translated as SN and SP varies. This is because the surface degradations (SN=SP=1e5cm/s) generate traps at the interface Si-SiO2 implying a higher surface base current which decreases the current gain. In the reverse active region(figure 2b), the same phenomenon is observed [ref.6,8]. But, once again a large sensitivity can be seen only for low reverse voltages applied to the component. This influence is also visible in the middle of the curve as opposed to the previous case.

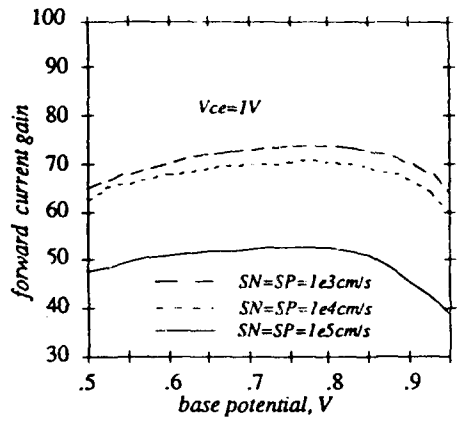


Figure 2a: Influence of the recombination velocities on forward current gain

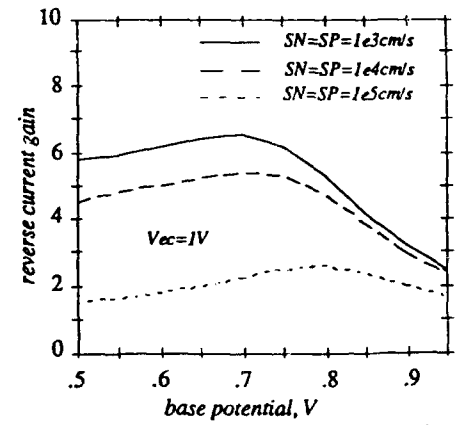


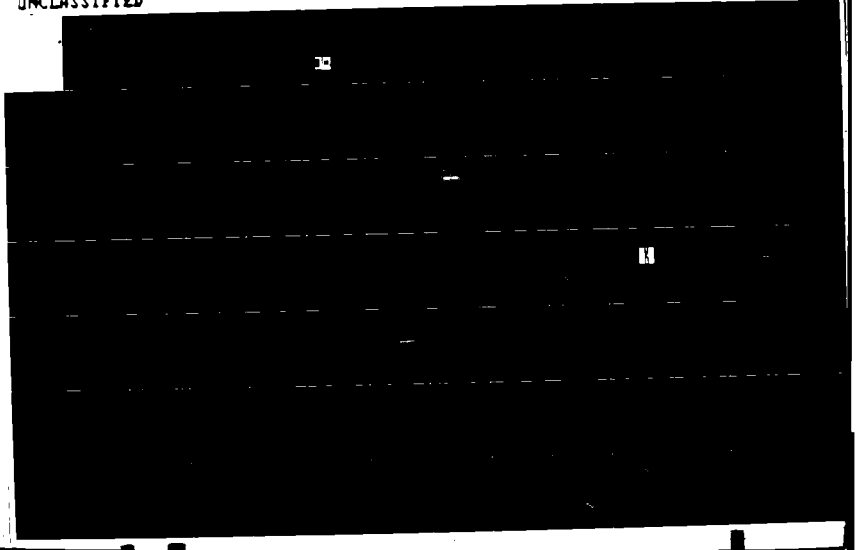
Figure 2b: Influence of the recombination velocities

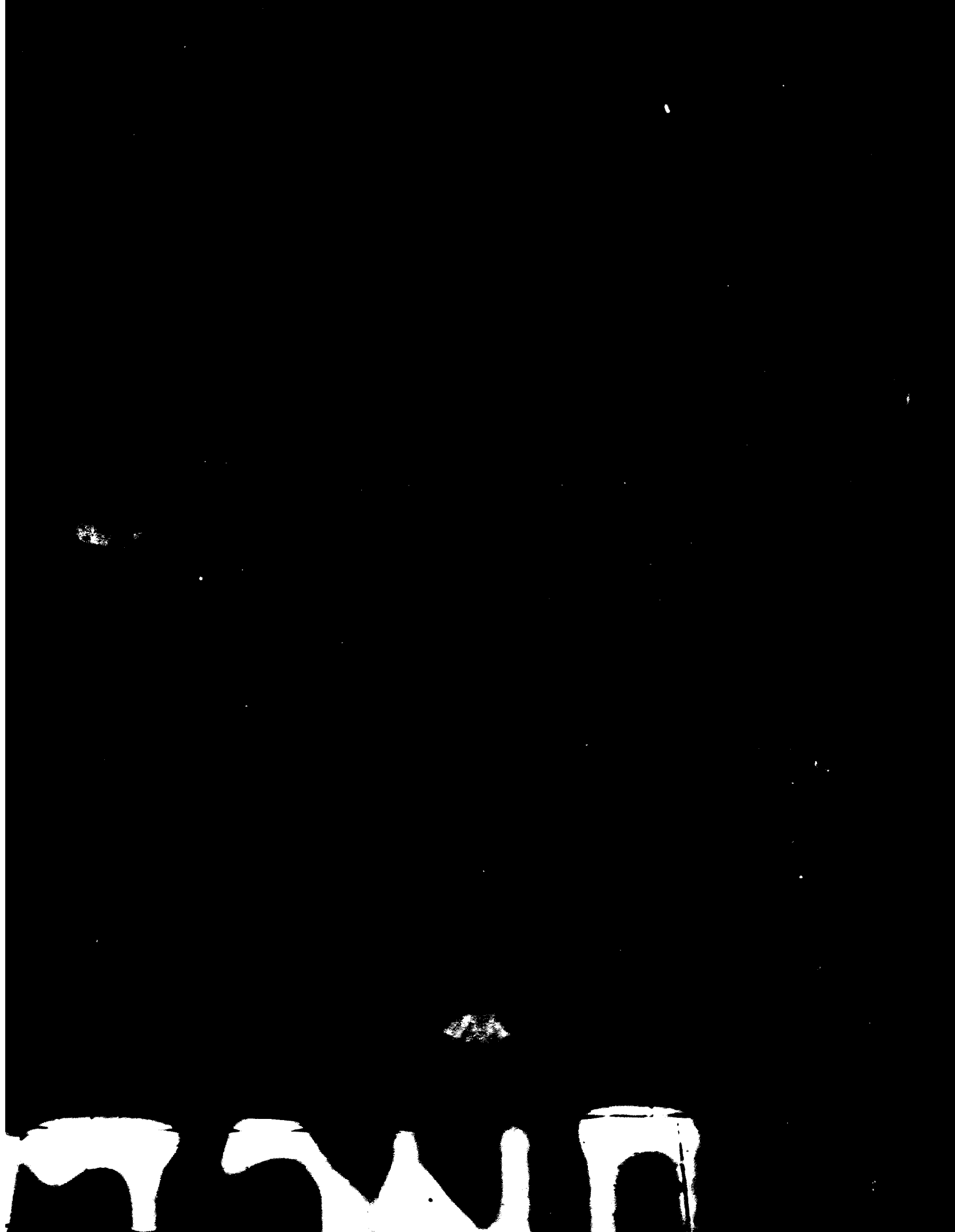
AD-A286 614 EUROPEAN SYMPOSIUM ON RELIABILITY OF ELECTRON DEVICES FAILURE PHYSICS AND..(U) GAREL ASSOCIATES NORTHAMPTON (UNITED KINGDOM) G M BRYDON 7 OCT 94 XB-ONREUR

2/

UNCLASSIFIED

ML





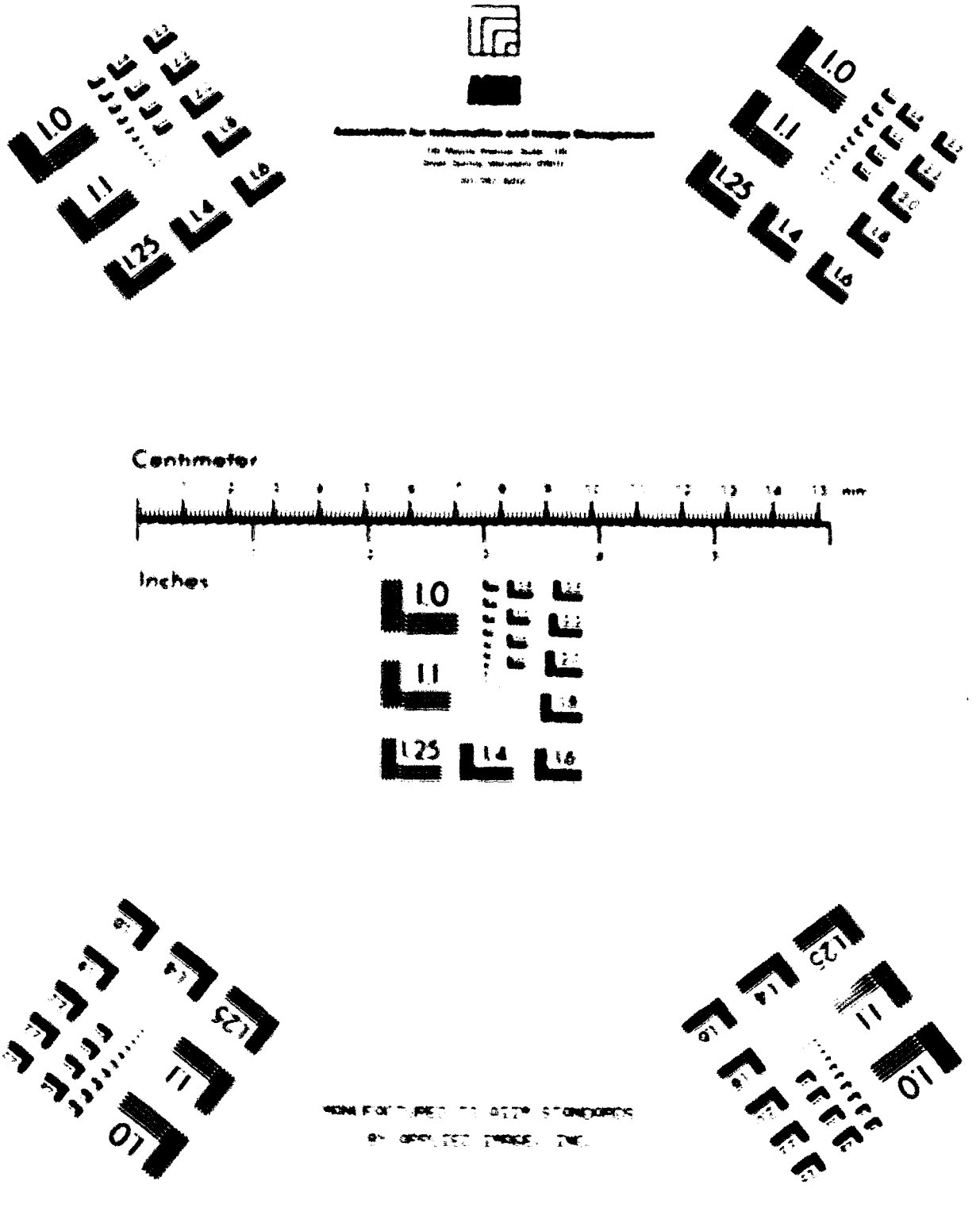

Centimeter
Inches
1.0
1.1
1.25
1.4
1.6

Figure 3 shows the influence of the lifetimes inside the material. It appears that the phenomenon may be observed whatever the polarizations is. This is due to the fact that these lifetimes directly act on the recombination current in the base which is one of the main components of the overall base current with the diffusion current due to the hole injected from the base to the emitter [refs.7,8]. Consequently this effect is quite different from the previous ones.

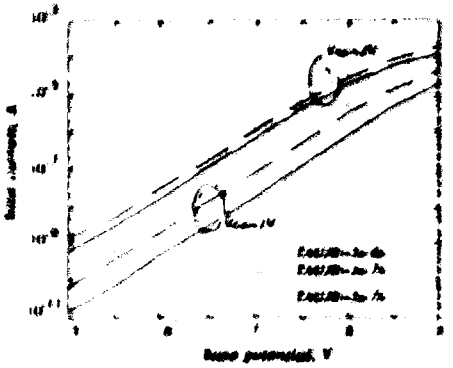


Fig 3: Influence of the lifetimes on base current

4. RESULTS AND CONCLUSIONS

In conclusion we are able to distinguish the effects of the GAN model parameters. Experiments have been conducted on several transistors which different geometries in order to demonstrate that obtaining the Gummel plots of one transistor is sufficient to predict the behaviour of other devices on the same die. Two types of transistors have been tested : one with the SUBILON technological process from PHILIPS and one with the BICMOS technological process from THOMSON. Figure 4a and 4b show excellent correlations between computed and experimental curves.

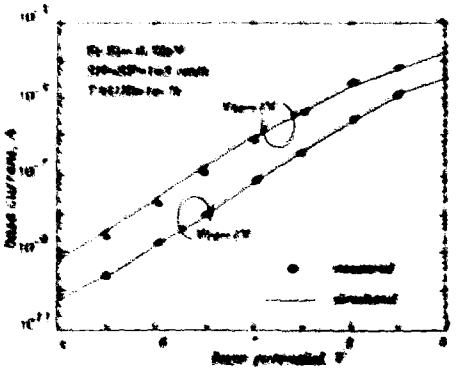


Figure 4a: Experimental and simulated results for SUBILON process

Classical optimization techniques minimizing the errors between computed and experimental curves lead to a set of parameters for the trap energy level and the surface state density. Lifetime parameters have been chosen according to the results of the authors[ref.3]. An Auger recombination process model is implemented in DAVINCI [ref.2], but modifying its parameters has not revealed a significant effect because of the low concentration of the dopant in the base. Different models for mobility have been tested but with no effect on the static characteristics: ARORA mobility model [ref.9] is used. We have taken into account the band gap narrowing effect [ref.10]. The obtained set of parameters has allowed the static process prediction for the other devices on the die. In the particular case of the BICMOS technological process from THOMSON, and in the case of a component operating inverted, the base current is very high. Indeed, in this configuration, the P+N junction which normally isolates the transistor is now not sufficiently reverse biased. So an additional current occurs in the base which overcomes the effect of the surface current and of the depletion region current. Consequently, determining Is, BN and SF will be done from the Gummel plots, when the transistor is biased in its forward mode, but with a loss of accuracy.

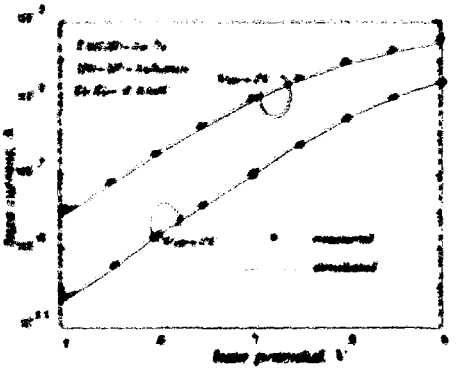


Figure 4b: Experimental and simulated results for BICMOS process

REFERENCES

1 M.A.Alsiddik and B.R.Jones "Correlation between trap characterization by low frequency noise , mutual conductance dispersion, oscillations and DLTS in GaAs MESFETs" Solid State Electronics Vol 35, 1992, pp 1713-1719

2 M.S Tyagi and R VAN Overstraeten "Minority carrier recombination in heavily-doped silicon" Solid State Electronics Vol 26, 1983, pp 577-597

3 D.J.Roulston, N.D.Arora, and S.G.Chamberlain "Modeling and measurement of minority-carrier lifetime versus doping in diffused layer of n+-p silicon diodes" IEEE Trans Electron Devices, vol.ED-29, Feb 1982, pp 284-291

4. S.M.Sze : "Physics of Semiconductor Devices", 2nd ed., John Wiley and Sons, New York, 1981, pp 56-57

5. C.T.sah, R.N.Noyce, and W.Shockley :" Carrier generation and recombination in p-n junctions and p-n junctions characteristics" Proc IRE, vol 45, Sept 1957, pp 1228-1243

6. C.T.sah :"Effect of surface recombination and channel on p-n junction and transistor characteristics" IRE Trans Electron Device , vol ED 9 Jan 1962, pp 94-108

7. D.J.Roulston:"Bipolar Semiconductor Devices" 1st ed., McGraw-Hill, New York, 1990

8. B.K.Jones and V.O.Kaya : "The non-ideal current in bipolar transistors" Solid State Electronics" Vol 30, 1987 , pp 987

9. N.D.Arora,J.R.Hauser, and D.J.Roulston, "Electron and Hole Mobilities in Silicon as a function of Concentration and Temperature" IEEE Trans. Electron Devices, vol. ED-29, Feb 1982, pp.292-295

10. J.W.Slotboom, "The pn-product in silicon,"Solid State Electronics 1977 vol 20 pp 279-283

ANALYSIS OF AVALANCHE BREAKDOWN OF n-CHANNEL MOSFETs USING SPECTROSCOPIC EMISSION MICROSCOPY

DSH Chan, LS Koh, WK Chim, JCH Phang

Centre for IC Failure Analysis and Reliability
Faculty of Engineering, National University of Singapore
10 Kent Ridge Crescent, Singapore 0511

ABSTRACT

The photon energy distributions from 1.5eV to 2.7eV of light emitted from nMOSFETs in avalanche breakdown were studied using an experimental spectroscopic emission microscope. The distributions were found to be non-Maxwellian under the high electric fields of breakdown and the Gaussian distribution function was found to give a better fit.

1. INTRODUCTION

Emission microscopy has been used to detect light emitted from stressed MOSFETs (Refs. 1,2,3,4). The spectral signature of light emission from nMOSFETs has been measured in an attempt to detect hot-carrier activity (Ref. 5) and to predict device degradation (Ref. 6). In an nMOSFET, the electric field near the corner of the Si-SiO_2 interface where the drain junction is directly under the gate is typically the highest. Carrier interaction with the field causes them to acquire kinetic energies larger than that corresponding to the ambient temperature. The presence of hot electrons can be monitored with the substrate current I_{sub}, gate current I_g, and light emission (Ref. 7). The emission spectrum due to hot-carriers can provide information on the electric field and electron energy distribution within a device.

2. EXPERIMENT

2.1 Apparatus

The experimental set-up is shown in Figure 1 and has three main components. These are the light collection optics, light dispersion optics and detector. The light collection optics consists of a MicroZoom Probe reflected light microscope with long working distance objectives and optical fibre light-guide. As the emitted light is of very low intensity and the area of emission extremely small, the highest power objective of 50X with the largest numerical aperture (NA) of 0.45, and 2X zoom in the probe body was used to maximize the collection efficiency. An adapter focuses the light such that it emerges from the trinocular as a converging beam so that it can be launched into the light guide at well within the NA of the light guide for the highest launching efficiency. The light guide is made of fused silica. The light dispersion optics and detector is the Optical Spectrum Analyzer from Monolight Instruments. The Optical Spectrum Analyzer is made up of a monochromator, photomultiplier unit and system controller. The monochromator is the Ebert-Fastie type with a continuously rotating grating. It has an aperture ratio (f#) of f/7, a nominal focal length of 75mm and a grating with a blaze wavelength of 500nm. The photomultiplier unit has a built in high tension power supply and a pre-amplifier. The photomultiplier tube is the R928 side-on type from Hamamatsu. The system controller controls the monochromator scanning speed, the signal gain and wavelength sampling interval. It also interfaces the system to the computer.

Calibration of the system is necessary as each optical component has a wavelength dependent transmission characteristic and the photomultiplier has a spectral responsivity. The set-up was calibrated by using the system to capture a spectrum $R_{STD}(\lambda)$ of a tungsten lamp of known spectral content $M_{STD}(\lambda)$. The correction factor $C(\lambda) = M_{STD}(\lambda)/R_{STD}(\lambda)$ can then be calculated. Since the tungsten lamp is an incandescent light source which produces a continuous spectrum from 400nm to beyond 1100nm at 2900K, the grating based monochromator will produce second-order spectra. Second order blocking filters are necessary for calibration which was performed over two ranges of wavelengths: 420nm to

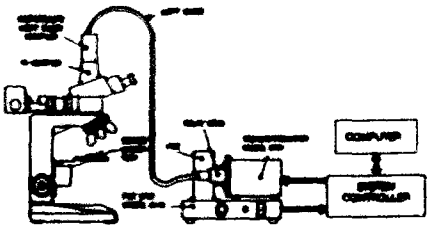


Figure 1 The spectral analysis system. The microscope and the PMT is placed in a light proof box.

620nm and with the second-order blocking filter from 620nm to 860nm. The lower wavelength limit is due to the optics which cannot transmit in the ultra-violet and the upper wavelength limit is due to the spectral responsivity of the photomultiplier tube. As there are two sets of correction factors resulting from the calibration, weighted spectral measurement was done over the first range without the second-order blocking filter and the second range with the second-order blocking filter. The two sets of spectral measurements were then merged at 620nm.

The device under test is placed under the microscope and biased using the HP 4145B Parameter Analyzer. The device under test, microscope, light-guide and photomultiplier unit were enclosed in a light proof black box to prevent ambient light from affecting the spectrum.

2.2 Device Description and Biasing

The devices used in this study were polysilicon-gate n-channel double diffused drain (DDD) MOSFETs test structures fabricated using 1.5μm technology on 6.0-9.0 Ωcm epitaxial layer on p-type substrate. The mask gate lengths were *1.0μm, 1.6μm and 2.0μm* and gate width was 50μm. The source-drain junction depth = 0.35μm, gate oxide thickness = 25nm and poly-Si-gate thickness = 0.4μm. The devices on the test structure have a common source and gate. Typical threshold voltages range from 0.65V to 0.76V at V_{ds} = 50mV and I_d = 0.1μA/μm of device width.

The three devices were biased into the snapback region at constant drain current of 10mA, 20mA and 30mA. V_g was varied from 1V in steps of 1V until the emission was too faint to be detected. The source and substrate were grounded. The I-V characteristic, threshold voltage V_t, and snapback characteristic were measured and recorded using the Parameter Analyzer before and after the experiment. The emission spectrum for each bias point was captured and the corresponding V_{ds}, V_g, I_{sub} and source current I_s were recorded. Figure 2 shows the emission image superimposed on the reflected image of a 50μm/1.6μm device in avalanche breakdown at I_d = 20mA and V_g = 3V. The measured V_d at this bias point is approximately 8.5V.

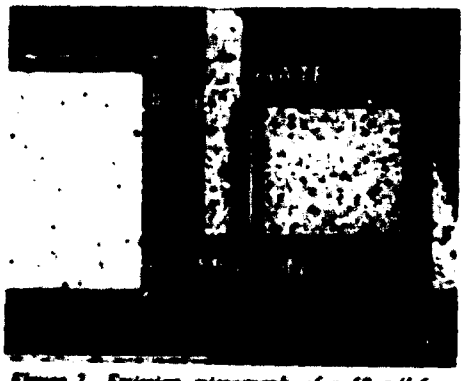


Figure 2. Emission micrograph of a 50μm/1.6μm nMOSFET in avalanche breakdown with I_d=20mA, V_g=3V and a measured V_d of 8.5V

2.3 Correction for Absorption in Polysilicon

The technical intensity spectrum obtained from the Optical Spectrum Analyzer has to be corrected for absorption. The emitted light from the channel of the nMOSFET passes through the gate oxide, polysilicon gate, BPSG passivation and Si_3N_4 passivation. SiO_2, BPSG and Si_3N_4 have large band-gap energies relative to the photon energies measured and thus there is very little absorption of light in these films. The polysilicon band-gap energy of 1.12eV is less than the photon energy range of 1.4eV to 2.7eV and absorption has to be corrected for in the emission intensity spectrum. The corrected intensity $I_o(\lambda)$ for absorption through the gate is given by

$$I_o(\lambda) = \frac{I(\lambda)}{\exp(-\alpha t_{poly})} \qquad (1)$$

where $I(\lambda)$ is the measured intensity at a particular wavelength λ, α is the absorption coefficient (Ref. 8) and t_{poly} is the thickness of the polysilicon gate.

2.4 Photon Energy Distribution

The wavelength of the emitted light is converted to its equivalent photon energy E using Planck's equation

$$E = \frac{hc}{\lambda} \qquad (2)$$

where h is Planck constant,
c is the speed of light and
λ is the wavelength.

Since a grating monochromator measures the light intensity per unit wavelength, $I_\lambda(\nu)$, the following conversion was carried out to obtain photons per unit energy $I_o(\nu)$ (Ref. 4)

$$I_o(\nu) = \frac{c}{\nu^2} I_\lambda\left(\frac{c}{\lambda}\right) \qquad (3)$$

where ν is the frequency.

3. PHYSICAL MECHANISM FOR MOSFET BREAKDOWN

The explanation of the physical mechanism of avalanche breakdown in nMOSFETs follows that of Hsu *et al* (Ref. 9). When an n-channel MOSFET is operated in the saturation region, the electric field near the drain junction can be large enough to cause impact ionization (1) as shown in Figure 3. The generated electrons are swept into the drain, while the holes move into the substrate. As the holes drift in the substrate towards the source (2), an ohmic drop develops that tends to forward bias the source/substrate junction. This junction then injects electrons into the base region of the parasitic lateral n-p-n transistor (3). Most of these injected electrons are then collected by the drain through the depletion region under the channel and at the drain/substrate junction (4). Those electrons collected through the drain region increase the negative-charge density in the channel near the drain region, and as a consequence also increases the electric field. This higher electric field raises the avalanche multiplication factor above its value without electron injection. Therefore, a lower drain-to-source bias is sufficient to sustain the same drain current level. This accounts for the observed negative resistance after breakdown.

The peak electric field E_m, determined from an analytical model (Ref. 9) to predict the drain current-voltage characteristic up to the current-controlled negative resistance region, is given by

$$E_m = \sqrt{A^3(V_d - V_{dsat})\left(V_d - V_{dsat} + \frac{2\Delta Q_m}{C_{ox}}\right) + E_{sat}^2} \quad (4)$$

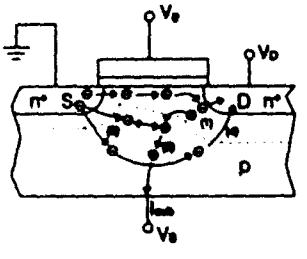


Figure 3 Schematic diagram illustrating the current paths in a nMOSFET.

where V_d is the drain voltage, V_{dsat} is the saturation voltage and E_{sat} is the electric field at which electron velocity saturation occurs. The parameter A is given by

$$A = \sqrt{\frac{\epsilon_{ox}}{\epsilon_{si}} \frac{1}{X_{ox}[X_j + 0.01(V_d - V_{dsat})(X_1 - X_j)]}} \quad (5)$$

where X_j is the junction depth, X_{ox} is the oxide thickness, ϵ_{ox} is the oxide permittivity and ϵ_{si} is the silicon permittivity. The factor 0.01 has units of V^{-1}. X_1 is given by

$$X_1 = \sqrt{\frac{2\epsilon_{si}(V_{dsat} + 2\phi_F)}{qN_{sub}}} \quad (6)$$

where q is the electronic charge and N_{sub} is the substrate doping. ϕ_F is the equilibrium electrostatic potential given by

$$\phi_F = \frac{k_b T}{q} \ln\left(\frac{N_{sub}}{n_i}\right) \quad (7)$$

where k_b is the Boltzmann constant, T is the lattice temperature and n_i is the intrinsic carrier concentration of silicon. In breakdown, the source injects electrons, represented by I_e of which a fraction $(k\gamma\alpha_T I_e)$ is collected by the drain high-field region, causing additional negative mobile charge ΔQ_m to be added to this region. ΔQ_m is given by

$$\Delta Q_m = \frac{k\gamma\alpha_T I_e}{W v_{sat}} \quad (8)$$

where γ is the injection efficiency of the source/substrate junction, α_T is the base transport factor, k is the fraction of the collected electrons that go through the drain high-field region, W is the gate width and v_{sat} is the electron saturation velocity. C_{ox} is the oxide capacitance per unit area given by

$$C_{ox} = \frac{\epsilon_{ox}}{X_{ox}} \quad (9)$$

While the calculations by Hsu *et al* (Ref. 9) shows a smooth change in the I-V characteristics, their actual experimental device characteristics show discontinuities in drain voltages especially for longer channel devices. The analytical expression seems unable to model these discontinuities.

4. RESULTS

A typical semi-log plot of the photon energy spectrum shown in Figure 4. Periodic modulation of period 0.3eV to 0.4eV due to interference effects in the polysilicon thin film as reported in (Refs. 10,11) was observed. In spite of this modulation, for the photon energy range of 1.5eV to 2.5eV, the photon energy distribution decreases fairly linearly and analysis using a temperature model for the electron energy distribution (Ref. 12) is possible. The carrier temperature T_c can be extracted from the slope of the photon energy distribution semi-log plot. The range of carrier temperatures obtained as shown in Figure 5 are higher than that for saturated nMOSFETs (Ref. 13) but the trend for changes in bias is similar. The intensity was found to be proportional to the nth power of the substrate current where $n \approx 0.88$ as shown in Figure 6.

Over an extended energy range of 1.5eV to 2.7eV, it was observed that at some bias the photon energy distribution was non-Maxwellian and can be better represented by a Gaussian distribution function (Ref. 10) $f(E) = N\ exp(-AE^2)$ as shown in Figure 4 or a hybrid of both (Ref. 14) $f(E) = N'\ exp(-A'E-B'E^2)$, where N and N' are normalisation factors, A, A' and B' are proportionality constants. The peak electric field in a nMOSFET under breakdown was calculated using equation (4) and it was observed that the Gaussian distribution function provided the better fit when the electric field was greater than $4x10^5$V/cm as shown in Figure 7. The hybrid distribution function generally provided the better fit after snapback in the I-V curve occurs.

5. CONCLUSIONS

Spectral analysis of light emitted from n-channel MOSFETs biased into avalanche breakdown was performed over a photon energy range of 1.5eV to 2.7eV using an experimental set-up for spectroscopic emission microscopy described in this paper. The photon energy distributions obtained for different device channel lengths and biases were studied using fits to Maxwellian, Gaussian and the hybrid of both distributions. The photon energy distribution was generally found to be non-Maxwellian under the high electric fields of avalanche breakdown and the Gaussian distribution function was found to give a better fit. The observations in this work raises the possibility of using the photon energy distribution to study the electron energy distribution within a nMOSFET .

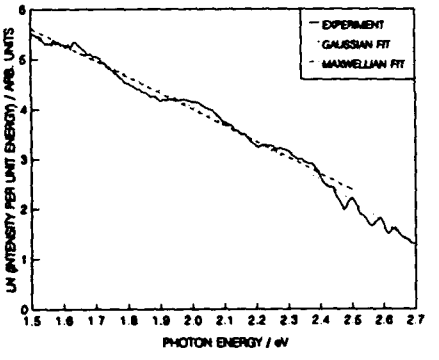


Figure 4. A typical semi-log plot of the photon energy distribution of light emitted from a 50μm/1.6μm device with I_d=30mA, V_g=3V and V_d=8.45V. The PED has been fitted with two distribution functions:
(a) Maxwellian (- - -) from 1.5eV to 2.5eV,
(b) Gaussian (.....) from 1.5eV to 2.7eV.

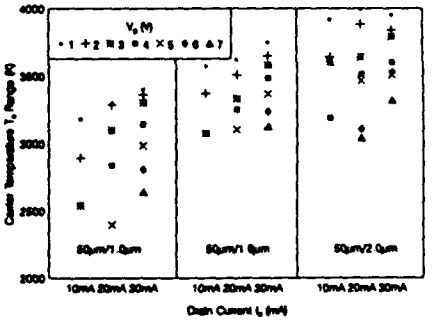


Figure 5. Carrier temperature T_c range for 1.0μm, 1.6μm, 2.0μm channel length devices at various breakdown current I_d. V_g=1V to 4V for I_d=10mA due to system sensitivity limitations, thus T_c range appears narrower, and the lowest value is greater than the corresponding T_c for I_d=20mA.

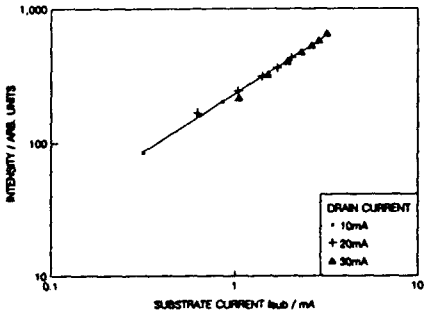


Figure 6. Log-log plot of intensity against substrate current for a 50μm/1.0μm device for all bias points.

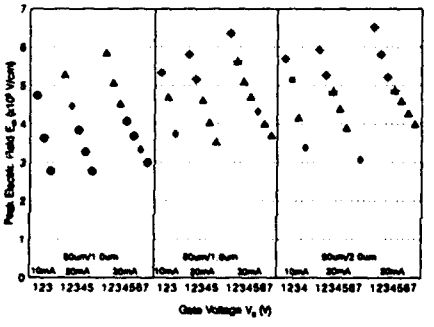


Figure 7. A plot of calculated peak electric field strength against bias points. The best fit distribution function has been assigned at each bias point. The symbols ●, ▲ and ■ represent the Maxwellian, Gaussian and hybrid distribution functions respectively. + indicates snapback in the I-V characteristics. The symbol ◆ indicate situations where the sum of the squared errors for different distribution functions are nearly equal and the best fit function could be either be between Maxwellian and Gaussian or between Gaussian and hybrid.

REFERENCES

1. Shade G.F. "Physical Mechanisms for Light Emission Microscopy", ISTFA, 1990, pp. 121-128.

2. Khurana N. "Second Generation Emission Microscopy and its Applications", ISTFA, 1989, pp. 277-283.

3. Khurana N. and Chang C.L. "Analysis of Product Hot Electron Problems by Gated Emission Microscopy", IEEE IRPS, 1986, pp. 189-194.

4. Tam S., Hsu S.C., Ko P.K. and Muller R.S. "Spatially Resolved Observation of Visible-Light Emission from Si MOSFETs", IEEE Electron Dev. Lett., Vol. EDL-4, No. 10, October 1983, pp. 386-388.

5. Ishizuka H., Tanaka M., Konishi H. and Ishida H. "Advanced Method of Failure Analysis Using Photon Spectrum of Emission Microscopy", ISTFA, 1990, pp. 13-19.

6. Tsutsu N., Uraoka Y., Nakata Y., Akiyama S. and Esaki H. "New Detection Method of Hot-Carrier Degradation Using Photon Spectrum Analysis of Weak Luminescence on CMOS VLSI", Proc. IEEE 1990 Int. Conference on Microelectronic Test Structures, Vol. 3, March 1990, pp. 143-148.

7. Hu C. "Hot Carrier Effects in VLSI Electronic Microstructure Science Vol. 18, Advanced MOS Device Physics", N.G. Einspruch, G.S. Gildenblat, Ed., Academic Press, Inc., 1989, pp. 119-160.

8. Lubberts G., Burkey B.C., Moser F. and Trabka E.A. "Optical Properties of Phosphorus-doped Polycrystalline Silicon Layers", J. Appl. Phys. 52(11), November 1981, pp. 6870-6878.

9. Hsu F.C., Muller R.S. and Hu C. "A Simplified Model of Short-Channel MOSFET Characteristics in the Breakdown Mode", IEEE Trans. Electron Devs., Vol. ED-30, No. 6, June 1983, pp. 571-576.

10. Sekido M., Taniguchi K. and Hamaguchi C. "Direct Observation of Gaussian-Type Energy Distribution for Hot Electrons in Silicon", Japanese Journal of Applied Physics, Vol. 30, No. 6, June 1991, pp. 1149-1153.

11. Hublitz K. and Lyon S.A. "Light Emission from Hot Carriers in Si MOSFETs", Semicond. Sci. Technol. 7, 1992, pp. B567-569.

12. Tam S., Hsu F.C., Ko P.K.and Muller R.S. "Spatially Resolved Observation of Visible-Light Emission from Si MOSFETs", IEEE Electron Device Letters, Vol. Edl-4, No. 10, October 1983, pp. 386-388.

13. Kölzer J., Boit C., Dallmann A., Deboy G., Otto J. and Weinmann D. "Quantitative Emission Microscopy", J. Appl. Phys. 71 (11), 1 June 1992, pp. R23-R41.

14. Bude J., Sano N. and Yoshii A. "Hot-carrier Luminescence in Si", Physical Review B, Vol. 45, No. 11, The American Physical Society, 15 March 1992, pp. 5848-5856.

UNDERSTANDING OF THE HOT-CARRIER DEGRADATION IN SUBMICRON MOSFET'S: FROM UNIFORM INJECTION TOWARDS THE REAL OPERATING CONDITIONS

G. Groeseneken, R. Bellens, G. Van den bosch, H.E. Maes
IMEC, Kapeldreef 75, 3001 Leuven, Belgium

Abstract

An overview is given of the present understanding of the hot carrier degradation problem in submicron MOSFET's. First, the degradation mechanisms, observed under for circuit operation somewhat artificial but well-controlled uniform substrate hot electron and substrate hot hole injection conditions are discussed. Then the more realistic case of static channel hot carrier degradation is treated, and some important process related effects are illustrated, followed by the behaviour under the - for real operation - most relevant case of dynamic degradation. Finally, the strategies for improving hot carrier reliability and the forecast of the hot carrier reliability problem for sub-0.25 μm technologies are briefly discussed.

1. INTRODUCTION

One of the main problems encountered when downscaling device geometries into the deep submicron range has been the hot carrier induced degradation of MOSFET's. This problem has been studied intensively during the past decade, under both static and dynamic stress conditions. In this period, it has evolved from a more or less academic research topic to one of the most stringent constraints guaranteeing the lifetime of submicron devices.

Under the influence of the high lateral fields in short channel MOSFET's, electrons and holes in the channel and pinch off regions can gain sufficient energy to surmount the energy barriers or tunnel into the oxide. This leads to the generation of traps, both at the interface and in the oxide, and to electron and hole trapping in the oxide, which will cause changes in transconductance, threshold voltage and drive current of the MOSFET.

In this paper, an overview of our present understanding of the hot carrier degradation problem is given. The mechanisms that are at the base of the hot carrier degradation are hot hole and hot electron injection. Both injection mechanisms lead to different degradation phenomena. In order to understand the degradation mechanisms under real operating conditions, it is mandatory to first understand the degradation for each type of injected carrier separately. Ideal tools to study these phenomena are the uniform substrate hot hole and hot electron injection techniques. In these techniques, the MOSFET's are stressed under artificial conditions, but the field and fluence conditions can be accurately controlled. These techniques were used to study the degradation of devices at both room temperature and at 77K in order to get better insight in the interface degradation mechanisms under both hot hole and hot electron injection. This is the subject of the first part of the paper.

In the second part of the paper, the degradation of MOSFET's under the more realistic conditions of channel hot carrier injection is reviewed. The different degradation mechanisms under static stress conditions are summarized, lifetime determination methods are briefly described and some important process related factors and effects will be highlighted.

Going to even more realistic conditions, the degradation under dynamic stress conditions is described, emphasizing the influence of the measurement setup, the comparison with static stress and the importance of post-stress and detrapping effects.

Finally, in the third part of the paper, some strategies that are under investigation for improving the hot carrier reliability for deep submicron transistors will be briefly discussed. These strategies are concentrated on the use of improved dielectrics on the one hand, and on alternative drain engineering techniques on the other.

2. DEGRADATION OF MOSFETS UNDER UNIFORM HOT CARRIER INJECTION

In this section, an overview is given of the effects of electron and hole injection – trapping, oxide and interface trap generation – as obtained from uniform substrate hot-carrier injection experiments. The efficiency of these processes, as well as their dependencies on oxide and silicon field, temperature and oxide thickness will be discussed. Some indicative values have been summarized in Table 1.

2.1 Effects of electron injection

Electrons injected into the gate oxide have a certain, though small probability (on average about 10^{-7}) of becoming trapped in electron traps present in the oxide bulk. Low-field electron trapping has been found to be independent of the silicon field (i.e. the energy at injection) except in oxides too thin for the injected hot electrons to be thermalized. The trapping is, however, sensitive to the injected current density (Refs. 1, 2). The saturated value has a maximum at low field and is determined by the steady-state balance of trapping and detrapping (Ref. 3).

At 77 K electron trapping is one to two orders of magnitude more effective due to additional trapping in energetically shallow traps that cannot be permanently filled at room temperature due to a very high thermal

	ELECTRONS		HOLES	
	295K	77K	295K	77K
trapping	10^{-7}	10^{-6}-10^{-5}	0.1-0.2	0.5-0.8
$\|E_{ox}\|$ dep.	↓	↓↓	↓	↓
trap generation	10^{-7}-10^{-6}	10^{-7}	0	0
$\|E_{ox}\|$ dep.	↑↑	↑↑		
D_{it} generation	10^{-9}-10^{-7}	10^{-11}-10^{-9} ↑↑	10^{-3}-10^{-2}	4.10^{-3}
|Eox| dep.	↑↑		↓	↓

Table 1 - Efficiency and oxide field dependence of degradation processes related to hot-electron and hot-hole injection. The numbers indicate trapping, oxide trap and interface trap generation efficiency, which is defined as the number of trapped charges or generated traps per injected carrier.

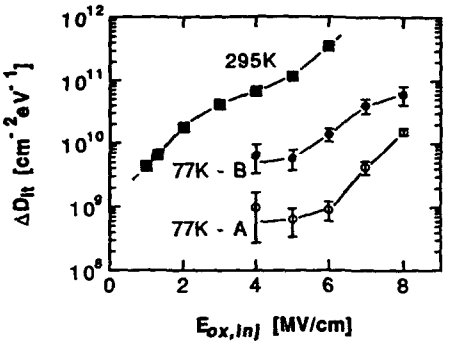


Fig. 1 - Total interface trap generation after injection of $6.25x10^{18}$ cm^{-2} electrons as a function of oxide field, at 295 and 77 K. Curves A and B are after injection at 77 K and after warmup to 295 K, respectively. (Ref. 9)

emission rate. Field detrapping from shallow traps is strongly promoted at higher oxide field (Ref. 3).

Electron trapping is less important in thinner oxides, since both the total number of traps decreases as well as their effect on the MOSFET I-V characteristics.

At higher oxide fields electron trapping no longer saturates at high fluence while the trapping rate increases with oxide field. This is ascribed to the creation (and filling) of additional electron traps by injected electrons that acquired a high energy in the oxide field. The existence of so-called "threshold" fields of enhanced electron trap generation at 1.5 MV/cm (the onset of electron heating in the oxide field) (Ref. 1) or 4 MV/cm (the introduction of holes into the oxide) (Ref. 4) have been reported, while other studies only observed a gradual increase in trapping without any clear threshold (Ref. 2). It has been remarked that the exact manner in which the trapping data are examined could lead to different conclusions, however (Ref. 5). The minimum energy to create a trap has been reported to be 2.3 eV (Ref. 1). The existence of a (field or energy) threshold and the field activation of electron trapping are important issues in the downscaling of devices and in accelerated lifetime tests (Refs. 1, 6).

As the trap occupancy decreases with E_{ox}, a consistent picture of trap generation phenomena can only be obtained after the application of an additional – short – "trap-filling" step after the actual high-field injection to effectively occupy all the possibly generated traps (Refs. 7, 2).

Electron trap generation is temperature activated: at 77 K this process is reduced (Ref. 8), and according to others even nearly inhibited (Ref. 3). The importance of taking into account the trap occupancy in revealing the occurrence of trap generation at 77 K has been demonstrated (Ref. 9).

The most important features of electron-induced interface trap generation are summarized in Fig. 1. This plot compares the interface trap generation ΔD_{it} (at a fixed fluence) at two injection temperatures 295 K and 77 K, as a function of oxide field (Ref. 9). At 295 K, the interface trap generation rate is seen to exhibit an approximately exponential field dependence over a wide oxide field range. This strong field dependence suggests a correlation with electron heating in the oxide field (Ref. 1). This is further supported by the observation that interface trap generation is independent of the injection conditions in the silicon (Ref. 2).

At 77 K, curve A is the direct generation during injection, while curve B is the total generation after injection and subsequent warmup to 295 K (at zero bias). As for the case of room temperature, both curve A and B show a strong field enhancement. Hot-electron induced interface trap generation is clearly a strongly temperature-activated process: going from 295 K to 77 K, the generation efficiency drops by more than two orders of magnitude. Curve B shows that during injection at 77 K also some latent damage is formed. However, the fact that curve B still lies well below the 295-K injection curve proves that at low temperature the main interface trap producing mechanism is truly suppressed and not just temporarily inhibited. The small remaining generation is believed to be due to other, less important processes (Ref. 9).

The approximately exponential field dependence of D_{it} generation observed in Fig. 1 –when plotted on a log scale– does not suggest the existence of threshold fields for enhanced trap generation, in agreement with other work (Ref. 2). Earlier studies, however, have reported threshold fields equal to those observed for electron trap generation (Refs. 1, 4), as discussed above. The fact that interface traps are formed at any oxide field is not inconsistent, however, with the existence of a threshold *energy* of trap generation.

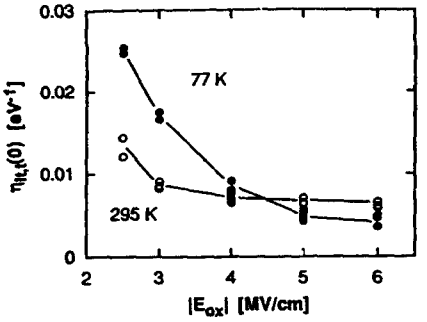


Fig. 2 - Initial interface trap generation efficiency per trapped hole versus oxide field, at 295 and 77 K. (Ref. 12)

2.2 Effects of hole injection

The trapping of holes injected into the gate oxide occurs with several orders of magnitude higher efficiency than that of electrons: the trapping efficiency reaches values as high as 10-20% (Refs. 10, 11, 12). Hole trapping is quite insensitive to the oxide field conditions. With increasing oxide field magnitude, a slowly decreasing trapping efficiency is observed, that has been attributed to a hole capture cross section that is linearly decreasing with $|E_{ox}|$ (Ref. 10). No dependence on silicon field nor injected current density has been observed, an observation that is consistent with the assumed fast thermalization of injected holes in the oxide (Ref. 10, 12).

At 77 K hole trapping is enhanced, the efficiency increasing to values up to 70% (Refs. 12, 13). Analysis of the trapping kinetics at 295 and 77 K has demonstrated that the enhanced trapping at 77 K is not due to additional trapping in shallow hole traps, but rather to a larger effective capture cross section at the latter temperature (Ref. 12).

At large injected hole density a clear saturation level of the trapping is observed, that is independent of the oxide field during injection (Ref. 5), *suggesting the absence of additional hole trap generation as a result of hole injection*, in contrast to the case of trap generation by electron injection.

Hole traps are mainly located near the Si-SiO_2 interface (Refs. 11, 14) so that their effect on the current-voltage characteristics of the transistor depends approximately linearly on oxide thickness. Furthermore, an essentially trapped-hole free layer exists at either interface due to detrapping of trapped holes by tunneling processes. As a result, gate oxides of thickness smaller than two times this charge-free layer no longer show significant permanent hole trapping. The critical thickness for which hole trapping disappears has been reported to be 4 nm (Ref. 14) or even as much as 7 nm (Ref. 11). This effect is of particular importance for the further downscaling of devices.

In contrast with the trap generation, the *interface trap generation is much more efficient by hole injection than by electron injection*. Typical generation efficiencies are in the range of 10^{-3}-10^{-2} (Refs. 10, 12, 15). As is also the case for hole trapping, the generation process was observed to be independent of silicon field (Ref. 12). A second feature common to hole trapping is the weak oxide field dependence of this interface trap generation process: the D_{it} generation efficiency slowly decreases with $|E_{ox}|$ (Ref. 12, 15, 16), again in contrast to the exponential oxide field dependence of interface trap generation by electron injection.

At 77 K the interface trap generation efficiency is also higher than at 295 K, clearly in strong contrast to electron-induced Dit-generation, as discussed previously, and also to radiation-induced interface trap generation (Ref. 12).

The features of hole trapping and hole-induced interface trap generation are strikingly similar, suggesting a causal relation between these two processes. Fig. 2 plots the interface trap generation efficiency *per trapped hole* at 295 and 77 K, revealing that this quantity (and therefore the generation process) is nearly temperature-independent (Ref. 12). In Fig. 2, about 1% of the trapped holes give rise to interface traps, but also higher values have been reported (Refs. 6, 15). Further evidence for a direct correlation between hole trapping and interface trap generation is provided by the oxide thickness dependence of the generation process. A linear relation between ΔN_{ot} and ΔD_{it} has been reported independent of oxide thickness, while D_{it} generation increases with oxide thickness for small t_{ox} but rapidly saturates for larger t_{ox}, consistent with the spatial distribution of trapped holes in the oxide (Ref. 15)

Apart from the *direct* generation of interface traps during hole injection discussed above, also *delayed* interface trap formation occurs after termination of injection (Refs. 12, 16). This important delayed component is believed to be a temperature activated process similar to that observed in irradiated devices (Ref. 12).

3. DEGRADATION OF MOSFETS UNDER REAL OPERATING CONDITIONS

3.1 Static degradation

Once the different types of damage that are created under influence of both electron and hole injection are determined, and the dependencies of their generation efficiencies on electrical fields are known, one can tackle the more realistic case of the channel hot carrier degradation under static conditions. First of all it is important to understand the different degradation mechanisms that are playing a role under various stress conditions and for different transistor types. Next, a method for the lifetime determination and extrapolation towards normal operating conditions has to be established. Finally, a number of processing related factors that can influence the degradation and the hot carrier lifetime will be discussed.

3.1.1 Degradation mechanisms

The hot carrier degradation mechanisms under static conditions have been extensively studied during the last decade, and there exists more or less a consensus on a consistent picture of the degradation mechanisms for both n-type and p-type MOSFET's. A full description falls beyond the scope of this paper, but can be found in (Refs. 6, 17, 18). Here we will only summarize the sign of the current change and the type of damage that is generated for three types of MOSFET's (conventional n-MOS, LDD n-MOS and conventional p-MOS), and for three different gate voltage ranges: low (V_g=V_t), medium (V_g=V_d/2) and high (V_g=V_d). This is shown in Table 2. The damage that is determining for the current change is indicated in italic.

V_g-range	n-MOS	nMOS LDD	p-MOS
low V_g	I_d ↑ *h-trapping* D_{it}-creation	I_d ↑ *h-trapping* D_{it}-creation	$\|I_d\|$ ↑ *e- trapping* D_{it}-creation
medium V_g	I_d ↓ *D_{it}-creation*	I_d ↓ *e-trapping* Dit-creation	$\|I_d\|$ ↑ *e-trapping* D_{it} creation
high V_g	I_d ↓ *e-trapping*	I_d ↓ *e-trapping*	$\|I_d\|$ ↓ *D_{it} creation*

Table 2 - Degradation mechanisms for three types of MOSFET's and for three stress gate voltage ranges.

For conventional n-MOS devices, the maximum degradation occurs in the medium voltage range, and is caused by mobility degradation due to interface trap generation. These interface traps are generated by the simultaneous injection of holes and electrons for these conditions, but it was shown in the previous section that the holes are more effective by about four orders of magnitude in generating interface traps. At low gate voltages, the current is increasing due to the efficient trapping of holes, leading to a channel shortening effect. For these conditions, interface traps are created as well, but their influence on the current is masked by the trapped holes. Their influence might become visible, however, after neutralization of these holes.

The main difference with LDD n-MOS devices is that for this type of MOSFET the maximum current degradation is dominated by a series resistance increase of the n⁻-region, due to electron trapping in the spacer above this region.

The change of the drain current in p-MOSFET's is analoguous to the hole trapping case in n-MOSFET's. Electron injection and trapping increases the (absolute value of the) drain current by channel shortening, and masks any effect of the generated interface traps. Only, this electron trapping occurs for almost all stress gate voltages, because the electrons are becoming hotter than holes in the same electrical field and because the energy barrier for injection of electrons is smaller than for holes, which leads to efficient electron injection over the complete gate voltage range.

3.1.2 Lifetime determination

Like for all reliability failure mechanisms, it is important to be able to predict the lifetime of a component or device under operating conditions. Such lifetime methods are always based on experiments, in which the failure mechanism is accelerated, either by temperature or by high voltages or high currents. During the experiment, a degradation parameter, relevant for the damage, is monitored, and the lifetime is defined as the time to reach a certain shift in this degradation monitor.

For the hot carrier degradation mechanism, several accelerated lifetime determination methods have been proposed in the past. They are, however, all based on acceleration of the degradation by increased (drain) voltages, since hot carrier degradation is one of the few mechanisms that is not accelerated by an increase of the temperature (Ref. 19).

Most accelerated lifetime determination methods are based on the lucky electron model. The damage of the device, in terms of generated interface traps ΔN_{it}, can be related to the electrical field E_m at the drain (Ref. 20):

$$\Delta N_{it}(t) = C_1 \left[\frac{I_d}{W} t \exp\left(- \frac{\phi_{it,e}}{q\lambda_e E_m} \right) \right]^n \qquad (1)$$

where W is the width of the device, $\phi_{it,e}$ is the energy an electron must possess in order to create an interface trap and λ_e is the hot electron mean free path. A measure for the electrical field E_m is the multiplication factor $M=I_{sub}/I_d$, given by:

$$M = \frac{I_{sub}}{I_d} = C_2 \exp\left(- \frac{\phi_i}{q\lambda_e E_m} \right) \qquad (2)$$

where ϕ_i is the impact ionization energy. By defining the lifetime as the time τ to reach a fixed amount of damage, and by combining (1) and (2), the lifetime can be written as:

$$\frac{\tau I_d}{W} = C_3 \left(\frac{I_{sub}}{I_d} \right)^{-\frac{\phi_{it,e}}{\phi_i}} = C_3 M^{-m} \qquad (3)$$

If one plots the lifetime as a function of the multiplication factor on a double logarithmic scale, according to (3) one obtains a straight line with slope m, where m is an indication of the energy of the electrons that are causing the damage.

In a simplified form of (1) the lifetime is related to the drain voltage, and thus the power supply voltage as

$$\tau = C_4 \exp\left(\frac{B}{V_d} \right) \qquad (4)$$

Plotting the logarithm of the lifetime τ as a function of $1/V_d$ yields again a straight line (Ref. 21). The latter

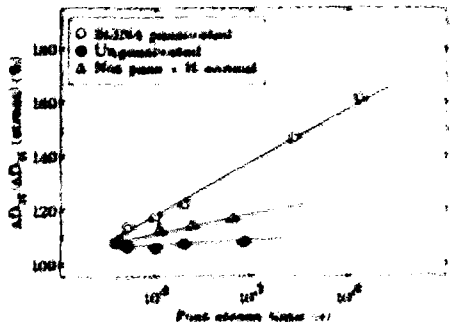


Fig. 1 : Post-stress interface trap generation for samples with nitride passivation, without passivation and without passivation but with N annealed (Ref. 23)

method has the advantage that the lifetime can be directly related to the power supply voltage, but it is only valid in a narrow range of gate voltages, near the maximum of the substrate current.

For monitoring the degradation, several parameters have been used: electrical parameters, like ΔV_t, $\Delta g_{m,max}/g_{m,max}$, $\Delta I_d/I_d$, which are measuring the change of the electrical characteristics, as well as more physical parameters, like the charge pumping current, which is more a measure of the real damage of the interface (Ref. 17, 22). The shift of the electrical parameters is determined both by the amount of damage itself and by the influence of the damage on the electrical characteristics. The advantage of using charge pumping in combination with the more conventional characteristics is that it provides information on the damage only, which permits to separate influence of the real damage from the influence of a certain damage on the electrical characteristics. Also by using the charge pumping current as a monitor, one can obtain information on the mechanisms that is at the basis of the damage generation (Ref. 17).

3.1.3 Influence of processing

When understanding the degradation mechanisms for the different types of MOSFET's, as discussed in subsection 3.1.1., and having at one's disposal the methods and models for the determination of the hot carrier lifetime, described in 3.1.2, it is important to understand the processing factors and the parameters that affect the hot carrier resistance of the devices. With this respect, three parameters are of major importance: 1) the amount of *hot carrier generation and injection* in the channel, 2) the amount of *physical damage*, generated by the injected hot carriers, and 3) the influence this damage has on the electrical characteristics of the MOSFET. All processing factors that can influence the hot carrier resistance have impact on one of these three parameters.

In section 4, it will be discussed that two processing factors are indeed the key factors in the strategy for improving the hot carrier reliability: on the one hand the doping profile of the drain junction, which should be optimized to minimize the lateral electrical field and consequently also the hot carrier generation in the channel, and to minimize also the influence of a certain damage on the electrical characteristics, and on the other hand the quality of the oxide, which should be optimized to minimize the *physical damage* under influence of the injected carriers.

Besides these two factors, however, a number of other effects can have an important influence on the hot carrier degradation. It is therefore not only important to be aware of these factors, but also to control them adequately in order to obtain a maximum device lifetime. In this subsection, three examples of such processing factors are discussed. In a first example, it is demonstrated that, for some types of passivation layers, the generation of interface damage can continue, even after termination of the hot carrier stress. In a second example, it is shown that for p-MOSFET's, the type of doping of the polysilicon gate can have a large effect on the hot carrier lifetime. Finally, in a third example it is demonstrated that the type of implanted and intermetal layers used in the backend process can have an important influence on the hot carrier degradation.

3.1.3.1 Post-stress effects

When the interface trap density of a MOSFET is monitored after termination of the hot carrier stress, a further and continuous increase is observed, even if all contacts are grounded during the post-stress period (Ref. 27). This post-stress interface trap generation proceeds logarithmically with time. Fig. 1 shows the post-stress interface trap generation for devices from two technologies that are identical, except for the passivation layer. The nitride passivated device shows a much larger post-stress effect than the unpassivated ones. When the unpassivated devices were annealed in hydrogen at 770°C during 30 minutes just before the hot carrier stress, the post-stress interface trap generation was significantly larger than that of the unpassivated devices without the anneal. This supports the hypothesis that the effect is related to the presence of hydrogen in the gate oxide.

The post-stress effect was investigated in more detail, and the influence of various parameters, such as the stress conditions, post-stress oxide field and temperature have been determined (Ref. 28). Based on this study, the following features were characterized: the post-stress interface trap generation increases logarithmically with time, and occurs only after hole injection conditions (i.e. low V_g and high V_d) during stress; its magnitude is proportional to the duration of the stress and depends on the passivation and the presence of hydrogen in the layers; it is enhanced for positive post-stress field polarities, and its rate can be altered by switching the oxide field from positive to negative polarities; it can be stopped by a brief electron injection, and it is suppressed when cooling down to 77K. A physical model, based on the release and motion of neutral and positive hydrogen by the detrapping of trapped holes has been proposed to explain these dependencies (Ref. 28). It will be shown in section 3.2 that these post-stress effects can have an important influence on the dynamic degradation of n-channel MOSFET's.

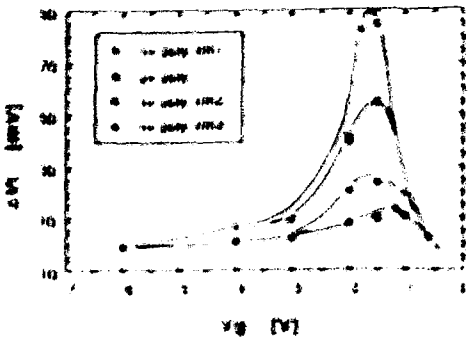


Fig. 4 Threshold voltage shift (reverse saturation) versus stress gate voltage after a stress of 1000s at Vds= 6V for three different n-type poly and one p-type poly devices (Ref. 27).

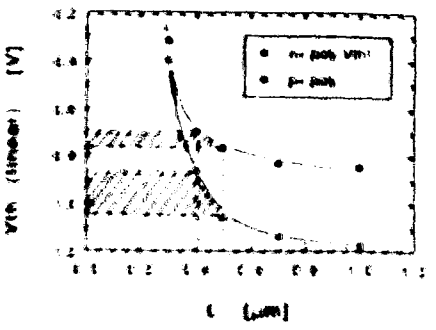


Fig. 6 Linear threshold voltage versus channel length for n+ poly and p+ poly devices (Ref. 23).

3.1.3.2 Influence of poly dopant for p-MOSFET's

When scaling down the technologies to the deep submicron level, p^+ poly-silicon is used more and more as the gate material of the p-channel devices in order to improve the short channel effects (Ref. 24). It is therefore mandatory to monitor the influence of this gate material on the hot carrier degradation performance (Refs. 26, 27).

Fig. 5 shows the gate voltage dependence of ΔV_{th} for a stress of 1000s at V_{ds}= -6V for four types of p-MOSFET's from a 0.3 µm CMOS process: 3 with n-type poly gates with different threshold voltage implants, and one with p-type poly gate. As expected, the maximum degradation occurs at the maximum of the gate current (cf. Fig. 1). The degradation of the devices with p-type gate, however, is smaller by about a factor of 5 as compared to the devices with an n-type gate.

In order to investigate this difference, a charge pumping analysis was performed. The result is shown on Fig. 7, where the shift in threshold voltage is plotted versus the shift in CP characteristics, which is directly proportional to the trapped electron charge (Ref. 22). From this figure it becomes clear that, for the same stress conditions, both types of MOSFET's have comparable damage (trapped charge) but the influence of the damage on the electrical characteristics is much higher for the n-type than for the p-type gate devices.

This larger effect of the same amount of trapped charges is completely explained by the larger short channel effect for the devices with n-type gates, as illustrated on Fig. 6. The same amount of trapped electron density leads to equal channel shortening in both types of devices. Due to the larger short channel effects for the n-type gates, however, this leads to a much larger shift in threshold voltage for the devices with the shortest gate lengths (Refs. 18, 23). By using p-type gates for the p-MOSFET's, the lifetime of the devices can be improved by about one order of magnitude. Therefore, p-type gates may be preferred for submicron technologies from the point of view of reliability as well.

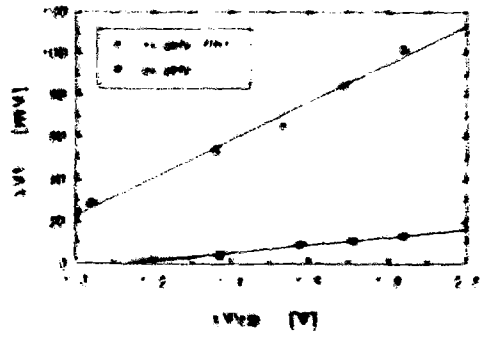


Fig. 5 Threshold voltage shift versus time for maximum degradation (Ref. 27).

3.1.3.3 Influence of passivation dielectric

As a final example, it has been found recently that the resistance of MOSFET's to hot carrier degradation is strongly affected by the materials covering the transistors (Refs. 26, 27, 28). The film composition and process conditions of both the interlayer (ILD) and the passivation (IMD) dielectric stack can play an important role and, if not adequately controlled, can lead to enhanced degradation.

Fig. 7 shows the results of an extensive study of the impact of ILD and IMD stack composition on hot-carrier lifetime for devices of a 0.5 µm TUM CMOS process (Ref. 28). As can be seen, a significant improvement is achieved with a Si-rich film instead of the standard PECVD TEOS oxide (PETEOS) in the ILD. The best results are obtained, however, by replacing the PETEOS by a phosphosilicate glass layer (PSG).

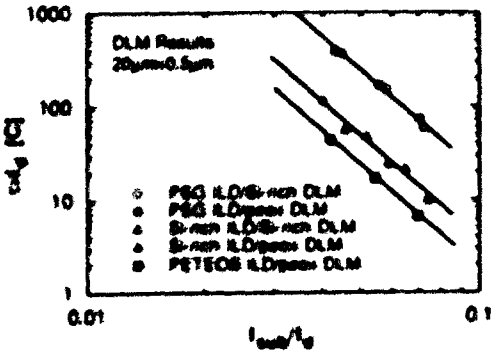


Fig. 7 - Influence of ILD and IMD composition on hot carrier lifetime (Ref. 28)

Analysis of the physical damage of the "good" (with PSG/PSOX) and "poor" (with PETEOS/PSOX) quality devices is shown on Fig. 8, which shows the trapped charge (ΔN_{ot}) and interface traps (ΔD_{it}) as a function of injected electron fluence under uniform substrate hot electron injection (Ref. 28). This clearly demonstrates that the major difference between both technologies is a more than 3 times higher interface trap generation rate for the poor quality devices. The reason for this higher generation rate is probably the presence of water-related species that diffuse into the gate oxide and result in an enhanced susceptibility to hot-electron induced damage. PSG layers seem to form an effective barrier against the diffusion of such water-related species (Ref. 28).

3.2 Dynamic degradation

In digital circuits, the devices are ushually stressed under static conditions. Although the static degradation conditions, treated in the previous section are already more realistic than the uniform injection conditions of section 2, they are still not relevant for the waveforms the devices see in real operation. Therefore, it is mandatory to study the degradation of the devices under the most realistic, i.e. dynamic conditions, and to know which effects have to be taken into account in order to extrapolate the results deduced under static stress conditions in a reliable way.

For a long time, there has been a lack of consensus on the degradation mechanisms that are occurring under dynamic stress conditions. It was shown recently, however, that this lack of consensus was mainly caused by problems with inductive ringing in the measurement setup. Since this effect was recognized and solved, the controversy regarding dynamic degradation has also largely disappeared, leading to the consensus that for most relevant cases, the degradation under dynamic operation behaves quasi-statically. It has to be kept in mind, however, that for some specific cases, additional effects, such as post-stress and detrapping effects can play an important role under dynamic operating conditions. For a more detailed overview of dynamic degradation effects, we refer to (Ref. 29).

3.2.1 Influence of measurement setup

A first prerequisite for the comparison of dynamic degradation results with those from static stress is that the measurement setup itself does not cause any additional degradation. It was found recently that setups, used for static stress measurements, are mostly not suited to perform dynamic degradation measurements since the parasitic inductances and capacitances of the 50Ω coaxial cables cause very large over- and undershoots of the drain and source voltages, as illustrated on Fig. 9 (Ref. 30). When the gate voltage is switched off in the presence of a high drain voltage, the voltage oscillations can result in a forward biased source diode in combination with an enlarged electric field at the drain. Both effects cause an enhanced degradation (Refs. 30, 31, 32). The largest part of the enhanced degradation effects published in the past as well as the lack of consistency among different research groups, can be ascribed to this effect.

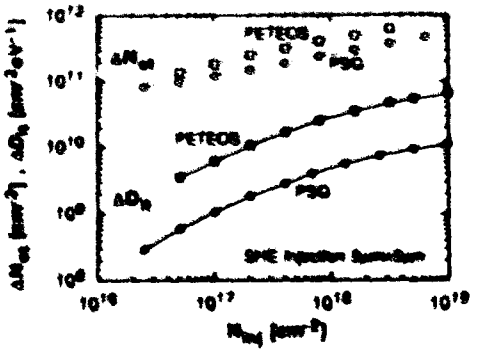


Fig. 8 - Physical damage (ΔN_{ot}, ΔD_{it}) in PSG and PETEOS-devices (Ref. 28)

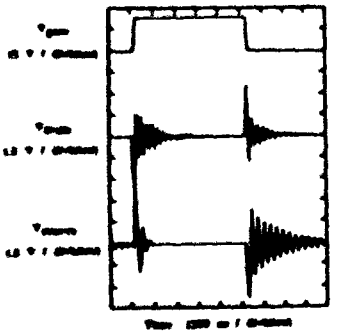


Fig. 9 - Fluctuations of V_s and V_d measured during dynamic operation without any precaution. F=500kHz, edgetime=10ns, V_d=7.5v, V_g=0/7.5V, W=100µm, L=1.7µm (Ref. 30)

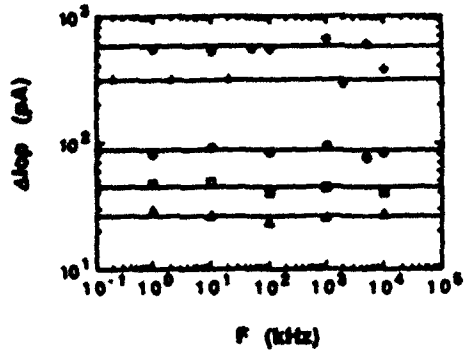


Fig. 10 - ΔI_{cp} versus F for nMOS transistors stressed with constant pulse shape conditions. o, Δ and □ are for conventional 2-μm transistors stressed at V_d = 7.25 V, V_g pulsed between 0 and 7.25 V, stress time = 6000 s, t_r = t_f = 10 % of the period of the V_g pulses. is for LDD 1.25-μm transistors stressed at V_d = 7.5 V, V_g pulsed between 0 and 7.5 V, stress time = 10000 s, t_r = t_f = 10 % of the period of the V_g pulses. + is for LDD 0.5-μm transistors stressed at V_d = 4.5 V, V_g pulsed between 0 and 4.5 V, stress time = 3000 s, t_r = t_f = 20 % of the period of the V_g pulses.

3.2.2 Comparison with static degradation

Taking into account the influence of these imperfections of the measurement setup, a renewed study of hot-carrier-induced dynamic degradation effects has been carried out showing that for all studied cases the degradation behaves quasi-statically. A typical measurement that is used to verify the quasi-static nature of the degradation behaviour and that is very sensitive to non-quasistatic effects occurring during the edges of the gate voltage pulse is the constant pulse shape experiment in which the rise and fall times are a constant portion of the period of gate voltage pulse. This means that when the frequency is increased, rise and fall times are decreased.
It is shown on Fig. 10 that for various technologies with different channel lengths no frequency dependence has been observed, indicating that no non-quasi-static degradation occurs during the switching of the gate voltage. Similar experiments with different waveforms did never reveal any enhanced dynamic degradation. Based on these experiments, the lifetime prediction methods deduced for static stress conditions can be easily adapted for dynamic operating conditions (Ref. 33).
This is illustrated on Fig. 11 where the lifetime curve (full line) is extrapolated based on static stress experiments and compared to results of dynamic stress conditions. The point labeled 'constant pulse shape' corresponds to measurements at different frequencies. ΔI_{cp} was taken as a monitor for the hot-carrier-induced degradation, in order to avoid problems with screening of the degradation by trapped oxide charge (Ref. 17).

3.2.3 Influence of post-stress and detrapping effects

In a last part of this section, the influence of two effects is discussed that are, however, often not taken into consideration and that can lead to a discrepancy between static and dynamic degradation results. Fig. 12 shows the result of a comparison between static and dynamic degradation. Curve C is measured after a dynamic stress where the drain voltage V_d = 7.75 V was constant and the gate voltage is pulsed from 0 to a varying gate voltage V_g. It is clear that the curves after static (A) and dynamic (C) stress do not coincide. This difference is due to the post-stress effect discussed above. Since in the experiment of Fig. 12, curve C, a positive charge is built up for the low gate voltage conditions, a post

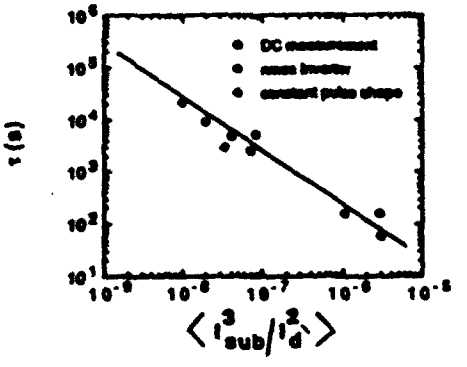


Fig. 11 - Lifetime (ΔI_{cp} = 20 pA) as a function of $\langle I_{sub}^3/I_d^2 \rangle$. Open circles are DC lifetime measurements, for the dynamic degradation experiments $\langle I_{sub}^3/I_d^2 \rangle$ is obtained from the pulse shape, the lifetime itself is measured.

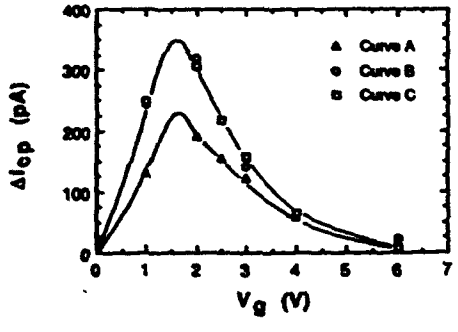


Fig. 12 - Curve A : ΔI_{cp} after static stress V_d = 7.75 V, Curve B : ΔI_{cp} after static stress and subsequent post-stress period of 9950 s, Curve C : ΔI_{cp} after dynamic stress at F = 25 kHz, width = 200 ns, t_r = t_f = 10 ns (Ref. 23)

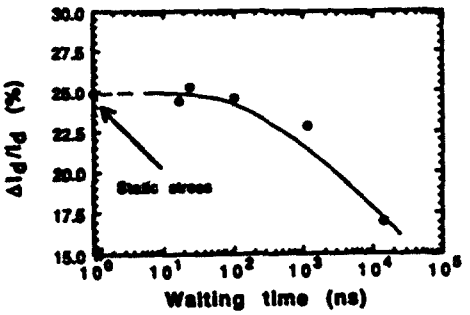


Fig. 13 - Dynamic degradation: V_g pulsed between 0 and -2.5 V and V_d = -10 V. T_r = t_f = 10 ns, width = 100 ns, number of pulses = 5.10^9, F = 5 MHz. Static degradation: V_g = -2.5 V, V_d = -10 V, stress time = 500 s.

stress effect will occur during the time $V_g = 0$. In order to take into account the post-stress effect, curve C was compared to the result of a static stress followed by a subsequent post-stress period, corresponding to the post-stress period of the dynamic stress. It is clear from Fig. 12 that curves B and C are identical for all gate voltage conditions.

A last effect that has to be taken into account - and that is of particular importance for p-channel transistors - is the detrapping of trapped charge. Fig. 13 shows $\Delta I_d/I_d$ versus the waiting time, i.e. the time when V_g=0V and $V_g = V_d$. It is clear that for increasing waiting (detrapping) time the degradation decreases. This can be explained by a decrease of the channel shortening effect due to a decrease of the trapped negative charge with time.

4. STRATEGIES FOR IMPROVING HOT CARRIER RELIABILITY

As already mentioned earlier, the hot carrier degradation problem has evolved from a mere academic topic towards a real bottleneck in the further downscaling of the technologies towards deep submicron geometries. Therefore, it is not only important to understand the degradation mechanisms, to model the degradation behaviour and to establish lifetime methods, but it becomes also mandatory to develop strategies for improving the hot carrier resistance of the devices. The key factors for improving the hot carrier reliability are on the one hand the *drain engineering*, which determines the amount of hot carrier generation for a given supply voltage and the influence of a certain physical damage on the electrical characteristics, and on the other hand the *oxide quality*, which determines the amount of physical damage for a given hot carrier injection fluence. A third important factor is the needed decrease of the power supply voltage for future deep submicron generations. In this section, each of these factors will be briefly discussed.

4.1 New drain engineering techniques

In view of a number of practical reasons, the most logical scaling scheme of device dimensions, which aimed at keeping the electrical fields constant, has been abandoned. In practice non-constant field scaling has been used. In this scheme the operation voltage does not scale with device dimensions and the electrical fields in MOSFET's increase significantly. In order to reduce these lateral electric fields for a given supply voltage and dimension, lightly-doped-drain (LDD) structures have been used successfully for the 1.2µm down to 0.7µm generations for a 5V supply voltage. In these devices the peak of the lateral field is lowered by reducing the doping concentration near the drain and by providing a smooth junction transition instead of an abrupt one. By self-aligned processing, the n^--region is located underneath a spacer oxide and connects the channel to the highly doped junction. During operation, electron trapping can occur in this (poor quality) spacer oxide, which is detrimental for device lifetime, as discussed in 3.1.1. Because the gate has only limited control over the underlying LDD-region, the channel current is indeed extremely sensitive to the buildup of negative charge in the spacer oxide. The degradation behaviour of these LDD devices is therefore quite different from that in the conventional transistors and characterized by a serious increase with time of the series resistance and a corresponding decrease of the channel current.

It was already predicted in 1986 that the LDD concept would not allow 0.5µm transistors to operate during 10 years at a 5V power supply (Ref. 34). Whereas reduction of the power supply down to 3.3V makes LDD still appropriate for the 0.5µm generation, it becomes a very questionable approach for the 0.35µm and even totally inadequate for the 0.25µm generations at this operating voltage. Since 1986 several advanced drain-engineered structures have therefore been proposed which attempt to overcome the major drawbacks of the LDD approach being the reduced current drive due to the n^+-gate offset and the spacer-induced degradation. In these advanced concepts the gate fully overlaps the drain. Examples are the ITLDD (Inverse-T gate LDD) (Ref. 35), the GOLD (Gate-Drain overlapped Device) (Ref. 36) and the LATID (Large-Tilt-angle Implanted Drain) concepts (Ref. 37). In the first two approaches (and in a number of variations to these concepts proposed since then) the n^- LDD and a self-alignment of the n^+ to-gate are achieved by a rather complex process sequence involving implantations through either a partially thinned polysilicon layer and/or special spacer structures. In the LATID device the n^- LDD source and drain regions are achieved by a phosphorus implantation under a tilt angle of around 45° and by rotating the wafer stepwise. The LATID structures are simpler to realize, provided the suited ion implanter is available. A drawback of all the advanced devices is the increased gate overlap capacitance. With these concepts lifetimes well over 10 years can be expected for 0.25µm devices operating at 3.3V (Ref. 37). The degradation behaviour of these

advanced structures is much more complex than for conventional and LDD structures and is still not very well understood. This is caused by the important impact of these advanced approaches on the current paths, on the location and value of the field peaks and on the (reduced) impact of generated damage (interface traps and trapped charges) on the electrical characteristics of the device. The analysis of this degradation behaviour will require more in-depth investigations in which accurate 2D-simulations will become indispensable.

4.2 Improved gate dielectrics

Besides the suppression of hot carrier generation by alternative drain engineering techniques, it is equally important to minimize the damage in the oxide for a given hot carrier injection fluence. In the past years, a lot of attention has been paid to the use of nitrided oxides and/or oxynitrides for improved dielectric reliability (Refs. 38-41). Such silicon oxynitride dielectrics are composed primarily of silicon dioxide, but with a small fraction of nitrogen buildup at the interface. There are different approaches to obtain such nitrided oxides. In the so-called ROXNOX-process (reoxidized nitrided oxide), the nitridation consists first of an oxidation in O_2, followed by a nitridation step in NH_3, and finally a reoxidation again in O_2 (Ref. 39). In another approach, N_2O is used as the nitridation gas. In this approach, the nitridation can be done by growing the oxide in pure O_2, followed by a nitridation step in N_2O, or the dielectric can be directly grown in N_2O ambient (Ref. 40). For both approaches, either furnace or RTP processes have been applied.

Concerning the influence of the nitridation on the charge trapping and interface trap generation, the picture is rather complicated. First of all, interface trap generation is found to be suppressed by the nitridation. This is attributed to the presence of nitrogen at the silicon-oxide interface, which reduces the number of strained Si-O bonds at the interface, which normally acts as interface trap precursors. Dielectrics with higher nitrogen content are therefore more effective in the suppression of the interface trap generation.

The charge trapping, on the other hand, is strongly enhanced for very heavy nitridations, especially for NH_3-nitrided oxides, which is explained by the presence of hydrogen in these oxides. The reoxidation step, which is intended to eliminate the H from the oxide, should therefore be carefully optimized (Ref. 39). This is particularly important in order to avoid an increased degradation of the p-MOSFET's, which are more sensitive to electron trapping. Recently, it was demonstrated that lightly nitrided oxides, performed by RTP in NH_3, show optimum conditions for hot carrier reliability at nitrogen concentrations as low as 0.5 at. % (Ref. 41).

In the case of N_2O-nitrided oxides, on the other hand, the nitridation is essentially hydrogen-free, and both the interface trap generation and electron trap density are reduced, leading to improvements in the lifetime with more than one decade (Ref. 40). The exact degradation mechanisms, hole and electron trapping and trap generation efficiencies, still need to be explained, however, and more research in this field is therefore still required.

4.3 What beyond 0.25μm generations ?

In the generations beyond 0.25μm (0.25μm - 0.07μm) hot carrier degradation can be expected to become a less important reliability problem as a consequence of the expected further reduction of the supply voltage and the gate insulator thickness. It is generally accepted that the critical issue when scaling down dimensions will shift from concerns related to hot carrier effects to those of suppressing short-channel effects. *If and only if* indeed a scaling scheme is pursued for the supply voltage going from 3.3-2.5V for the 0.25μm generation down to the 1.5-1.0V range for the 0.1-0.07μm generations, the corresponding reduction of the lateral field peak will indeed significantly alleviate the hot carrier problem. However, it is important to point to existing misconcepts regarding hot carrier generation and injection. It is quite often assumed that reduction of the drain voltage below 3.3-2.7V will automatically eliminate the problem of carrier injection because carriers would no longer gain sufficient energy to overcome the $Si-SiO_2$ energy barrier. Moreover, a reduction of this voltage below the threshold energy for impact ionization (~ 1.6eV) is often believed to also yield an elimination of electron-hole pair creation by impact ionization. Both assumptions have been unambiguously disproved by many recent studies. Impact ionization at room temperature for voltages well below the threshold energy for impact ionization and even below the band gap energy (down to 0.6V) has indeed been observed in deep submicron 0.1μm Si MOSFETs (Refs. 42, 43). Gate currents have been measured directly by Chung *et al.* (Ref. 42) in 0.1μm devices at drain voltages well below the $Si-SiO_2$ energy barrier value (down to 1.75V). Moreover gate currents which were deduced from threshold voltage shifts in 0.3μm floating gate devices have been reported for drain biases as low as 1.4V (Ref. 44). There are different possible explanations for the existence of hot electron currents at these low voltages (Ref. 42) (on which we will however not further elaborate in this paper). The quasi-equilibrium model states that the electron gas is in quasi-thermal equilibrium with the electric field at all times. This postulates that the electrons have a certain probability to acquire any energy in the drift field even under low drain bias (Ref. 6). The low voltage substrate and gate currents are due to electrons existing in the high energy region of the energy distribution. The relationship between gate current (I_G) and substrate current (I_{sub}) measured in these deep-submicron devices at low voltages was found to remain the same as the one derived for the micron device generations (Ref. 42):

$$\frac{I_G}{I_D} = \left(\frac{I_{sub}}{I_D}\right)^{\phi_B/\phi_i} \qquad (1)$$

in which ϕ_B is the effective barrier height and ϕ_i is the critical energy for impact ionization.

The degradation of deep-submicron devices at reduced voltages has been reported by several groups (Refs. 42, 45-47). Again, the existing degradation models derived for the 1μm generation and below, still seem to remain more or less applicable in the deep-submicron regime at low voltages. Reduction of the lateral fields will thus yield much smaller substrate currents and therefore, according to these models, considerably higher lifetimes. However, it also turns out that at a given substrate current, the hot carrier degradation worsens as L_{eff} is reduced (Ref. 42, 45), an observation which deserves close attention. This can be explained by the non-scalability of the degraded region by which a relatively larger part of the channel is affected which in turn leads to a stronger impact on the device characteristics. Reduction of the gate insulator thickness will on the one hand lead to a larger substrate current (for a given drain bias) but this is counteracted by less trapping in the insulator, as mentioned in 2.1, and therefore less MOSFET degradation. The overall degradation is therefore not a straightforward function of the gate insulator thickness. Just as an example of hot carrier degradation in deep-submicron devices, for the double punch-through stopper transistor with effective channel length of 0.07μm reported by Koyanagi *et al* (Ref. 46), the maximum supply voltage to guarantee a hot carrier lifetime (defined for a 10% g_m degradation) of 10 years in this device was found to be 1.7V. This result demonstrates on the one hand that considerable hot carrier degradation indeed occurs even at these low voltages but on the other hand that a reduction of the operation voltage down to 1-1.5V for this generation can indeed provide sufficient immunity.

5. CONCLUSIONS

In this paper, it was demonstrated that the hot carrier reliability problem has evolved from a more or less academic topic towards a real bottleneck for the further downscaling of MOSFET-technologies. The main degradation mechanisms have been reviewed, under uniform and non-uniform static and dynamic conditions. Some important processing related effects have been illustrated, and the strategies for the improvement of the hot carrier lifetime have been briefly discussed.
With respect to the future, it must be concluded that even for a consistent reduction of the supply voltage and the corresponding reduction of the generation and injection of hot carriers in 0.1μm devices, this reliability issue still requires to be considered with care. Although it is believed that the advanced drain engineering approaches may somehow be relaxed for deep-submicron technologies *in case of appropriate voltage supply reduction*, the hot carrier problem clearly can not be ignored and should be accounted for.

6. ACKNOWLEDGEMENTS

Some of the results discussed were obtained partially with the support of Alcatel Bell, Antwerp, within a project with the IWT (Flemish Institute for Science and Technology), and partially within the ADEQUAT-project (JESSI BT1B/ESPRIT 7236). The authors also acknowledge Dr. L. Deferm and his co-workers of the IMEC CMOS Process Integration group for processing support.

7. REFERENCES

1. D. J. DiMaria and J. W. Stasiak, "Trap creation in silicon dioxide produced by hot electrons", *J. Appl. Phys.*, vol. 65, p. 2342, 1989

2. A. v. Schwerin and M. M. Heyns "Oxide field dependence of bulk and interface trap generation in SiO_2 due to electron injection", *Proc. Insulating Films on Semiconductors (INFOS)*, p. 263, 1991

3. T. Nishida and S. E. Thompson, "Oxide field and temperature dependences of gate oxide degradation by substrate hot electron injection", *Proc. Intern. Reliability Physics Symposium (IRPS)* , p. 310, 1991

4. M. M. Heyns, D. Krishna Rao and R. F. De Keersmaecker, "Oxide field dependence of Si-SiO_2 interface state generation and charge trapping during electron injection", *Proc. Insulating Films on Semiconductors (INFOS)* , p. 327, 1989

5. M. M. Heyns and A. v. Schwerin, "Charge trapping and degradation of thin dielectric layers", *Proc. Insulating Films on Semiconductors (INFOS)*, p. 73, 1991

6. P. Heremans, R. Bellens, G. Groeseneken, A. v. Schwerin, W. Weber, M. Brox and H. E. Maes, "The mechanisms of hot carrier degradation", *Hot carrier design considerations in MOS devices and circuits*, Van Nostrand Reinhold, New York, Chapter 1, 1992

7. Y. Nissan-Cohen, J. Shappir and D. Frohman-Bentchkowsky, "Trap generation and occupation dynamics in SiO_2 under charge injection stress", *J. Appl. Phys.*, vol. 60, p. 2024, 1986

8. D. J. DiMaria, "Temperature dependence of trap creation in silicon dioxide", *J. Appl. Phys.*, vol. 68, p. 5234, 1990

9. G. Van den bosch, G. Groeseneken and H. E. Maes, "Direct and post-injection oxide and interface trap generation resulting from low-temperature electron injection", *J. Appl. Phys.*, vol. 74, p. 5582, 1993

10. A. v. Schwerin, M. M. Heyns and W. Weber, "Investigation on the oxide field dependence of hole trapping and interface state generation in SiO_2 layers using homogeneous nonavalanche injection of holes", *J. Appl. Phys.*, vol. 67, p. 7595, 1990

11. L. Lipkin, A. Reisman and C. K. Williams, "Hole trapping phenomena in the gate insulator of As-fabricated insulated gate field effect transistors", *J. Appl. Phys.*, vol. 68, p. 4620, 1990

12. G. Van den bosch, G. Groeseneken, *H. E. Maes, R.* B. Klein and N. S. Saks, "Oxide and interface degradation resulting from substrate hot-hole injection in metal-oxide-semiconductor field-effect transistors at 295 and 77 K", *J. Appl. Phys.*, vol. 75, p. 2073, 1994

13. G. Van den bosch, G. Groeseneken, P. Heremans, M. Heyns and H. E. Maes, "Hole trapping and hot-hole induced interface trap generation in MOSFET's at different temperatures", *Proc. ESSDERC*, p. 477, 1992

14. Q. D. M. Khosru, N. Yasuda, A. Maruyama, K. Taniguchi and C. Hamaguchi, "Spatial distribution of trapped holes in the oxide of metal oxide semiconductor field-effect transistors after uniform hot-hole injection", *Jap. J. Appl. Phys.*, vol. 30, p. 3652, 1991

15. Q. D. M. Khosru, N. Yasuda, K. Taniguchi and C. Hamaguchi, "Oxide thickness dependence of interface trap generation in a metal-oxide-semiconductor structure during substrate hot-hole injection", *Appl. Phys. Lett.*, vol. 63, p. 2537, 1993

16. A. v. Schwerin and M. M. Heyns, "Homogeneous hole injection into gate oxide layers of MOSFET's: injection efficiency, hole trapping and Si/SiO_2 interface state generation", *Proc. Insulating Films on Semiconductors (INFOS)*, p. 283, 1991

17. P. Heremans, R. Bellens, G. Groeseneken, H.E. Maes, "Consistent model for the hot carrier degradation in n-channel and p-channel MOSFET'S", *IEEE Trans. El. Dev.*, vol. 35, p. 2194, 1988

18. R. Woltjer, A. Hamada, E. Takeda, "Time dependence of p-MOSFET Hot carrier degradation measured and interpreted consistently over ten orders of magnitude", *IEEE Trans. El. Dev.*, vol. 40, p. 392, 1993

19. P. Heremans, G. Van den bosch, R. Bellens, G. Groeseneken, H.E. Maes, "Temperature dependence of channel hot carrier degradation in n-channel MOSFET's", *IEEE Trans. El. Dev.*, vol. 37, p. 980, 1990

20. C. Hu, S.C. Tam, F.C. Hsu, P.K. Ko, K.W. Terrill, "Hot electron induced MOSFET degradation - model, monitor and improvement", *IEEE Trans. El. Dev.*, vol. 32, p. 375, 1985

21. E. Takeda, N. Suzuki, "An empirical model for device degradation due to hot carrier injection", *IEEE El. Dev. Lett.*, vol. 4, p. 111, 1983

22. P. Heremans, J. Witters, G. Groeseneken, H.E. Maes, "Analysis of the charge pumping technique and its application for the evaluation of MOSFET degradation", *IEEE Trans. El. Dev.*, vol. 36, p. 1318, 1989

23. R. Bellens, E. de Schrijver, G. Van den bosch, G. Groeseneken, P. Heremans, H.E. Maes, "On the hot carrier induced post-stress interface trap generation in n-channel MOS transistors", *IEEE Trans. El. Dev.*, vol. 41, p. 413, 1994.

24. F. Matsuoka, H. Iwai, H. Hayashida, K. Hama, Y. Toyoshima, K. Maeguchi, "Analysis of hot carrier induced degradation mode on p-MOSFET's", *IEEE Trans. El. Dev.*, vol. 37, p. 1487, 1990

25. C. Monsérié, R. Bellens, G. Groeseneken, H.E. Maes, "Comparative study of hot carrier degradation in p^+ and n^+ poly p-MOSFET's of a 0.5μm CMOS technology", *Proceedings ESSDERC*, p. 829, 1993

26. N. Lifschitz and G. Smolinsky, "Hot carrier aging of the MOS transistor in the presence of spin-on-glass as interlevel dielectric", *IEEE El. Dev. Lett.*, vol. 12, p. 140, 1991

27. V. Jain, D. Pramanik, S.R. Nariani, C. Hu, "Internal passivation for suppression of device instabilities induced by backend processes", *Proc. Intern. Reliability Physics Symposium (IRPS)*, p. 11, 1992

28. G. Van den bosch, L. Deferm, L. Forester, T. Collins, "Improved hot carrier reliability in a 0.5μm TLM CMOS process by back-end process optimization", *Proceedings ESSDERC*, 1994

29. W. Weber, M. Brox, R. Bellens, P. Heremans, G. Groeseneken, A. v. Schwerin, H.E. Maes, "Hot carrier degradation during dynamic stress", *Hot carrier design considerations in MOS devices and circuits*, Van Nostrand Reinhold, New York, Chapter 4, 1992

30. R. Bellens, P. Heremans, G. Groeseneken, H.E. Maes, W. Weber, "The influence of the measurement setup on enhanced AC hot carrier degradation of MOSFET's", *IEEE Trans. El. Dev.*, vol. 37, p. 310, 1990

31. E. Takeda, R. Izawa, K. Umeda, R. Nagai, "AC hot carrier effects in scaled MOS devices", *Proc. Intern. Reliability Physics Symposium (IRPS)*, p. 118, 1991

32. K. Mistry, B. Doyle, "The role of electron trap creation in enhanced hot-carrier degradation during AC stress", *IEEE El. Dev. Lett*, vol. 11, p. 267, 1990

33. W. Weber, "Dynamic stress experiments for understanding hot-carrier degradation phenomena", *IEEE Trans. El. Dev.*, vol. 35, p. 1476, 1988

34. M. Kakumu, M. Kinugawa, K. Hashimoto and J. Matsunaga, "Power supply voltage for future CMOS VLSI in half and sub micrometer", *IEDM Tech. Dig.*, p. 399, 1986

35. T. Huang, W.W. Yao, R.A. Martin, A.G. Lewis, M. Koyanagi and J.Y. Chen, "A novel submicron LDD transistor with inverse-T gate structure", *IEDM Tech. Dig.*, p. 742, 1986

36. R. Izawa, T. Kure and E. Takeda, "Impact of gate-drain overlapped device (GOLD) for deep submicron VLSI", *IEEE Trans. El. Dev.*, vol. 35, p. 2088, 1988

37. T. Hori, J. Hirase, Y. Odake and T. Yasui, "Deep-submicrometer Large-Angle-Tilt Implanted Drain (LATID) Technology", *IEEE Trans. El. Dev.*, vol. 39, p. 2312, 1992

38. C.G. Sodini, K. S. Krisch, "Silicon Oxynitride gate dielectrics for scaled CMOS", *IEDM Tech. Dig.*, p. 617, 1992

39. T. Hori, "Nitrided gate oxide CMOS technology for improved hot carrier reliability", *Proc. Insulating Films on Semiconductors (INFOS), Microelectronic Engineering*, vol. 22, p. 245, 1993

40. A. Ditali, V. Mathews, P. Fazan, "Hot carrier induced degradation of gate dielectrics grown in nitrous oxide under accelerated aging", *IEEE El. Dev. Lett.*, vol. 13, p. 538, 1992

41. H.S. Momose, T. Morimoto, Y. Ozawa, K. Yamabe, H. Iwai, "Electrical characteristics of rapid thermal nitrided oxide gate n- and p-MOSFET's with less than 1 atom% nitrogen concentration", *IEEE Trans. El. Dev.*, vol. 41, p. 546, 1994

42. J. Chung, M-C. Jeng, J.E. Moon, P.K. Ko and C. Hu, "Low-voltage hot-electron currents and degradation in deep-submicrometer MOSFETs", *Proc. Int. Rel. Phys. Symp. (IRPS)*, p. 92, 1989

43. L. Manchanda, R.H. Storz, R.H. Yan, K.F. Lee and E.H. Westerwick, "Clear observation of sub-band gap impact ionziation at room temperature and below in 0.1μm Si MOSFETs", *IEDM Tech. Dig.*, p. 994, 1992

44. E. Sangiorgi, B. Ricco and P. Olivo, "Hot electrons and holes in MOSFETs biased below the Si-SiO_2 interfacial barrier", *IEEE El. Dev. Lett.*, EDL-6, p. 513, 1985

45. S. Cristoloveanu, "Physical Mechanisms of Hot-Carrier-Induced degradation in deep-submicron MOSFETs", *Proc. ESSDERC*, p. 797, 1993

46. M. Koyanagi, "Technology and Device design constraints for Low Voltage Low Power sub-0.1μm CMOS Devices", *Proc. ESSDERC*, p. 935, 1993

47. T. Mizuno, A. Toriumi, M. Iwase, M. Takahashi, H. Niiyama, M. Fukumoto and M. Yoshimi, *IEDM Tech. Dig.*, p. 695, 1992

PRE-BREAKDOWN CHARGE TRAPPING IN HIGH FIELD STRESSED OXIDES

L Dong, M J Tunnicliffe and V M Dwyer

Department of Electronic and Electrical Engineering
Loughborough University of Technology
LEICS, LE11 3TU, U.K.

Tel: +44 509 222856, Fax: +44 509 222854

ABSTRACT

Constant voltage stresses were applied to MOS capacitor structures and the resulting tunnelling current profiles were recorded. It was found that hole trapping occurred only during the first few seconds of the stress, while the electron trapping occurred throughout the stress period. A quantitative model for charge trapping during the build-up stage is presented and is shown to agree well with the experimental current profiles.

1.INTRODUCTION

Gate dielectric breakdown is a major cause of MOS circuit failure, especially in very large scale circuits. It has been widely accepted that oxide breakdown is triggered when the accumulation of trapped holes in the oxide reaches a critical value. A proper understanding of charge trapping is therefore vital if an accurate breakdown model is to be developed.

The build-up stage of the dielectric breakdown starts when electrons enter the oxide conduction band by means of the Fowler-Nordheim tunnelling mechanism. The tunnelling current is exponentially dependent upon the cathode electric field, which is affected by the trapped charges. The investigation reported here focuses on the tunnelling current profile, from which the charge trapping model is proposed.

The phenomenon of the charge trapping in thermally grown SiO_2 has been a subject of many investigations. Liang and Hu (Ref.1) presented a mathematical model of electron trapping and trap generation in silicon dioxide. Chen et al. (Ref.2) reported that holes were trapped after they were generated by high-energy electrons, while Jenq (Ref.3) suggested hole trapping was the result of direct tunnelling between the anode and the trap sites.

New experimental data is presented in this paper and a mathematical model of charge trapping in SiO_2 is proposed.

2. EXPERIMENTAL RESULTS

Experimental studies were limited to a single 4" <100> p-type Si wafer, containing 215x268μm MOS capacitor structures. The oxide thickness of the capacitors had a mean value of 41.42nm and a standard deviation of 0.161nm.

Constant voltage stress was generated by a Hewlett Packard HP4145B parametric analyzer and applied to the devices-under-test (DUT) between the gate and substrate terminals. Throughout the experiments, the gates were stressed negatively with respect to substrate and the oxide current was monitored by the analyzer as a function of time.

During the first set of experiments, relatively low stress voltages (-39V,-39.5V and -40V) were applied to the devices. The resulting tunnelling currents (Fig.1) increase from their initial values to a maximum, before decaying with time.

Subsequent experiments, performed using higher stress voltages (-42V and -43V), showed tunnelling currents which appeared to decay from the very beginning (Fig.2).

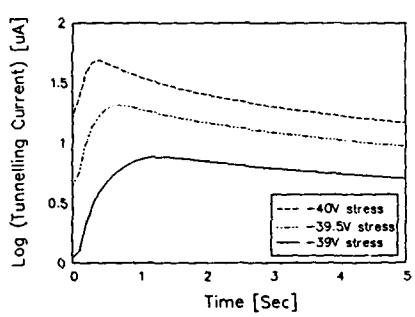


Figure 1: Tunnelling current profiles for stresses -39V, -39.5V and -40V.

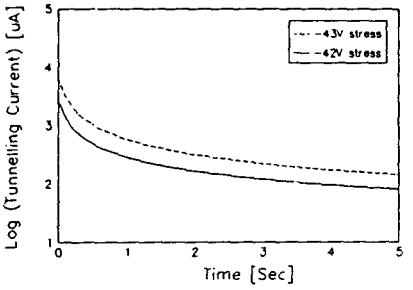


Figure 2: The oxide tunnelling current profiles for stress -42V and -43V.

The tunnelling current profiles may be characterised by two parameters: (i) the maximum current and (ii) the time at which this current is reached. The variation of these parameters with stress voltage is shown in Figs.3 and 4.

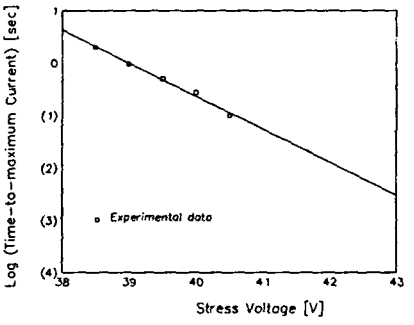


Figure 3: The time of maximum tunnelling current as a function of stress voltage.

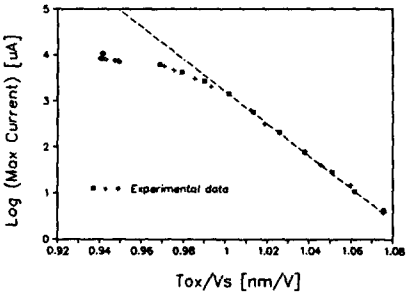


Figure 4: Maximum current vs. stress field.

The logarithm of the maximum tunnelling current varies linearly with 1/E at lower fields (in accordance with standard tunnelling theory (Ref.4)), and saturates at higher fields (Fig.4).

3. MATHEMATICAL MODELLING

The oxide current is determined by the Fowler-Nordheim equation (Ref.4):

$$J(t) = A E_{cat}^2(t)\, e^{-\frac{B}{E_{cat}(t)}} \qquad (1)$$

where J(t) is the tunnelling current density, E_{cat} is the cathode field and A and B are constants. The profile of tunnelling current in Fig.1 can be understood from a consideration of the charge trapping mechanism and its effect upon $E_{cat}(t)$.

3.1 Negative charge trapping

The non-saturating tunnelling current decay observed in Figs.1 and 2 suggests a continuous generation of electron trap sites. From first order rate equation, the trapped electron density Q_{ot}^- can be determined using the model of Liang and Hu (Ref.1)

$$Q_{ot}^- = qN_{op}\left(1-e^{-\int_0^t \sigma J dt/q}\right) + \int_0^t gJ(t)dt - \frac{qg}{\sigma_g}\left(1-e^{-\int_0^t \sigma_g J dt/q}\right) \qquad (2)$$

where σ and σ_g are the respective capture cross-sections of pre-existing traps and newly generated traps, and, N_{op} is the pre-existing trap density and g is the trap generation rate.

3.2 Positive charge trapping

The increase of oxide current in Fig.1 can be explained as a consequence of positive charge trapping. Since the current increase is very rapid, it is believed to be formed by direct hole tunnelling between the anode and the trap sites near the SiO_2/Si interface (Ref.3). The increment of trapped charge in time interval dt is given by

$$dQ_{ot}^{+}(t) = J_{hole}(t)dt \qquad (3)$$

where J_{hole} is the hole tunnelling current. The latter may reasonably be assumed to have a F-N type dependence upon anode field E_{an} and to be proportional to the density of available hole traps in the anode (Ref.5), i.e.

$$J_{hole} = Ce^{-\frac{D}{E_{an}}}\left(1 - \frac{Q_{ot}^{+}}{qN_{int}}\right) \qquad (4)$$

where N_{int} is the total density of hole traps and C and D are constants dependent upon structure of the anode/oxide/trap system. The factor in parenthesis causes the trapping to slow as the available number of states decreases, and to tend to 0 as $Q_{ot}^{+} \rightarrow qN_{int}$.

The oxide field profile changes when charges are trapped within the dielectric. If we assume sheet charges located at their respective centroids, then the electric field at cathode will be given by:

$$E_{cat} = \frac{V_{ox}}{T_{ox}} - \left(1 - \frac{x_n}{T_{ox}}\right)\frac{Q_{ot}^{-}}{\epsilon_{ox}} + \left(1 - \frac{x_p}{T_{ox}}\right)\frac{Q_{ot}^{+}}{\epsilon_{ox}} \qquad (5)$$

in which Q_{ot}^{+} and Q_{ot}^{-} are the magnitudes of the trapped hole and electron densities, x_p and x_n are the respective centroids of Q_{ot}^{+} and Q_{ot}^{-} and T_{ox} is the oxide thickness. The voltage across the bulk Si resistance ($\approx$250Ω) is much smaller than the stress voltage increment ($\geq$0.5V) and was therefore ignored in the model.

The effect of trapped holes on the anode field can be expressed as:

$$E_{an} = \frac{V_{ox}}{T_{ox}} - \frac{Q_{ot}^{+}}{\epsilon_{ox}}\frac{x_p}{T_{ox}} + \frac{Q_{ot}^{+}}{\epsilon_{ox}}\frac{x_n}{T_{ox}} \qquad (6)$$

Simultaneous numerical solution of equations (1-6) allows the tunnelling current to be calculated. Parameters were given values consistent with earlier publications, with the exception of N_{int}, which was adjusted in order to obtain an optimal correlation with the experimental data. The results are shown in Fig.5 and Fig.6.

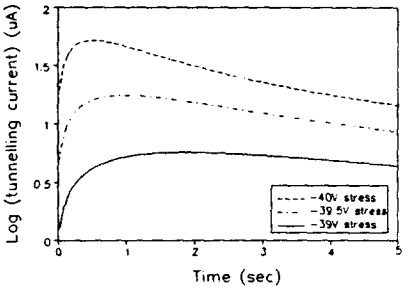


Figure 5: Modelling results of tunnelling current profiles for lower voltage stresses.

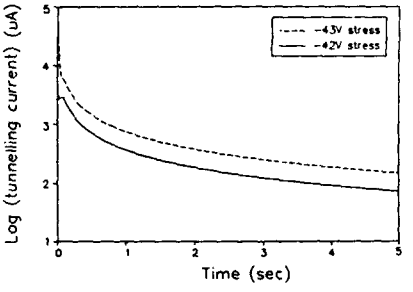


Figure 6: Modelling results of tunnelling current profiles for higher voltage stresses.

The results in Fig.2 and Fig.3 can also be explained by this model. Firstly, since the density of available traps is fixed, the time required to fill them decreases linearly with increasing current and hence exponentially with increasing voltage (Eqn.4). Secondly, the rapid hole-trap filling at high fields (~10ms for -42V stress) lies beyond the time-interval of the parametric analyzer (which was set to 100ms throughout these experiments). Hence the maximum recorded current lies within the decay-portion of the current profile, where the electron trapping compensates the trapped hole charge. The *apparent* maximum current therefore increases less

rapidly with increasing field, as shown in Fig.3.

4. CONCLUSION

Dielectric breakdown in MOS devices appears to be triggered by the accumulation of holes in oxide trap states. The build up of these trapped holes may be monitored by studying the current through the oxide. Fig.1 clearly show the effects of charge trapping.

This paper presents a simple physical model to describe such effects. In the model, the initial current rise is dominated by the tunnelling of holes from the silicon valence band directly into trap states located near the Si/SiO_2 interface. This mechanism soon appears to saturate, after which the slower electron trapping begins to dominate, causing the observed tunnelling current decay.

Despite the fact that hole tunnelling into such traps is poorly understood, and the relative simplicity of the model, a reasonable agreement is found between calculated and measured currents.

5. REFERENCES

1. M-S.Liang and C.Hu, "Electron Trapping in Very Thin Thermal Silicon Dioxides", IEDM Tech. Dig. (Washington D.C), pp.396-399, 1981.

2. I.C.Chen, S.E.Holland and C.Hu, "Hole Trapping and Breakdown in Thin SiO_2", IEEE Trans. on Electron Device Letters, Vol.EDL-7, No.3, pp.164-167, 1985.

3. C.S.Jenq, T.R.Ranganath, C.H.Huang, H.S.Jones and T.T.L.Chang, "High-field Generation of Electron Traps and Charge Trapping in Ultra-thin SiO_2", IEDM Tech. Dig. (Washington D.C), pp.388-391, 1981.

4. M.Lenzlinger and E.H.Snow, "Fowler-Nordheim Tunnelling into Thermally Grown SiO_2", J. Appl. Phys., Vol.40, No.1, pp.278-283, 1969.

5. V.M.Dwyer, M.J.Tunnicliffe, L.Dong, to be published.

INVESTIGATION OF ACCURACY OF WORST-CASE HOT-CARRIER RELIABILITY LIFETIMES AS PREDICTED BY BERKELEY'S MODEL

Alan Mathewson Paolo O'Sullivan Paul Hurley Alan Mathewson
National Microelectronics Research Centre Ireland
Tel +353 21 276471 Telefax +353 21 [illegible] e-mail amathews@ucc.ie

ABSTRACT

Hot carrier effects pose a significant reliability problem in modern MOS processes. An accurate method of predicting hot carrier lifetime is essential for the development of fine geometry MOS technology. A hot carrier degradation model developed by C. Hu et al at the University of Berkeley is widely used to predict device lifetime at given operating conditions from the results of accelerated tests. This paper demonstrates a new method of performing hot carrier stress measurements which satisfies the key demands of this model. This method combines adjusting device drain voltage in order to maintain a constant value of substrate to drain current. This method is employed to show that the Berkeley model makes a maximum lifetime prediction which is about an order of magnitude too short at accelerated stress conditions. This result casts doubt on the reliability of the Berkeley model for use in circuit reliability simulation and for use in setting industrial reliability benchmarks. A new understanding of the importance of the gate oxide voltage during hot carrier reliability characterisation using the Berkeley model is also discussed.

1 INTRODUCTION

Hot carrier effects are a significant reliability issue in fine geometry MOS processes [1, 2]. Consequently, an accurate method of predicting hot carrier lifetimes is essential for the development of fine geometry MOS technology.

The intense electric fields in the drain region of the device result in high energy free carriers termed 'hot carriers'. A finite number of these carriers can be injected into the gate oxide resulting in fixed oxide charge N_{ox} and the generation of interface states N_{it}. This is observed experimentally as a long term shift in the threshold voltage and transconductance.

A hot carrier degradation model [3] developed by C. Hu et al at the University of Berkeley, the 'Berkeley Model' is widely used to predict device lifetime at given operating conditions from the results of accelerated tests. The model predicts device lifetime τ in terms of the drain current I_d and substrate current I_b as

$$\tau = \frac{CW}{I_d} \left(\frac{I_b}{I_d}\right)^{-m} e^{\frac{\phi_{it}}{qE_{ox}\lambda}}$$

where W is device width, and C and m are process dependent parameters.

2 BERKELEY MODEL DEVICE LIFETIME PREDICTION

Hot carrier lifetime predictions are usually [4] expressing conditions are made based on the results of accelerated stress tests. The substrate degradation mechanisms are accelerated via high drain biases. Hot carrier stressing has traditionally been carried out at the gate voltage which produces I_b max, the maximum substrate current since I_b is considered to be the primary measure of hot carrier damage [5].

The definition of lifetime used in this work

to the time to observe 10% degradation in device maximum transconductance.

Values for the model parameters n and C are obtained from lifetime measurements at a range of $\frac{I_b}{I_d}$ ratios. Using the model, a device lifetime prediction is made at any desired bias condition from measurements of I_b and I_d at the given condition.

Two approaches are possible in making a lifetime prediction with the Berkeley model. The simpler method is to assume that there is negligible error in taking I_b and I_d to remain constant during the lifetime of a device. This is not generally true during the lifetime of a device. The alternative, to model the evolution of I_b and I_d during a stress, is difficult to perform with any accuracy.

A test procedure is developed in this paper, however, which allows Berkeley model lifetime predictions to be verified without making significant assumptions. This procedure involves keeping the $\frac{I_b}{I_d}$ ratio constant during stressing. This method permits a direct check of the accuracy of the model. The establishment of model accuracy under these conditions is a prerequisite for successful lifetime predictions under DC conditions.

3 CONSTANT-RATIO STRESSING

An inspection of the model shows that two parameters can potentially vary during a stress, namely I_d and $\frac{I_b}{I_d}$ (the latter is referred to as the 'current ratio'). Clearly, $\frac{I_b}{I_d}$ and I_d cannot be simultaneously held constant. However, as the current ratio is raised to a power of between 3.0 [1] and 5.5 [6], it is $\frac{I_b}{I_d}$ which has the greater impact on device lifetime.

It was thus decided to control only the current ratio and to monitor the effect on I_d.

The current ratio is controlled by continuous drain voltage adjustment during stressing. This method is referred to as the 'constant-ratio' method.

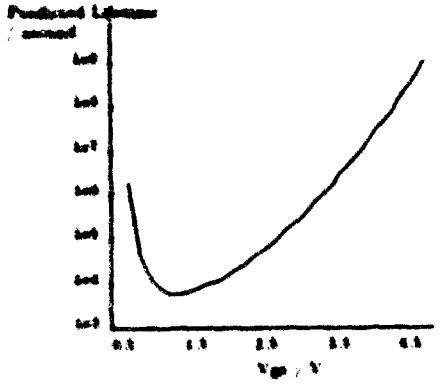


Figure 1 Berkeley lifetime contour plot of device lifetime prediction versus V_{gs} at a fixed V_{ds}.

4 GENERATION OF BERKELEY PREDICTION PLOT

A hot carrier characterisation of a batch of devices was performed via constant-ratio tests at a range of current ratios. The devices used were $W/L = 20\mu m/1.5\mu m$ n-MOSFETs. All the tests were carried out with V_{gs} set to initially give $I_{b(max)}$. The derived values of n and C were used to make a prediction of device lifetime as a function of V_{gs} at an elevated value of V_{ds}, by measuring I_b and I_d during a V_{gs} sweep and applying equation 1. The set of predictions is referred to as the Berkeley lifetime contour and is plotted in Fig. 1. The predicted lifetimes are sufficiently short to allow a partial verification of the Berkeley lifetime contour in a reasonable time.

The Berkeley lifetime contour suggests that the minimum hot carrier lifetime of a device occurs at a gate voltage considerably below the $I_{b(max)}$ condition. This has been reported before [6]. Hu et al. suggest, when presenting the Berkeley model [1], that degra-

dation is fastest at the $I_{b(max)}$ condition. A contradiction thus exists which suggests that, while confidence in reliability predictions at $I_{b(max)}$ is high [1], the model was not well verified at low V_{gs}. Since this is the region of minimum predicted lifetime, an understanding of model accuracy at low V_{gs} is important. (It is reported that the degradation mechanisms at gate voltages near threshold are significantly different to those at $I_{b(max)}$ [7].)

5 EVOLUTION OF V_{ds} AND I_d DURING CONSTANT-RATIO STRESSING

Results are now presented to demonstrate the suitability of the constant-ratio method when working with the Berkeley model.

V_{ds} and I_d are recorded during stressing to check that these variables do not vary excessively during constant-ratio stressing. Fig. 2 shows the evolution of V_{ds}, I_d and $\frac{I_b}{I_d}$ during a stress at V_{gs}=1.4V while the ratio is maintained at an approximately constant value. Except for a few minor excursions, the value of $\frac{I_b}{I_d}$ remains within a +/- 0.1% band.

Fig. 3 shows plots of V_{ds} and I_d as functions of time during a stress at the $I_{b(max)}$ condition. The current ratio is controlled to a similar degree as at V_{gs}=1.4V.

The changes in I_d and V_{ds} observed during the constant-ratio stresses are tabulated in Table 1. The variations in I_d during comparable constant-voltage (i.e. constant-V_{ds}) tests are also included in the table for comparison.

The table shows that the I_d variation is at least as small during constant-ratio as during constant-voltage stressing. The impact of drain current variation on lifetime is thus a less significant factor in constant-ratio stressing, underlining its suitability as a hot-carrier reliability test method.

The variations in V_{ds} and I_d during stressing at $I_{b(max)}$ are significantly less than those

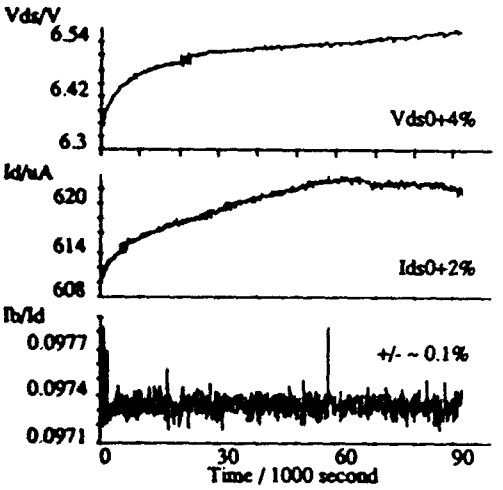


Figure 2: V_{ds}, I_d and $\frac{I_b}{I_d}$ during constant-ratio stress at V_{gs}=1.4V.

V_{gs} Condition	V_{ds} C.R.	I_d C.R.	I_d C.V.
$V_{g(low)}$	4.0%	2.7%	3.7%
$I_{b(max)}$	0.8%	0.6%	0.6%

Table 1: Variation of V_{ds} and I_d during hot-carrier stresses at constant-ratio (C.R.) and constant-voltage (C.V.) conditions.

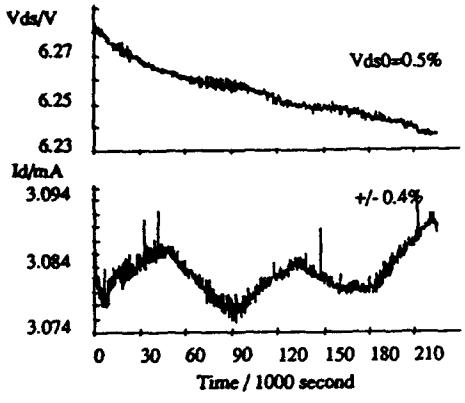


Figure 3: V_{ds} and I_d during constant-ratio stress at $I_{b(max)}$.

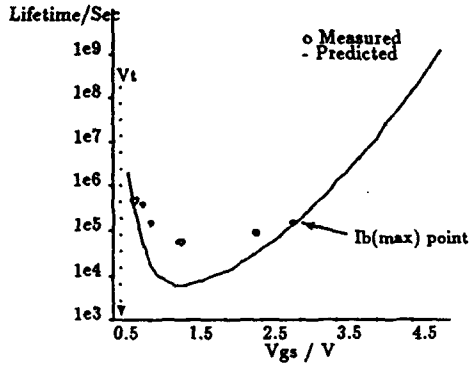


Figure 4: Measured device lifetimes superimposed on relevant Berkeley lifetime plot

at $V_{g(low)}$. This emphasises the desirability of performing stresses at the $I_{b(max)}$ condition during a hot-carrier characterisation of the Berkeley m and C parameters. (Although V_{ds} does not appear explicitly in eqn. 1, the smaller the V_{ds} change the more confidence there is that the nature of the degradation mechanisms does not change.)

The I_d variation during the constant-ratio test is very similar to that observed during a constant-voltage test with the same initial voltage conditions. In fact, constant-ratio and constant-voltage stresses at $I_{b(max)}$ are very similar (i.e. less than 1% difference) in all relevant respects, including device lifetime. Constant-voltage stressing at $I_{b(max)}$ can thus be used as an acceptable substitute for constant-ratio stressing as required.

6 STRESS LIFETIMES AT LOW V_{gs}

Hot-carrier constant-ratio stresses were performed at several V_{gs} points. The initial V_{ds} for each stress was that for which the Berkeley lifetime contour was derived. Measured device lifetime versus gate voltage is plotted in Fig. 4, superimposed on the Berkeley lifetime contour.

In the set of measured points, the minimum measured lifetime occurs at the same V_{gs} as the Berkeley minimum lifetime prediction.

7 COMPARISON OF BERKELEY PREDICTIONS AND LIFETIME MEASUREMENTS

The measured lifetimes show that the Berkeley equation does not model the V_{gs} dependence of lifetime well. Agreement at $I_{b(max)}$ is good, since the values of m and C in the Berkeley model were extracted at this condition. However, the minimum lifetime prediction is more than an order of magnitude too short.

The measured lifetimes lie along a smooth curve and tend towards the Berkeley prediction as V_{gs} tends to the threshold voltage of the device and as V_{gs} tends to the $I_{b(max)}$ con-

dition.

The pessimism of the model minimum lifetime prediction has been verified at less accelerated conditions on a separate set of W/L=25μm/1.5μm n-MOSFET's. A Berkeley lifetime contour at a certain V_{ds} for these devices predicted that minimum device lifetime at V_{gs}=1.3V would be 14% times as long as that at $I_b(max)$. In fact, a test of some forty days duration showed that the actual mean lifetime of a number of devices stressed with V_{gs}=1.3V was 87% of that of the mean $I_{b(max)}$ lifetime. This strongly supports the assertion that low V_{gs} lifetimes are pessimistic.

8 ANALYSIS OF RESULTS

The results show that, at accelerated stress conditions, a Berkeley model lifetime prediction at low gate voltage can be over an order of magnitude shorter than the actual measured lifetime at the same condition.

The prediction which appears, from the results, to be that most affected by the model inaccuracy is the minimum predicted lifetime at a given V_{ds}. This has implications for the use of the model in the field of industrial reliability standards. Attempts to bring minimum lifetime predictions above a desired value may lead to expensive over-engineering and/or the rejection of viable processes.

Conversely, if the widely-held assumption that the maximum reliability hazard exists at the $I_{b(max)}$ condition is made when using the Berkeley model, processes with substandard hot-carrier reliability may, incorrectly, be considered acceptable.

The suitability of the Berkeley model for use in transient circuit reliability simulation is also suspect, given its inaccuracy.

The inaccuracy of the model at low gate voltage precludes the possibility of low-V_{gs} stressing during hot-carrier reliability characterisations. Such a stressing condition may have appeared desirable since the minimum device lifetimes occur here. However, the results presented in this paper strongly suggest that all characterisation stresses should be performed at the $I_{b(max)}$ condition when using the Berkeley model.

Since the Berkeley model fails to model well the V_{gs} dependence of lifetime, a superior model could be expected to account for device gate voltage. Such models have recently appeared in the literature [6].

9 CONCLUSIONS

A method of hot-carrier stressing which maintains a constant ratio of substrate to drain currents has been successfully demonstrated. This allows test results to be used in an extraction of the Berkeley hot-carrier model parameters with minimum error.

Lifetime predictions made with the Berkeley model are seen to be pessimistic, typically by about an order of magnitude, at and about the gate voltage which gives the minimum lifetime prediction for a given V_{ds}.

In hot-carrier characterisations using the Berkeley model, hot-carrier stressing should be performed at the maximum substrate current condition.

Minimum device hot-carrier lifetime does not occur at the maximum substrate current condition, as has been reported [1]. This confirms more recent observations [6].

References

[1] C.Hu, S.Tam, F.Hsu, P.Ko, T.Chan, K.Terril, "Hot-Electron-Induced MOSFET Degradation - Model, Monitor and Improvement", *IEEE Trans, Electron Devices*, Vol.32, p.375, 1985

[2] F.Hsu, J.Hui, K.Chiu, "Hot-Electron Degradation in Submicron VLSI",*IEDM Tech Digest*, pp.48-51, 1985

[3] T.Tsuchiya, T.Kobayashi, S.Nakajima, "Hot-Carrier-Injected Oxide Region and

Hot-Electron Trapping as the Main Cause in Si nMOSFET Degradation", *IEEE Trans.Electron Devices*, Vol.34, p.386, 1987

[4] T.Poorter, P.Zoestbergen, "Hot-Carrier Effects in MOS Transistors", *IEDM*, pp.100, 1984

[5] M.Shimaya, S.Ogawa, N.Shiono, "A New Hot-Carrier Induced Degradation Mode Under Low Gate and Drain Bias Stresses in n-Channel MOSFET's", *Microelectronic Engineering*, pp.289, 1993.

[6] R.Woltjer, G.M.Paulzen, "Improved Prediction of Interface-Trap Generation in NMOST's", *IEEE Electron Device Letters*, Vol.15, No.1, p.4, 1994

[7] K.R.Hofmann, C.Werner, W.Weber, G.Dorda, "Hot-Electron and Hole-Emission Effects in Short n-Channel MOSFET's", *IEEE Trans.Electron Devices*,Vol.32,No.3,pp.691-698,1985

EVALUATION OF THE HOT-CARRIER-INDUCED OFFSET VOLTAGE OF DIFFERENTIAL PAIRS IN ANALOG CMOS CIRCUITS

R. Thewes[1,2], M. Kivi[3], K. Goser[2] and W. Weber[1]
[1] Siemens AG, Corp. Research and Development, ZFE BT ACM 23, D 81730 München, Germany
[2] Universität Dortmund, Lehrstuhl Bauelemente der Elektrotechnik, D 44227 Dortmund, Germany
[3] University of Liverpool, Department of Electrical Engineering and Electronics, Liverpool L69 4BX, U. K.

ABSTRACT

Using a specifically developed measurement setup and a test structure typical for analog applications, high precision measurements of the stress-induced offset voltage degradation of differential pairs are presented. Extrapolation to operating conditions yields valuable information for analog design in the sub-micron CMOS regime.

1. INTRODUCTION

As process development proceeds into the sub-micron regime the assurance of a reliable *digital* operation is no longer sufficient to guarantee stability of analog applications [1-4]. Investigations are demanded that account for the specific analog environment, such as use of specific analog circuit layouts, characterization of typical analog device and circuit parameters and application of typical analog operating points in the low gate voltage saturation regime for characterization and stress. In this work the degradation of the important offset voltage of p- and n-channel differential input stages is investigated which requires the detection of tiny changes in the matching of two transistors in the sub-mV regime.

2. EXPERIMENTAL SETUP

The test structure consists of two transistors designed for optimum matching as commonly used in analog applications (Fig. 1) which are implemented into the measurement setup as shown in the dashed rectangle in Fig. 2. Exactly as in real circuits the common source node of the differential pair is coupled to a current source I_{bias}. The gate of T_1 is connected to a fixed voltage, V_{G1}, while the gate voltage of T_2, V_{G2}, is varied around V_{G1} by using a precision summing amplifier and adding a tuneable voltage $V_{measure}$. For $V_{measure}$ varied over a wide range only a small difference voltage V_{G2}-V_{G1} results as the amplifier input gain for $V_{measure}$ is low (A<<1). Scanning $V_{measure}$ from a negative to a positive value, an equilibrium situation occurs eventually, with $I_1 = I_2$. Then V_{G2}-V_{G1} is equal to the input offset voltage which can be calculated easily from the known value of $V_{measure}$.

A sufficiently high sensitivity to detect the equilibrium case is obtained by a procedure as described in the following: a current source $I_4 = 0.5 \cdot I_{bias}$ is connected in parallel to voltage source V_{D2}. At equilibrium condition, current I_3 through V_{D2} changes polarity, i.e. measurement of I_3 in the region of interest can be performed with high resolution.

The same procedure as described for T_2 is repeated for T_1 by changing the two coupled switches. This compensates for matching tolerances of the current sources I_{bias} and I_4. A typical plot of current I_3 vs. input voltage difference is plotted in Fig. 3 where the cross-over of the two curves yields the offset voltage.

This setup allows direct measurement of the offset voltage to a precision of 10 µV without requiring adjustments of the external sources. With the same setup any dc electrical stress and the direct and complete characterization of all electrical transistor parameters is possible by programming current source I_{bias} as a voltage source for the source node of T_1 and T_2. In fact we determine linear and output current characteristics, the transconductance g_m and the differential output resistance r_{DS} both in saturation, and the threshold voltage both in the linear and the saturation regime.

3. RESULTS

Measured data of 2 µm p-MOSFETs from a 0.8 µm-5V-CMOS process are depicted in Figs. 4 - 6. Stresses are applied to one transistor of the pair only, thus including the worst case that occurs e.g. in comparators. In Fig. 4a the hot-carrier induced offset voltage shows changes of several mV after stress times of some 10^4 s, while the average initial offset voltage for virgin pairs is below 2 mV in these samples. Characterization of single transistor parameters shows negligible changes of the threshold voltage while the drain current shows a pronounced degradation effect (Fig. 4b). Note, that the operating points for drain current characterization correspond to the bias conditions during offset voltage measurement.

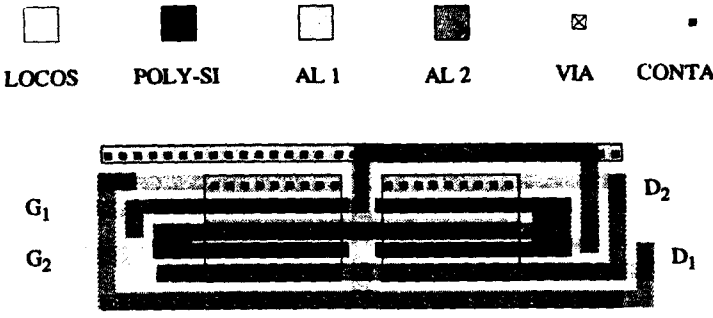


Fig. 1: Test structure of a differential pair designed for optimum matching as commonly used in analog applications. Each transistor is divided into two parts, which are arranged and coupled crosswise.

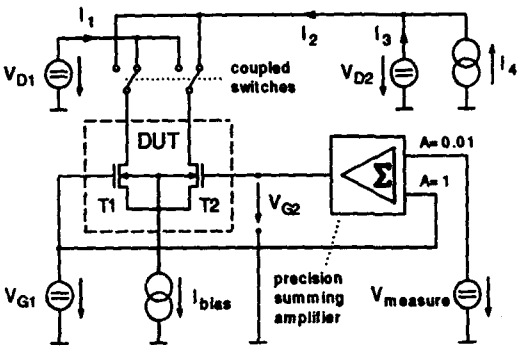


Fig. 2: Measurement setup. The test structure in Fig. 1 is emphasized by dashed lines.

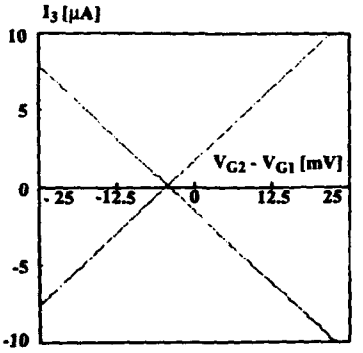


Fig. 3: Typical plot of current I_3 vs. voltage difference V_{G2}-V_{G1} to determine the offset voltage. I_3 is measured for both possible switch configurations as shown in Fig. 2. The cross-over yields the offset voltage.

The offset voltage degradation in Fig. 4a and the drain current degradation in Fig. 4b show a strong correlation. This correlation holds for different characterization points. As an example in Fig. 5 a higher $|I_{bias}|$ and a correspondingly higher $|V_{GS}|$ are compared. We obtain reduced offset voltage but nearly constant current degradations.

Further experimental p-MOS-data are shown in Fig. 6 where the stress drain voltage dependence is presented. Down to $|V_{D,stress}| = 7$ V data are available. The procedure for extrapolation to operating conditions is depicted in Fig. 7, where slope and intercept on the time axis define the degradation behavior (cf. Fig. 6) [5]. This leads to very high lifetimes at operating conditions in this specific process but other processes may exhibit more pronounced degradation effects.

In Figs. 8 and 9 data of 2 µm n-channel devices from a 1 µm-5V-CMOS process are presented. A typical hole stress condition is chosen in Fig. 8 while Fig. 9 presents data for a condition with maximum interface state generation. As in the case of p-MOS a correlation between offset voltage and drain current degradation *in the corresponding operating points* is found. Very little degradation of those parameters is found for the data in Fig. 9 leading to increased scatter on the logarithmic scale used.

Note, that the commonly used *linear mode drain current* (triangles, dashed lines) shows no correlation with the offset voltage: In the case of maximum linear mode drain current degradation (Fig. 9), only a weak offset voltage change is obtained, whereas in the case of relatively strong offset voltage degradation (Fig. 8) a small linear mode drain current degradation results (cf. [3]).

In general we observe smaller offset voltage changes and greater extrapolated lifetimes as compared to the p-channel case. In fact, a detailed study shows that p-channel device degradation is stronger in saturation than in linear mode, while in n-channel devices the inverse behaviour occurs [2,3]. Whereas usually

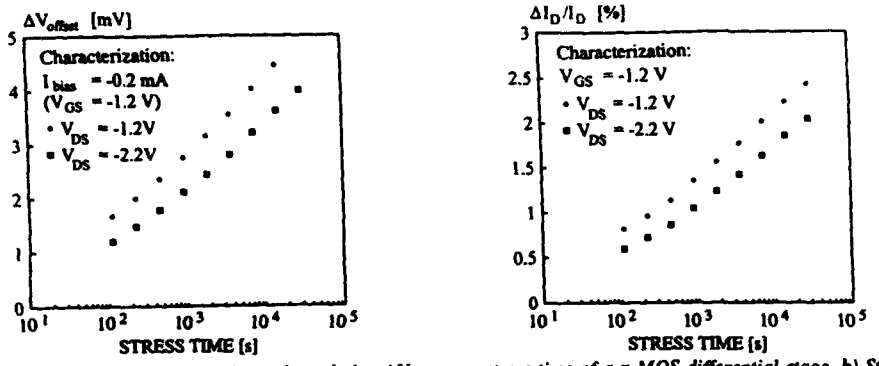


Fig. 4: a) Stress induced offset voltage degradation ΔV_{offset} vs. stress time of a p-MOS differential stage. b) Stress induced drain current degradation $\Delta I_D/I_D$ in the saturation regime vs. stress time. Stress conditions: $V_{G,stress}$ = -1.2 V, $V_{D,stress}$ = -8 V. Device: conventional drain profile, L = 2 µm. Process: 0.8 µm, 5 V, t_{ox} = 15 nm.

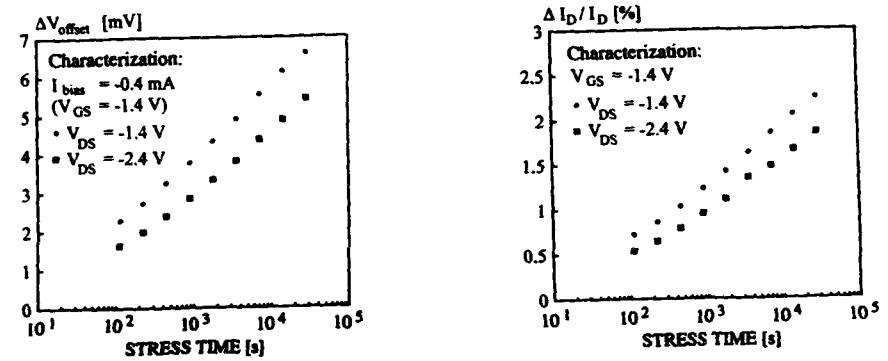


Fig. 5: Same parameters and stresses as in Fig. 4 but for different characterization operating points as given in the figure.

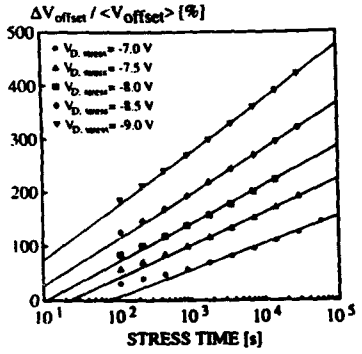


Fig. 6: Stress induced offset voltage degradation ΔV_{offset} / $\langle V_{offset} \rangle$ of p-MOS differential stages vs. stress time for different stress drain voltages and $V_{G,stress}$ = -1.2 V. $\langle V_{offset} \rangle$ stands for the average initial offset voltage and amounts to approximately 2 mV. Characterization with I_{bias} = -200 µA, V_{DS} = -1.2 V. Device data cf. Fig. 4.

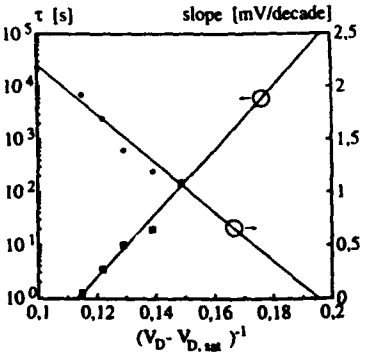


Fig. 7: Lifetime prediction: Stress drain voltage dependence of slope and intercept τ [5] of the offset voltage degradation of p-MOS differential stages. Characterization with I_{bias} = -200 µA, V_{DS} = -1.2 V. Device data cf. Fig. 4.

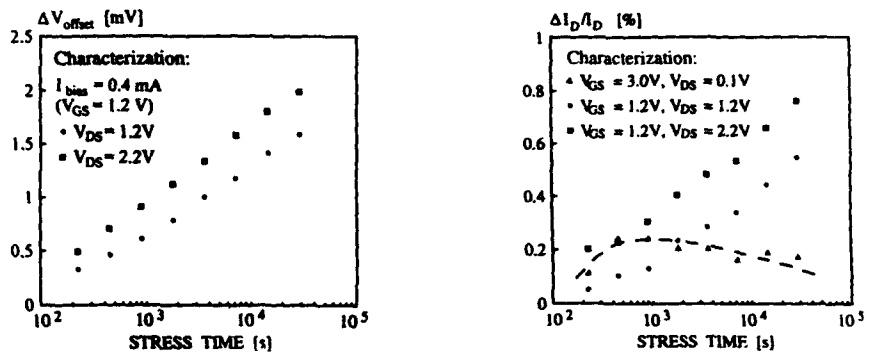


Fig. 8: a) Stress induced offset voltage degradation ΔV_{offset} vs. stress time of an n-MOS differential stage. b) Stress induced drain current degradation $\Delta I_D/I_D$ vs. stress time. Stress conditions: $V_{G,stress}$ = 0.9 V, $V_{D,stress}$ = 8 V (hole stress condition). Device: LDD, L = 2 µm. Process: 1 µm, 5 V, t_{ox} = 20 nm.

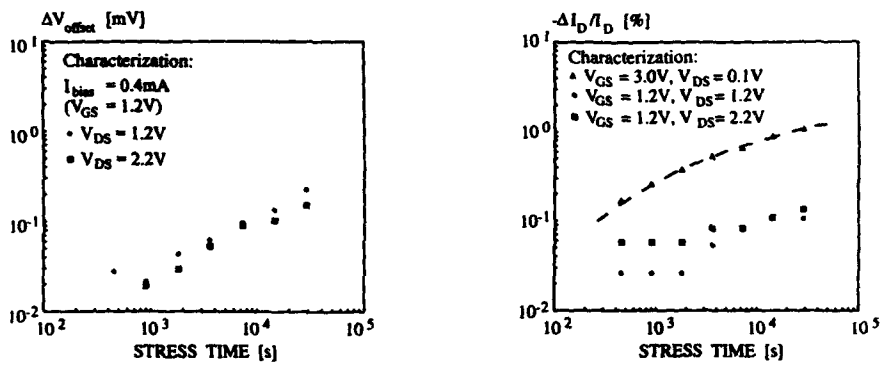


Fig. 9: a) Stress induced offset voltage degradation ΔV_{offset} vs. stress time of an n-MOS differential stage. b) Stress induced drain current degradation $\Delta I_D/I_D$ vs. stress time. Stress conditions: $V_{G,stress}$ = 3.5 V, $V_{D,stress}$ = 8 V (interface state generation stress condition). Device as in Fig. 8.

n-channel degradation is dominant, this fact can lead to a prevailing degradation of those circuit parameters that depend on p-MOS operation in the saturation region. An important one of these is indeed the p-MOS differential stage offset voltage.

4. CONCLUSIONS

In this contribution it has been proven that for analog applications a completely different approach is demanded than for the digital world. We realized an experimental environment relevant for differential stages by stressing an input pair of transistors with a typical analog layout. In an effort to characterize a parameter of relevance for analog applications the offset voltage was measured to high precision by a carefully chosen data acquisition method. A strong correlation was found between the stress-induced offset voltage and the drain current change in the corresponding operating points in p- and n-MOSFETs. However the linear mode parameter used in standard reliability tests showed uncorrelated behavior which is thus demonstrated to be inappropriate here.

REFERENCES

[1] J. E. Chung, K. N. Quader, C. G. Sodini, P.-K. Ko, and C. Hu, IEDM Tech. Dig. 1990, pp. 553-536

[2] R. Thewes, M.Brox, G. Tempel, W. Weber and K. Goser, IEDM Tech. Dig., 1992, pp. 531-534

[3] R. Thewes, W. Weber and K. Goser, Proc. ESSDERC 1993, pp. 85-88

[4] V.-H. Chan, B. W. Scharf, and J. E. Chung, Proc. CICC 1993, pp. 30.1.1-30.1.4

[5] M. Brox, E. Wohlrab, and W.Weber, IEDM Tech. Dig., 1991, pp. 525-528

NORMALISED 1/f NOISE: A MORE SENSITIVE DIAGNOSTIC TOOL FOR HOT-CARRIER DEGRADATION IN SUBMICRON MOSFET'S

Xiaosong Li and L.K.J. Vandamme
Electrical Engineering Dept., Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
Tel. #31-40-473242, FAX #31-40-448375

ABSTRACT

This study shows that 1/f noise measurements are a more sensitive diagnostic tool than dc measurements to detect hot-electron degradation in submicron n-channel MOSFETs.

1. INTRODUCTION

We report the hot-carrier degradation of n-channel MOSFETs using the dc characteristics and 1/f noise measurements. The damage caused by hot-carrier injection becomes visible as a reduction of the transconductance and/or a shift in the threshold voltage, in a decrease of the drain current and in a change of the 1/f noise. It seriously limits the reliability and lifetime of a MOSFET. All our results are obtained on wafer level. The n-MOSFETs used here are enhancement mode types with LDD junctions, made in a 0.5 μm technology by IMEC, Belgium. The gate area has an optical mask width W=100 μm and a length L varying from 0.45 μm to 10 μm. The gate oxide thickness is 11.3 nm. The dope concentration is $1.6*10^{18}$ cm^{-3} in the low-doped region and $1.4*10^{20}$ cm^{-3} in the high-doped drain and the source regions. The devices were stressed under the condition V_{GS}=2.0 V and V_{DS}=5.0 V, during one hour.

2. HOT-CARRIER EFFECTS ON DC

A group of devices with the same channel width but with different lengths was investigated. Each device has its independent gate, source, and drain terminal. The drain current I_{DS} as a function of the gate source voltage V_{GS} was measured in the ohmic region at V_{DS}=50 mV, before and after electrical stress. Since hot-carriers induce damage that is mainly located on the drain side, an interchange of drain and source (the so-called reverse mode (Refs. 1-3) changes the device characteristics. Hence, all characteristics of a post-stressed device are measured in the normal mode and in the reverse mode. To ensure that the difference between these two modes is due to stress, we always select a virgin device with identical I_{DS} vs V_{GS} and I_{DS} vs V_{DS} characteristics in the reverse mode and in the normal mode.

Fig. 1 shows I_{DS} vs V_{GS} obtained from a device with a channel length L=1 μm. Fig.2 illustrates the ratio I_{DSa}/I_{DSb} vs V_{GS} in the subthreshold region, where I_{DSa} is the drain current after stress and I_{DSb} the one before stress. Under the present stress condition for a 1 μm device and a 10 μm device, we have found a negative shift in the threshold voltage V_T. This implies that positive charges are trapped in the gate oxide (Refs. 4, 5). These excess positive charges may come from the holes, which are generated due to avalanche multiplication and are trapped during stress. The subthreshold gradient shows a change (see Fig. 2).

The drain current I_{DS} can be written as a function of the internal drain-source voltage V_{ds} for $V_{ds} \le V_G^* - I_{DS}R_{sS}$ (Ref. 6)

$$I_{DS} = \frac{\beta(V_G^* V_{ds} - \frac{1}{2}V_{ds}^2)}{1 + \beta R_{sS} V_{ds}} \qquad (1)$$

where

$$\beta = \frac{\mu_{00} W C_{ox}}{l(1 + \theta V_G^* + \theta_C V_{ds})} \qquad (2)$$

with θ and θ_c the mobility reduction

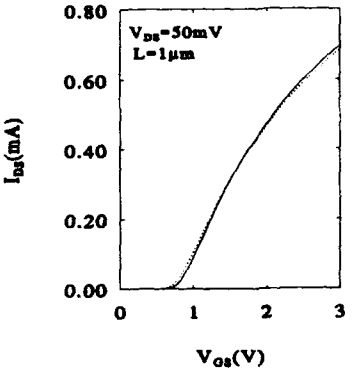


Fig. 1: The measured I_{DS} vs V_{GS} before (solid line) and after stress (dashed line). The normal mode and the reverse mode coincide.

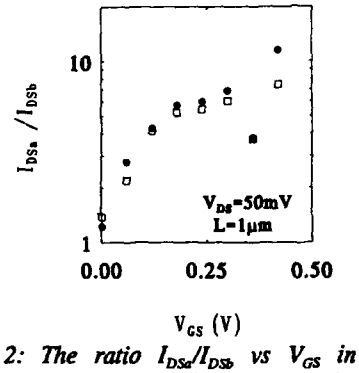


Fig. 2: The ratio I_{DSa}/I_{DSb} vs V_{GS} in the subthreshold region: • for the normal mode and □ for the reverse mode.

coefficients due to gate voltage and velocity saturation, respectively. The internal drain-source voltage V_{ds} is related to the external drain-source voltage V_{DS} by

$$V_{ds} = V_{DS} - I_{DS}(R_{sS} + R_{Dd}) \qquad (3)$$

with R_{sS} and R_{Dd} the series resistances on the source and drain, respectively.

In eqn. (1), the effective gate voltage V_G^* is kept at the same value before and after stress. The series resistance R_{sS} on the source side can be assumed to be independent of stress. So, the drain current I_{DS} will change with β and/or V_{ds}. The ratio I_{DSa}/I_{DSb} is then expected to be independent of V_{ds} and V_{DS} provided that $R_{sS}+R_{Dd}$ does not change after stress. As shown in Fig. 3, the observed ratio increases with V_{DS}, where the drain current I_{DSan} is for the normal mode and I_{DSar} for the reverse mode. This implies a decrease in the internal drain-source voltage V_{ds} due to stress. Therefore, in accordance with eqn. (3), there is an increase in R_{Dd}. Owing to the positive charges trapped in the gate oxide above the low-doped drain region, the electrons in the channel are more closely attracted to the interface than before stress. This has two consequences: the current path is more constricted and the local mobility becomes smaller due to surface scattering. This could be an explanation for the increase in the series resistance R_{Dd} on the drain side, associated with a negative shift in the threshold voltage.

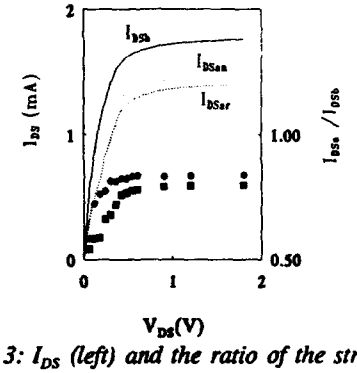


Fig. 3: I_{DS} (left) and the ratio of the stressed I_{DS} to the virgin I_{DS} vs V_{DS} (right, • normal mode and ▪ reverse mode) obtained at V_G^*=0.60V.

For comparison, the hot-carrier stress effects are also investigated for a MOSFET with a channel length L=10μm. A negative shift in the threshold voltage and an increase in R_{Dd} are observed too. However, unlike the short-channel device, the threshold voltage decrease in the reverse mode (0.12 V) is larger than that in the normal mode (0.03 V), and there is little change in the low field

mobility μ_{00}. This indicates that the hot-carrier damage is localised more on the drain side than for a short-channel device.

3.DEGRADATION INCREASES 1/f NOISE

The difference in 1/f noise before and after stress is more outspoken than dc characteristics as shown in Figs. 4 and 5 for L=1 μm and 10 μm, respectively.

For the same bias conditions, the reduction in the drain current varies from the normal mode to the reverse mode with channel length. It leads to increase or decrease the 1/f noise in a post-stressed device. However, when the 1/f noise is normalised for frequency *f*, *current* I_{DS}, and charge carrier number N, expressed by the α parameter for the ohmic region (Ref. 7)

$$\frac{S_I}{I_{DS}^2} = \frac{\alpha}{Nf} \quad , \tag{4}$$

or the relative noise fS_I/I_{DS}^2 for saturation, a systematic increase in α (see Fig. 6) or in fS_I/I_{DS}^2 (see Table 1) is observed after stress. The channel current always flows nearer to the interface on the source side than on the drain side. When the damaged part is on the source side, we expect a higher 1/f noise. So, a degraded device is often noisier in the reverse mode than in the normal mode. The degradation degree is determined by the current density and the electric field. These two quantities differ for devices with different geometry under the same bias conditions. Hence, the damaged part decreases with increasing channel length.

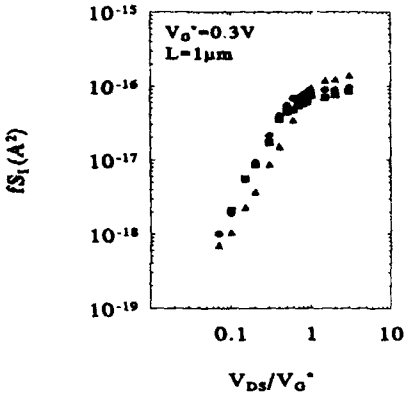


Fig. 4: The 1/f noise in the drain current vs V_{DS}/V_G^* (▲ virgin state, • after stress for the normal mode and ■ for the reverse mode).

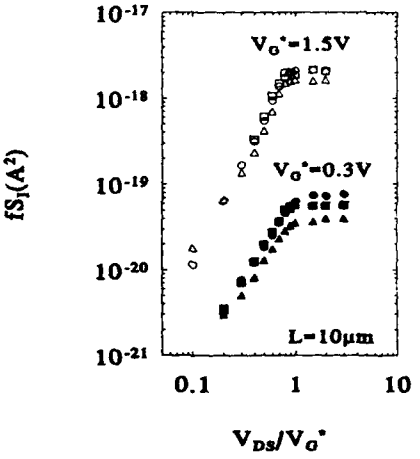


Fig. 5: The 1/f noise in the drain current vs V_{DS}/V_G^* (▲/△ virgin state, after stress: •/○ for the normal mode, ■/□ for the reverse mode).

4. CONCLUSIONS

The processes involved in bias-stress are complicated and can be summarised as follows:

(i) depending on the stress conditions, positive and negative shifts in threshold voltage are possible; we found negative shifts in both long and short devices,

(ii) owing to the nonuniform distribution of the trapped oxide charges in the gate oxide layer, the threshold voltage can have a nonuniform value along the channel,

(iii) the mobility reduction coefficient θ reduces for short channel and remains for long

Table 1. Change of fS_I/I_{DS}^2 in saturation in a MOSFET (L=10 μm) due to stress.

bias condition		normalized quantity fS_I/I_{DS}^2		
V_G^* (V)	V_{DS}	virgin state	stressed, normal mode	stressed, reverse
0.30	0.90	$1.4*10^{-11}$	$3.0*10^{-11}$	$6.1*10^{-11}$
0.60	1.80	$5.6*10^{-12}$	$1.2*10^{-11}$	$2.4*10^{-11}$
1.50	3.00	$1.7*10^{-12}$	$2.7*10^{-12}$	$3.3*10^{-12}$

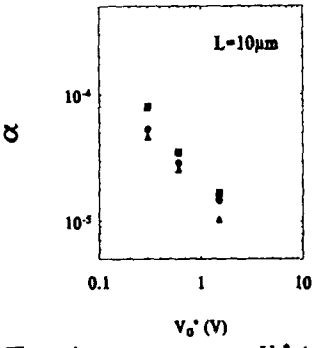


Fig. 6: The noise parameter α vs V_G^* (▲ virgin state, • after stress for the normal mode and ■ for the reverse mode).

channel. There is an increase in series resistance R_{Dd} on the drain side,
(iv) the position of the depth of the conducting channel on the drain side can shift with aging, so that a noisier or less noisy part of the semiconductor can be probed,
(v) in the ohmic region, the low field mobility μ_{00} only decreases 4%, the 1/f noise parameter α increases by about 100%. This makes a 1/f noise analysis a more sensitive tool than a mobility or a transconductance measurement.

5. ACKNOWLEDGEMENT

This research is supported by the European ADEQUAT project. We wish to thank Prof. H.E. Maes for helpful discussions and Dr. R. Bellens from IMEC Belgium for supplying the wafer and relevant information.

6. REFERENCES

1. C.M. Hu, S.C. Tam, F.-C. Hsu, P.K. Ko, T.Y. Chan and K.W. Terill, "Hot-electron-induced MOSFET degradation-model, monitor, and improvement", IEEE Trans. Electr. Devices ED-32, pp. 375-385, 1985.
2. Z.H. Fang, S. Cristoloveanu, and A. Chovet, "Analysis of hot-carrier induced aging from 1/f noise in short-channel MOSFETs," IEEE Electron Device Lett., vol. EDL-7, pp. 371-373, 1986.
3. M. Stegherr, "Flicker noise in hot electron degraded short channel MOSFETs," Solid-St. Electron., vol. 27, pp. 1055-1056, 1984.
4. X. Gao, S. Yee and H. Mollenkopf, "Positive charge in silicon dioxide due to high electric field FN injection," in: C.V. Thompson and J.R. Lloyd, ed., Mat. Res. Soc. Symp. Proc., vol. 265, pp. 249-254, 1992.
5. M.V. Fischetti, "Generation of positive charge in silicon dioxide during avalanche and tunnel electron injection," J. Appl. Phys., vol. 57, pp. 2860-2879, 1985.
6. Xiaosong Li and L.K.J. Vandamme, "An explanation of 1/f noise in LDD MOSFETs from the ohmic region to saturation," Solid-St. Electron., vol. 36, pp. 1515-1521, 1993.
7. F.N. Hooge, T.G.M. Kleinpenning and L.K.J. Vandamme, "Experimental studies on 1/f noise," Rep. Prog. Phys., vol. 44, pp. 479-532, 1981.

HOT CARRIER DEGRADATION IN SUBMICRON CMOS TECHNOLOGIES: A COMPARATIVE STUDY

J. Vandenbroeck (*), W. Kemp (**) and A. C. C. Swaen (***)

(*) Alcatel BTM
F. Wellesplein 1
B-2018 Antwerpen
Belgium

(**) Alcatel SEL
Lorenzstrasse 10
D-7000 Stuttgart
Germany

(***) Alcatel STR
Haarlerstrasse 20
CH- 8200 Zug
Switzerland

NMOS hot carrier degradation results of 17 submicron CMOS processes were compared. The dc lifetimes based on 10% gm degradation ranged from a few days to more than 100 years. Even on fast technologies, excellent lifetimes can be obtained.
Influence of the back-end processing and equipment on the hot carrier behaviour turned out to be enormous, overwhelming the effect of good drain engineering techniques.

INTRODUCTION

Performance and reliability are of utmost importance in modern telecommunications products. The ever advancing VLSI technology and the increasing number of components comprising the heart of today's systems places greater emphasis on reliability which means the individual FIT rate (Failures in Time) of the integrated circuits needs to decrease continuously in order to achieve this. careful monitoring of all degradation phenomena that might occur in present submicron technologies used in VLSI and ULSI circuits is therefore extremely important. One of these phenomena is hot carrier degradation in MOS transistors which is considered as one of the most important degrading mechanisms that can occur in submicron devices.
For a long time now a lot of studies have been attributed to this subject, and both dc and ac degradation phenomena seem to be well understood. Hot carrier reliability is expressed as a certain lifetime for obtaining 10% degradation of the most sensitive transistor parameter, generally the maximum linear transconductance ($g_{m,max}$). In the past, a transistor lifetime of 10 years could be achieved easily. These dc lifetimes were obtained on technologies with geometries in the range of 1.5 to 1µm by adapting the source and drain profiles of the nMOS transistors using Double Diffused Drain (DDD) and Lightly Doped Drain (LDD) profiles.
However, today there seems to be indications that this previously accepted dc lifetime (10 years for 10% shift in gm,max) is not always easy to obtain and that the total circuit lifetime with respect to hot carriers needs to be assessed using simulation and/or proper design techniques. This of course puts an extra stress on the designers of the telecommunication chips, since they now have to perform additional simulations, taking into account special design rules/guidelines etc. Moreover, in digital ICs based on standard cell designs, these cells are to be considered as "black boxes", leaving a large uncertainty concerning the reliability of the products under consideration.
To prove the validity of the statement with respect to the decreasing dc hot carrier lifetime of the advancing technologies, we investigated the hot carrier degradation results of seventeen different submicron CMOS processes (fifteen 5V processes and two 3.3 V processes), coming from 7 different suppliers. In this paper we will restrict ourselves to the results obtained on nMOS transistors. It will be shown that the influence of the processing can have an enormous impact on the hot carrier behaviour and that by taking the different processing effects into account, good dc lifetimes can be demonstrated.

RESULTS

All processes evaluated had a drawn gate length in the range of 1.0 to 0.6µm (with 5V operating voltage) and 0.5um (with 3.3V operating voltage). Their nMOS transistor current drive capabilities (IDSS drain current/um width @ VGS=VDS=5V) ranged between 0.3 and 0.5 mA/um, whereas the oxide thickness values ranged between 13 and 20 nm for the 5V processes, and 9 and 12 nm for the 3.3V processes.
All evaluated technologies had standard dry gate oxides, and used 2 or 3 levels of

metallisation.
Hot carrier degradation was performed on wide nMOS transistors with the minimum designed channel length. In general the corresponding effective electrical channel length (calculated from the maximum transconductance in the linear operating region of a long and wide device, and the device under test) was the typical one guaranteed by the process. On some technologies, results were also obtained on devices having the minimum electrical channel length as given by the process, by measuring on a deliberately shorter designed transistor (generally 0.1 µm less than the minimum allowed poly width (channel length) following the layout rules). Hot carrier results were always collected from fully processed, passivated wafers, by stressing the devices at close to the maximum substrate current. During stress gm, max (maximum gm defined as the maximum slope of the IDS-VGS curve of the linear operating region of the transistor at VDS=0.1V) was monitored as a function of time. Lifetime extrapolations for 10% gm degradation to both the nominal and the maximum supply voltage were obtained using the methods published in (1).

An overview of the results for the different technologies is presented in table I. Typical characteristics of the investigated technologies, such as the nominal feature size, the specified nominal current drive capabilities and the nominal gate oxide thickness values are shown as well. This provides an idea about the performance of the investigated technologies. It also indicates whether the electrical effective channel length of the transistors, on which the degradation was identified and for which the lifetimes were extrapolated, can be considered typical or worst case (minimum) for the technologies under consideration.
For the 5 V technologies, the results are shown in figures 1, 2 and 3 where the lifetimes obtained at the typical effective length, extrapolated to the nominal supply voltage, are plotted as a function of the nominal effective length for the different technologies (fig1), the specified gate oxide thickness values (fig2) and the specified current drive capabilities values IDSS (fig3).

DISCUSSION

The influence of the effective channel lengths for the same technology can be a factor of 2 (technology number 14) to 12 (technology number 7), whereas the influence of 10 % on the supply voltage can give a difference in lifetime between a factor of 3 (technology number 1) to more than 100 (technology numbers 2,3). These factors may seem rather exaggerated, but they are based on data provided by the suppliers, which could not always be verified by our own measurements.
Since most results were obtained on transitors with the nominal effective length we will restrict the rest of the discussion to this data. For comparative reasons the results extrapolated to the nominal supply voltage (5V for the 1.0 to the 0.6µm technologies and 3.3V for the 0.5µm technologies) will be used.
From the curves of figures 1 to 3 some general trends can be observed. The first observation is that the lifetimes decrease with decreasing effective channel length. However, technologies 1 and 3, having an effective nMOS transistor length of 0.5 and 0.45µm respectively have better dc lifetimes than many of the longer transistors in the 0.6 to 0.7µm range. In reference to the different 0.7µm transistors (technology numbers 7,9,11,12) a difference in lifetimes from 2 years (technology number 9) to 50 years (technology number 11) was also noted.
The influence of the oxide thickness, as observed from this study, is not so clear, but generally it can be expected that thin transistors with thinner gate oxides will have a shorter lifetime compared to transistors

having thicker gate oxides. This is somehow contradictory to previous studies (2) and (3), but it needs to be stressed that in these studies only the gate oxide thickness was changed, leaving the transistor structure the same. This, of course, was not the case in our study. Moreover the large scatter observed in our data indicates that the overall correlation is not good and that technologies with small and large gate oxide thickness values can show both good (technology numbers 3,10) and bad lifetimes (technology number 8).

The correlation between the hot carrier lifetime and the current drive capability of the nMOS transistors for the technologies under consideration, is very weak. A small trend showing increasing lifetimes for technologies with smaller drive currents can be observed. Again techology 3 is a clear exception to this.

We may thus conclude that for comparable technologies from the point of view of performance, an enormous difference in lifetimes is observed, ranging from a few days to more than 100 years.

Another important finding from this study is that almost identical technologies can have completely different results when they are run in a different wafer fab. This is the case for technologies 7, 8 and 9, which almost have the same process flow, but are processed at different locations.

These phenomena lead us to conclude that clearly the influence of the processing is much more important than generally thought of and that there can be a strong dependency on the process equipment itself.

This study shows that the processing steps which turned out to be extremely important, are those that could have an influence on the trapping behaviour and interface state generation of the gate oxides. In this respect we can name all steps associated with the metallisation, interlevel dielectric deposition, contact and via etching, annealing and passivation. This has already been reported in previous papers. For instance in (4) the influence of the passivation is discussed, showing that a plasma silicon oxide under the plasma nitride blocks the hydrogen from the nitride to penetrate in the gate oxide, limiting the excessive degradation due to the nitride passivation. References (5) and (6) discuss the enhanced hot carrier degradation due to the presence of plasma TEOS intermetal films and a possible improvement due to the addition of an ECR (Electron Cyclotron Resonance) - SiO2 film. In all cases moisture or hydrogen-related components were indicated to be the responsible factors for the enhanced degradation during the back-end wafer fab processing. This was again confirmed by (7) where the authors simulated the back-end process induced moisture penetration by performing a steam stress on wafers with one metal layer. A solution to this problem can be the employment of a silicon-rich intermetal oxide films (9) having higher dangling bond density than stoichiometric oxides.

In (8) the influence of plasma processing steps, possibly introducing radiation damage is discussed, with the possibility for improvement by annealing the damage out at a higher sinter temperature.

As a matter of fact, all of these factors have been considered to be relevant when analysing the technologies. Some technologies improved from a few days dc lifetime to more than 300 years after improving their interlevel metal oxides. On some technologies the impact of the plasma etching of the second metal layer has been found to be very strong. On other process technologies the anneal step of the barrier metal seemed to be important for having a good dc hot carrier behaviour.

The influence of the LDD structure used by most technologies (some technologies employed LATID (Large Tilt Angle Implanted Drain) but unfortunately these did not show a very good lifetime) turned out to be less significant than the influence of the processing steps previously mentioned.

CONCLUSION

Contrary to what is tended to be stated nowadays a good dc hot carrier behaviour (more than 10 years extrapolated lifetime for 10 % gm degradation at the nominal circuit supply voltage) can be obtained by taking into account the enormous influence that the wafer processing can have on the trapping properties of the gate oxides and interface states and subsequently the lifetime associated with hot carrier degradation. A careful study of the influence of the equipment during metallisation, as well as the optimisation of some key processing steps just before, during and after metallisation can result in a dramatic improvement of the hot carrier lifetime, even by several orders of magnitude.

ACKNOWLEDGEMENT

Alcatel would like to express its gratitude to the 7 companies which have been willing to communicate their results and related processing information to Alcatel. The authors would also like to thank B. Poppeliers of Alcatel Bell and S. Kingston of Alcatel SEL who have performed extensive hot carrier degradation experiments on the technologies discussed.

REFERENCES

(1) C.[illegible], S.Tam, F.C.Hsu, P.K.Ko, T. Y. Chan, K.W.Terrill, "Hot-Electron Induced MOSFET Degradation, Model, Monitor and Improvement", IEEE Trans. Electron Devices, Vol ED-32, No 2, 375(1985)

(2) M. Yoshida, D. Tohyama, K.Maeguchi, K. Kanzaki, "Increase of Resistance to Hot Carriers in Thin Gate Oxide MOSFETs", IEDM Tech. Dig., 1985, pp 254-257

(3) Y.Toyoshima, H.Iwai, F.Matsuoka, H.Hayashida, K.Maeguchi, K.Kanzaki, "Analysis on Gate Oxide Thickness Dependence of Hot Carrier Induced Degradation in Thin Gate Oxide nMOSFETs", IEEE Trans. Electron Devices, Vol ED-37, No 6, 1496(1990)

(4) S.Yoshia, K.Okuyama, F.Kanai, Y.Kawate, M. Motoyoshi, H.Katto, "Improvement of Endurance to Hot Carrier Degradation by Hydrogen Blocking p-SiO", IEDM Tech. Dig., 1988, pp 22-25

(5) M.T.Takagi,I. Yoshii, K.Hashimoto, "Characterisation of Hot-Carrier-Induced Degradation of MOSFETs Enhanced by H2O Diffusion for Multilevel Interconnection Processing", IEDM Tech. Dig., 1992, pp 703-706

(6) N. Shimoyama, K. Machida, J. Takahashi, K. Murase, K.Minegishi, T. Tsuchiya, " Enhanced Hot-Carrier Degradation Due to Water-Related Components in TEOS/ 03 Oxide and Water Blocking with ECR-SiO2 Film", IEEE Trans. Electron Devices, Vol ED-40, No 9, 1682(1993)

(7) Y.H. Lee, S. Hui, B. Sabi, "Evaluation of Moisture Induced Hot Carrier Degradation of n-and p-MOSFETs Using Steam Stress", IEDM Tech. Dig., 1993, pp 523-526

(8) M.L. Chen, C.W. Leung, W.T. Cochran, W. Juengling, C. Dziuba, T. Yang, "Supp[illegible]ion of Hot-Carrrier Effects in Subm[illegible]ometer CMOS Technology", IEEE Trans Electron Devices, Vol ED-35, No 12,10(1988)

(9) V.Jain, D.Pramanik, S.Nariani,C. Hu "Internal Passivation for Suppression of Device Instabilities Induced by Backend Processes", IRPS 92 proceedings, pp 11-15

5 V processes

Technology number	Feature size (μm)	Leff of stressed transistor (μm)	typ	min	tox typ (nm)	IDSS (μA/μm) VDD,typ. Leff,typ specified	Voltage Vdd,typ (V)	Vdd,max (V)	lifetime d = days y = years
1	0.6	0.5	X		15			5.5	30 y
1	0.6	0.5	X		15		5		80 y
2	0.8	0.55	X		16.5		5		200 d
2	0.8	0.55	X		16.5			5.5	1d
3	0.6	0.45	X		13	420	5		600 y
3	0.6	0.45	X		13	420		5.5	5 y
4	1	0.8	X		20	300	5		45 y
4	1	0.8	X		20	300		5.5	1 y
5	0.8	0.6	X		15	420	5		1 y
5	0.8	0.6	X		15	420		5.5	30 d
6	1	0.75	X		19	300		5.5	180 d
6	1	0.75	X		19	300	5		5 y
7	0.7	0.7	X		17.5	350		5.5	10 y
7	0.7	0.6		X	17.5	350		5.5	1 y
7	0.7	0.7	X		17.5	350	5		25 y
7	0.7	0.6		X	17.5	350	5		2 y
8	0.7	0.6	X		17.5	400	5		50 d
9	0.7	0.7	X		17.5	340	5		2 y
10	0.7	0.78	X		20	350	5		300 y
10	0.7	0.78	X		20	350		5.5	10 y
11	0.7	0.5		X	17	365		5.5	1 y
11	0.7	0.7	X		17	365		5.5	4 y
11	0.7	0.5		X	17	365	5		15 y
11	0.7	0.7	X		17	365	5		50 y
12	0.7	0.7	X		17	350	5		4 y
13	1	0.91	X		20	300	5		60 y
13	1	0.91	X		20	300		5.5	4.5 y
14	0.6	0.55	X		13.5	450	5		4.5 y
14	0.6	0.55	X		13.5	450		5.5	160 d
14	0.6	0.45		X	13.5	450	5		2.5 y
14	0.6	0.45		X	13.5	450		5.5	80 d
15	0.8	0.55		X	17.5	370		5.5	1 y
15	0.8	0.7	X		17.5	370		5.5	5 y

3.3 V processes

16	0.5	0.5	X		12	330	3.3		10 y
17	0.5	0.4	X		9	380		3.6	17 y
17	0.5	0.4	X		9	380	3.3		400 y

Table I: Overview of the extrapolated lifetimes obtained on the different technologies, as a function of the effective length of the stressed transistors and the voltage to which the lifetimes are extrapolated. Information about the technologies is given by their nominal feature size, oxuide thickness value and current drive capability of the nMOS transistor.

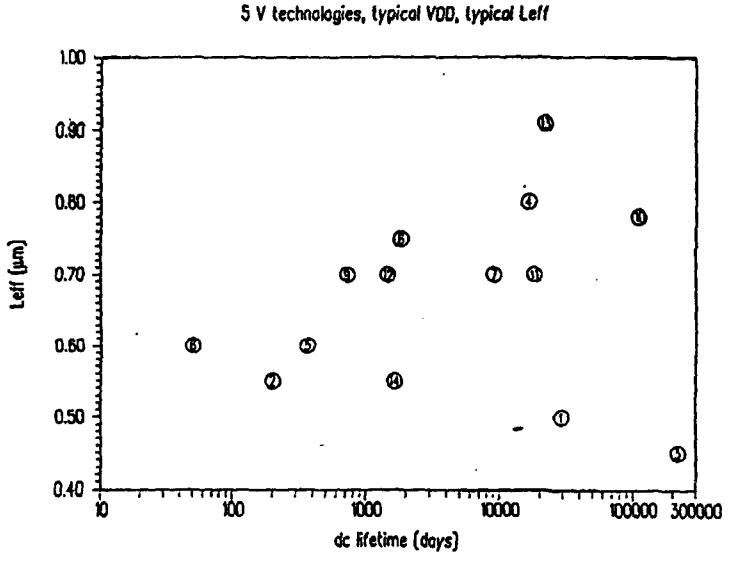


Fig. 1: DC hot carrier lifetimes, extrapolated to 5 V, obtained on the NMOS transistor with typical effective length. The results show the dependence of lifetime on the effective channel length.

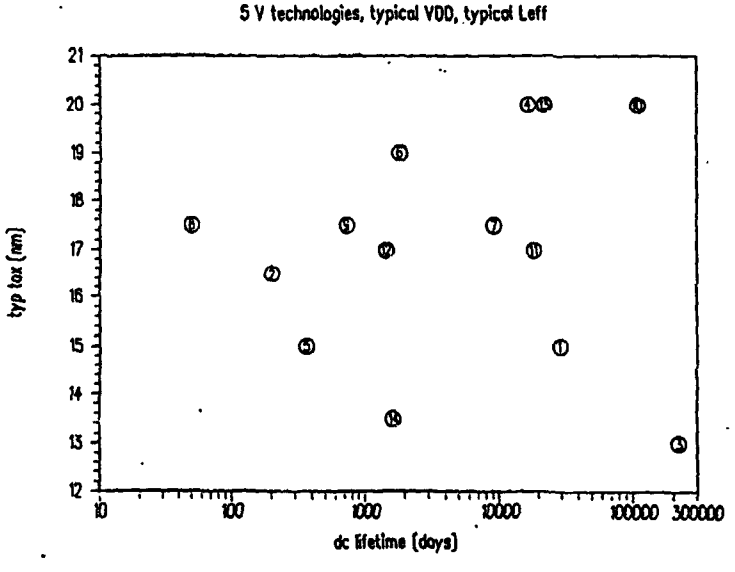


Fig. 2: DC hot carrier lifetimes, extrapolated to 5 V, obtained on the NMOS transistor with typical effective length. The results show the dependence of lifetime on the nominal gate oxide thickness.

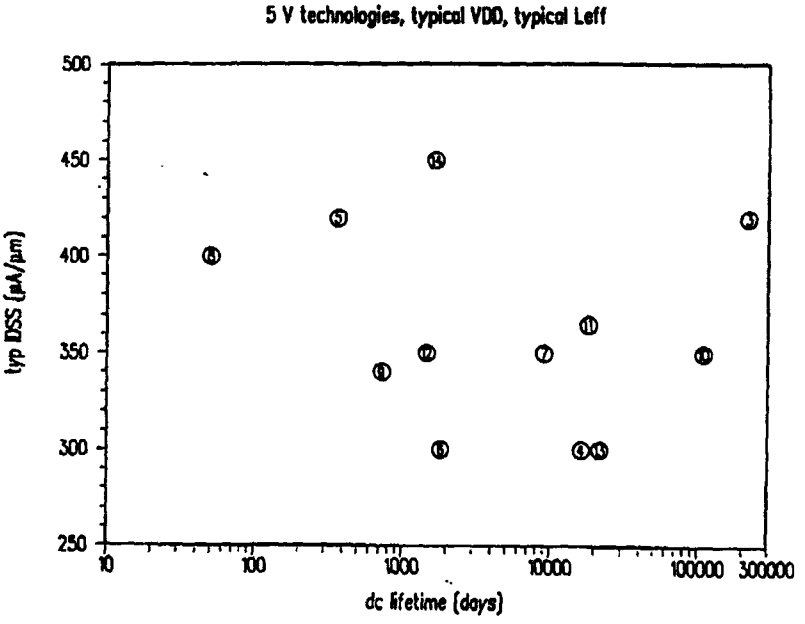


Fig. 3: DC hot carrier lifetimes, extrapolated to 5 V, obtained on the NMOS transistor with typical effective length. The results show the dependence of lifetime on the nominal NMOS IDSS current of the technologies.

DEFECT MONITORING AND LAYOUT RELATED YIELD AND RELIABILITY PREDICTION FOR VLSI INTERCONNECTS

A. Dreizner, H. Kück, K. Lukat, J. Pahlitzsch
Fraunhofer Institute of Microelectronic Circuits and Systems-IMS2-Dresden

ABSTRACT

Based on in-process defect monitor data and layout related Monte Carlo calculations we predict the yield of IC metal layers and the probability that missing metal defects reduce the IC reliability. The influence of such defects on electromigration´ failure distributions is investigated by experiments on submicron test structures.

1. INTRODUCTION

For VLSI circuits low ppm early failures are required.
Both yield and reliability are determined by process induced defects dominated by metallization defects (Ref.1). The relationship between yield and reliability and the influence of the defect density has been discussed in (Ref.2).
For the layout specific prediction of yield it is necessary *firstly* to know the defect density and the defect size distribution of the process and *secondly* to simulate the influence of randomly distributed defects on the electrical function of IC´s.
Yield modeling and prediction has been well documented in the literature (Refs. 3,4).
The tools described in this paper contain mask sets for defect monitor chips, programs for automatic chip testing, software for extraction of defect distribution parameters and software for Monte Carlo calculations in order to estimate the defect size depending failure probability of IC´s.

For state of the art VLSI metallization processes all early failures due to electromigration are expected to be defect related. These defects are due to missing metal which reduces the line width and causes increasing local current density.
This paper shows how the probability of missing metal defects can be predicted for IC layouts by embedding Monte Carlo simulations in the CADENCE design system.
In this context the influence of the defect induced reduction of line width on electromigration induced failures has to be known. Previous experimental work involving such defects has been carried out by Kemp at al. (Ref.5) who measured 3 µm wide stripes.

In chapter 4 test structures for electromigration tests of VLSI typical submicron lines with designed defects of 50 and 80% are described. First results of electromigration measurements will be available in June 1994.

2. DEFECT SIZE DISTRIBUTION

The test chip comprises 24 metal comb-meander-comb sub-chips for design rules with lines and spaces of 1.0/0.6µm, 1.4/1.0µm and 1.8/1.4µm with an area of 7.35 mm² each and we have fabricated test wafers using the test chip (Ref.6).
Using an automatic parametric test system HP 4062UX the resistivity of the metal lines is measured. In this way opens and shorts are detected. Because of the different geometries it is possible to classify the shorts and opens in 6 and 3 different size defect classes respectively.
Furthermore the measurement allows the determination of the chip coordinates of defects for defect analysis.
The number of defect sub-chips of class i failing by shorts as measured by the test system is given by

$$N_i = n\,D\int_{w_i}^{\infty} A_{crit,i}(x)S(x)dx \qquad (1)$$

where n is the total number of tested modules, x the defect size, S(x) the defect size distribution, D the defect density, w_i the space between the metal lines and $A_{crit,i}(x)$ the critical area. The sub-chip will fail, if the defect is centered in the critical area.
To determine S(x) and D we use the approximation (Ref.4)

$$S(x) = \frac{2(m-1)x_0^{(m-1)}}{(m+1)x^m} \qquad (2)$$

where x_0 is the fixed cutoff defect size. We calculate the parameters m and D by fitting the experimental data by (1) and (2).
Fig. 1. shows the product of defect density and defect size distribution depending on defect size for different lots of test wafers.

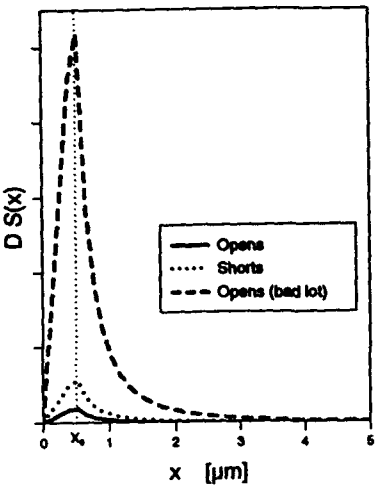


Fig.1 Defect size distribution for shorts and opens for "normal" and "bad" lots

3. PROBABILITY OF FAULT AND OF MISSING METAL DEFECTS

To determine the yield of good chips for a known defect distribution of the fabrication process it is necessary to analyse the layout of the chip. Analytical approaches to calculate the critical area from the layout data are proposed in (Ref.7).
In this paper we describe a new method which is based on statistical Monte Carlo analysis.

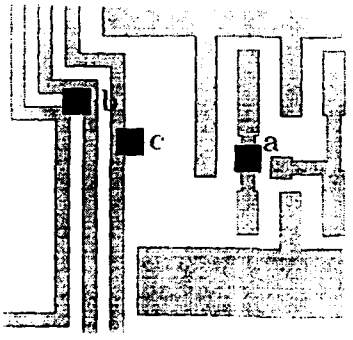


Fig.2 Typical generated defects in a processor layout a) open, b) short, c) missing metal

To calculate the probability K(x) which describes that a defect of size x will result in fault of the chip some thousands of metal layer defects are randomly generated in the chip area. The defects are assumed to have a square shape. For each generated defect it is checked whether the defect results in a fault or not. The ratio of defects resulting in a fault is K(x).
We use this method for shorts and opens and for the first time for missing metal defects.
In case of missing metal defect the amount of residual line width is estimated. If this width is more than δw we assume that no fault occurs. Opens are not counted as metal missing defects.
Fig. 2 shows a part of a processor layout designed in our institute with generated short, open and missing metal defect. The smallest lines and spaces are 2µm the chip area is nearly 1cm².
Fig. 3 shows the resultant K(x) curves for shorts, opens and missing metal defects. In our model calculations we assumed $\delta w = 1\mu m$.

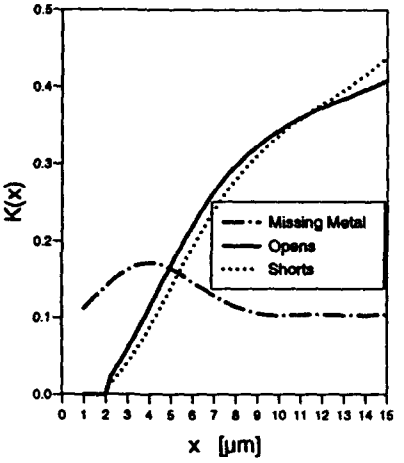


Fig. 3 Processor layout related fault probability for shorts, opens and missing metal defects

K(x) for missing metal begins to increase for $x=1\mu m$ when K(x) for opens and shorts is still zero.

The yield Y is related to the layout effective defect density D_{eff} and the chip area A:

$$Y=\exp(-D_{eff}A) \qquad (3)$$

with

$$D_{eff}= D\int_0^{\infty} S(x)K(x)dx \qquad (4)$$

Table 1 contains for the processor and several defect types the predicted data for Y and D_{eff}.
It can be seen that a low result in yield for opens (bad lot) correlates with a catastrophic high density of layout effective missing metal defects.
This fact ist caused by the more than proportional increase of S(x) in the x-range between 1 and 2µm and the non-zero contribution of K(x) for missing metal defects in this range.
For high defect densities there is a considerable probability that more than one electrical effective defect hits the chip. That's why in the case of the bad lots in contrast to equation (3) the binominal model was used to calculate the yield of defect free chips.
About 15% of the bad lot chips are affected with more than 10 electrical effective missing metal defects.

Table 1:

defect type	D_{eff} [cm^{-2}]	Y [%]
shorts	0.3	78
opens, normal lot	0.1	94
opens, bad lot	2.2	20
missing metal, normal lot	0.3	78
missing metal, bad lot	9.0	0.04

Based on data like shown in table 1 it is possible to discuss allowable yield boundaries in respect to ppm early failure goals.

4. MISSING METAL DEFECTS AND ELECTROMIGRATION BEHAVIOUR

Depending on the defect data of the fabrication process and IC integration level a definable number of metal missing "reliability defects" is expected.
It is to investigate which amount of relative line width reduction can cause electromigration induced early failures.
Lloyd (Ref.8) measured the failure distribution of 12,5µm wide stripes, Menon (Ref.5) of 3µm stripes of materials with different grain sizes. The results were different and depend on material, grain structure, line geometry and the amount of δw.
We designed test structures with VLSI typical lines in the range from 0.6 to 3.0 µm with 50 and 80% metal missing defects. Fig. 4 shows a typical example of a designed defect with line width of 1 µm and 50% missing metal.
In contrast to (Ref.5) also teststripes with more than one defect were designed. In this way there is a sufficient probability that defects hit grain boundaries or triple points. By comparing failure distributions for defect-free, one-defect and multi-defect test strips defect induced electromigration failure mechanisms can be analysed and we can prove which amount of line removal alters failure distributions.

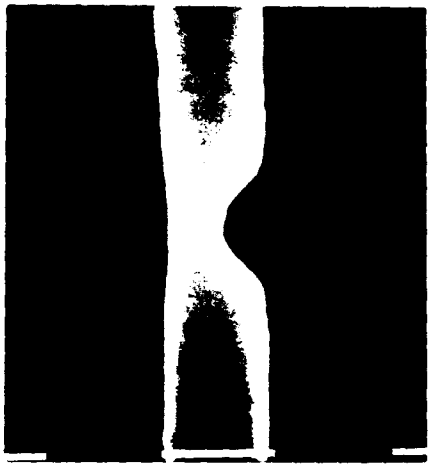


Fig.4 Typical designed defect, line width 1µm (white scale: 1µm)

It is reported by Hu (Ref.9) that circuit simulation allows to flag metal lines that pose reliability hazards due to high current densitiy.
The connection of such activities with the models and methods described in this paper would result in a nearly realistic simulation of the electromigration induced early failure problem.

5. CONCLUSIONS

Model calculations and electromigration measurements on stripes with missing metal reliability defects showed for IC metal layers a definable correlation between yield and early failure risk.
The proposed models and algorithms are essential first steps on the way to circuit simulation with respect to reliability and early failures of interconnects.

6. REFERENCES

1. H. Husten, C.P. Patrick, "Reliability defect detection, monitoring and screening during processing", Proceedings of ESREF 92, 1992, pp. 33-36.

3. A.V.Ferris-Prabhu,"Introduction to semiconductor device yield modeling", Artech House Bosten London , 1992.

4. Ch. H. Stapper , "The effects of wafer to wafer defect density variations on integrated circuit defect and fault distributions" IBM J. Res. Develop., Vol 28, No4, Juli 1984, pp. 461-474.

5. S.S Menon, A.K. Gorti and K.F. Poole, "Influence of grain size on defect-related early failures in VLSI Interconnects", Proc. IRPS, 1992 pp. 373-378.

6. A. Dreizner, H. Kück, K. Lukat, J. Pahlitzsch, "Schnelle Testverfahren zur Charakterisierung von Ausbeute und Zuverlässigkeit hochintegrierter Leitbahnsysteme", Proc. GME-Fachtagung ,1993 pp. 491-496.

7. A.R. Dalal, P.D. Franzon, M. J. Lorenzetti, "A layout-driven yield predictor and fault generator for VLSI", IEEE Trans. on Semiconductor Manufacturing , vol 6, No1, 1993.

8. J.R. Lloyd, P. M. Smith, G.S. Procob, "The role of metal and passivation defects in electromigration induced damage in thin film conductors, Thin Solid Films vol 93, 1982, pp. 385-395.

9. C. Hu, "IC reliability simulation", IEEE J. of Solid-State Circuits, vol. 22, No. 3, march 1992.

WRITE/ERASE DEGRADATION AND DISTURB EFFECTS IN SOURCE-SIDE INJECTION FLASH EEPROM DEVICES

D. Wellekens, J. Van Houdt, L. Faraone*, G. Groeseneken and H.E. Maes

IMEC, Kapeldreef 75, B-3001 Leuven, Belgium

* on leave from the University of Western Australia, Nedlands WA 6009, Australia

ABSTRACT

An in-depth analysis of the write/erase degradation of source-side injection Flash EEPROM devices is performed, which reveals two mechanisms underlying this degradation: a decrease of the charge per cycle on the floating gate, accompanied by the series effect of oxide and interface charges locally trapped above the channel. In addition, the main disturb effects are characterized and shown to be non-critical for reliable cell operation.

1. INTRODUCTION

Flash EEPROM devices which rely on source-side injection (SSI) for 5V-only programming have received a lot of interest during the past few years (Refs. 1-5). Most of these concepts use a triple-gate structure which exhibits an inherent immunity to soft-write and overerase effects (Refs. 1, 3, 5).

In this paper, two important reliability aspects are investigated on the High Injection MOS (HIMOS) device. (Refs. 3-6), which is a typical example of a SSI cell. Firstly, the different mechanisms contributing to the Write/Erase (W/E) degradation are studied by combining UV-erasure and IV-measurement with a charge pumping (CP) analysis. It is found that the degradation is due to the combination of a positive shift of the threshold voltage (V_t) both in the high and low V_t-state caused by oxide and interface charge at the split point, and a decrease of the charge injected per cycle onto the floating gate in both V_t-states. Secondly, the main disturb effects that can affect the margins for proper read-out of the cell, have been identified and characterized. These effects are shown to be well-controlled and have adequate margin.

2. DEVICE STRUCTURE AND OPERATION

As a typical example of a SSI device, the recently reported HIMOS cell has been studied. This device combines fast 5V-only programming features with a low development entry cost through the combination of a split-gate structure and a coupling capacitor which increases the channel hot-electron injection (CHEI) efficiency by several orders of magnitude (Ref. 4). Two perpendicular cross-sections of the HIMOS device are shown in Fig. 1.

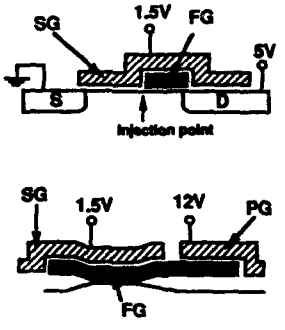


Fig. 1: Two perpendicular cross-sections of the HIMOS cell with typical programming voltages.

The memory cell is fabricated in a standard 0.7-μm double polysilicon process. The oxide under the floating gate (FG) is 8.5 nm thick and the interpoly dielectric layer is a thermally grown polyoxide with a thickness of 30 nm. The cell is programmed by enhanced CHEI at the split point and erased by Fowler-Nordheim (FN) tunneling of electrons from the FG towards the drain junction. The externally applied terminal voltages for each of the different operating modes are summarized in Table 1.

	source	drain	SG	PG
write	0	5	1.5	12
erase	-	5	0	-12
read	0	2	3	0

Table 1: Typical operating voltages for the HIMOS device.

In order to allow a proper investigation of the W/E-degradation, a device with a large capacitance between the FG and the program gate (PG) has been used (cell area = 50 μm^2). In this case the displaced charge per

cycle is much larger than for a smaller cell with the same V_t-window and thus the degradation is enhanced. An arbitrary V_t-window from -3V to +3V is chosen. The threshold voltage is defined at a fixed drain current of 1 μA, a drain bias (V_d) of 2V and select gate voltage (V_{sg}) of 3V. These conditions correspond to the read-out operation. Once the mechanisms leading to the W/E-degradation have been identified, a practical case will be considered using a cell that has been developed for use in a virtual ground array (VGA) architecture (Ref. 6). This configuration results in a cell that provides microsecond 5V-only programming in a 0.7-μm technology for a competitive cell area of 16.5 μm^2 and a PG voltage of only 12V. For this particular cell, the implications of repeated W/E-cycling on the position and magnitude of the V_t-window will be discussed. This cell has also been chosen for studying the main disturb mechanisms affecting the HIMOS cell.

3. WRITE/ERASE DEGRADATION

3.1 Endurance behaviour

The HIMOS cell is written by enhanced CHEI at the split point. This results in charge trapping and generation of interface traps (Ref. 4). Erasure is performed by FN-tunneling at the drain, which also has been known to result in charge trapping and creation of interface traps.

The standard procedure for degradation studies is an endurance measurement in which the high and low threshold voltages are monitored as a function of the number of W/E-cycles. However, such a measurement reflects the combined effect of all possible mechanisms contributing to the degradation. A typical endurance characteristic is shown in Fig. 2. As can be seen, the low threshold voltage (V_{tl}) shows a large increase with the number of W/E cycles suggesting a strong degradation of the read-out current, whereas the high threshold voltage (V_{th}) appears to be little affected by the repeated W/E-operations. In order to disclose the true nature of the degradation, two techniques have been used for separating the impact of possible damage within the cell channel and the impact of changes in FG charge.

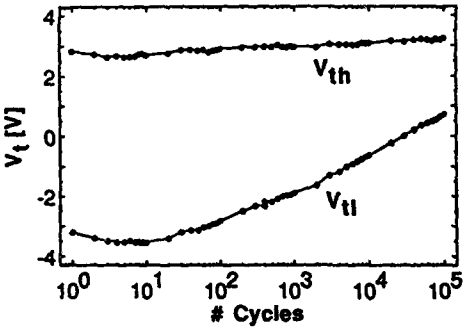


Fig. 2: HIMOS endurance characteristic under read-out conditions: a cell with a large coupling capacitor was used in order to accelerate the W/E-degradation.

3.2 Experimental procedures

One way of separating the different effects contributing to the W/E-degradation is by eliminating the charge on the floating gate. This can be done by illuminating the device with ultraviolet (UV) light, prior to and at different stages of cycling, in order to discharge the FG. This mechanism is self-limiting and the cell always ends up with the same reference charge on the FG. In this way, it is possible to reveal only the damage that may have occurred within the transistor channel.

A second powerful method is the charge pumping (CP) technique (Ref. 7). In a CP-measurement, the magnitude of the CP-current (I_{cp}) is proportional to the density of interface traps D_{it}, and the voltages corresponding to the transition edges of the CP-current are related to the threshold voltage and the flatband voltage. A CP-measurement is very useful, since it is sensitive to areal non-uniformities in both V_t and D_{it} (Ref. 7).

3.3 Analysis

Fig. 3 shows the I_d-V_{pg} characteristics after UV-erasure for different W/E-cycles. The bias conditions correspond to the read-out conditions, which were also used in the V_t-endurance measurement (Fig. 2). Three observations can be made from Fig. 3.

1) Cycling induces a very large positive V_t-shift after UV-erasure. This means that V_t of the FG channel has been locally increased by oxide and/or interface charges above the channel. These charges cause a series effect on the V_t-measurement. Since programming of the cell occurs at the split point, whereas erasure occurs at the drain outside the channel, the V_t-increase can be expected to be at the split point.

2) After cycling, a degradation of the subthreshold slope is observed. This subthreshold slope degradation is a measure of the increase in D_{it} and has to be accounted for when choosing V_{th} in order to minimize leakage through written cells after cycling.

3) The V_t-shift at 1μA observed in Fig. 3 is much smaller than the shift of V_{tl} in Fig. 2, which is measured at the same current level. This means that, besides the series effect, the charge per cycle that is removed from the FG diminishes with cycling. Likewise, in the written state V_{th} in Fig. 2 almost remains constant, while a large shift occurs after UV-erasure. This again indicates a decrease of the charge per

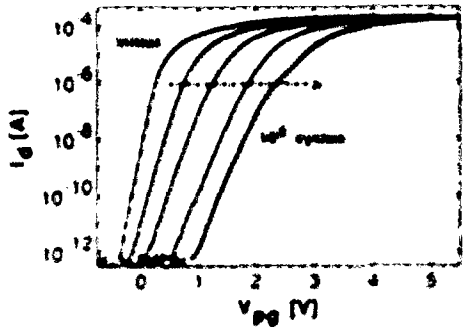


Fig. 3 IV characteristics of the cell after UV erasure at different stages of W/E cycling.

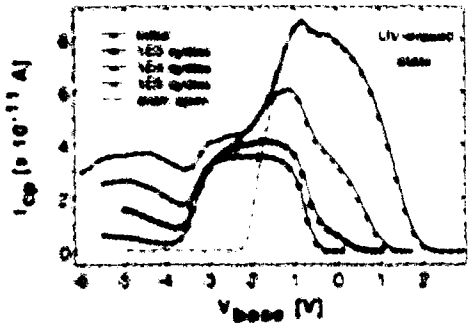


Fig. 4 Charge pumping characteristics of a cell in the UV erased state at different stages of W/E cycling.

cycle transferred to the FG. By comparing V_T at I_{DS}A in Fig. 3 to V_{CP} and V_{FB} in Fig. 2 at a particular number of cycles, the changes of the charge on the floating gate in the low and high V_T states can be derived. Instead, the V_T difference between the UV erased state and, for example, the written state is proportional to the charge on the FG in the written state. So, since the cell always returns to the same reference state after UV erasure, the change of this charge can be evaluated during cycling by measuring the difference in V_T between both states.

An alternative method to obtain similar information is provided by the CP technique. Since it has been shown previously that damage caused by the write operation is located entirely under the FG (Ref. 4), only that part of the channel was examined. Fig. 4 shows the CP characteristics in the UV erased state at the beginning of the W/E cycling and after 10^3, 10^4 and 10^5 cycles. Before W/E cycling, the rising edge of a CP curve is related to V_{FB} which is uniform along the FG channel. If only local damage occurs due to negative charge near the split point, this gives rise to a local increase of V_T. As a result, the interface traps in the damaged part of the channel are pumped at higher base levels, giving rise to a "shoulder" at these base levels. However, interface traps in the undamaged part of the channel are still pumped at the same base level. Therefore, after trapping of negative charge, the rising edge of the CP curve corresponds to V_T of the undamaged part of the channel. Based on these considerations, three main conclusions can be drawn from Fig. 4.

1) In the UV erased state, all rising edges of I_{CP} are at the same position. This demonstrates that after UV erasure, the device always ends up with the same reference charge on the FG.

2) The shoulder in I_{CP} at higher base levels is indicative of local electron trapping, while the increase of this shoulder is indicative of generation of additional interface traps. The dashed line in Fig. 4 represents the

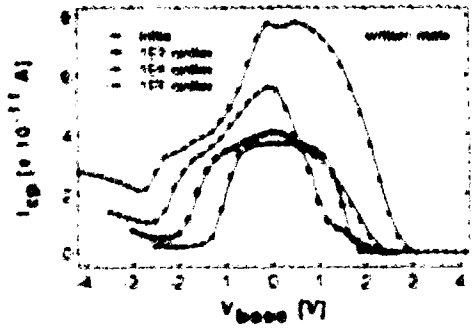


Fig. 5 Charge pumping characteristics of a cell in the written state at different stages of W/E cycling.

CP characteristics at 10^5 cycles with the drain contact open. In this case a shift of the rising edge to higher base levels is clearly observed. This proves that negative charge is present in the split point, which causes an additional potential barrier. Only when this barrier has been overcome, can the FG channel be flooded with electrons from the source. Therefore, the difference between the rising edge of the CP characteristic with source and drain connected and the CP characteristic with open drain is a measure of the series effect caused by damage in the split point.

3) An additional feature of the CP characteristics is the low voltage tail, which increases with the number of cycles. This tail was found to be related to the erase operation and may be due to charges generated by band to band tunneling occurring in the deep depletion layer during the erase operation.

On Fig. 5 the results are shown for the written state. In this case the rising edge of the characteristic shifts to

lower turn-levels as the number of cycles increases. Since this edge defines V_t of the undamaged part of the FG channel, the shift to lower levels indicates that less charge is injected since the floating gate as the number of cycles increases.

3.4 Discussion

Both from IV-measurements combined with UV-erasure, and from CP-measurements, it can be concluded that the degradation of SSI devices is caused by two main mechanisms:

1) During cycling, trapping of negative charge and generation of interface traps occurs at the split point. These charges cause a series effect and shift the threshold V_t (FG charge = 0) of the FG channel to the positive direction.

2) Besides this series effect, there is a reduction in the amount of charge per cycle transferred to and from the FG as the number of cycles increases.

Therefore, the change of V_t in the high and low threshold cases can be expressed as

$$\Delta V_{tl} = \Delta V_{tl}(Q_{fg}) + \Delta V_t (Q_{sp}) \qquad (1)$$

$$\Delta V_{th} = \Delta V_{th} (Q_{fg}) + \Delta V_{th} (Q_{sp}) \qquad (2)$$

where ΔV_{tl} and ΔV_{th} are the changes in threshold voltages in the low and high threshold cases, $\Delta V_t (Q_{fg})$ is the change of V_t due to changes in FG charge, and $\Delta V_t (Q_{sp})$ is the change of V_t due to changes above the channel in the split point. In fact, $\Delta V_t (Q_{sp})$ is the change of V_t which is seen after UV-erasure. Of course, $\Delta V_{tl} (Q_{sp}) = \Delta V_{th} (Q_{sp})$ and will simply be indicated as $\Delta V_t (Q_{sp})$.

$\Delta V_{tl} (Q_{fg})$, $\Delta V_{th} (Q_{fg})$ and $\Delta V_t (Q_{sp})$ were derived both from IV and CP-measurements and are shown in

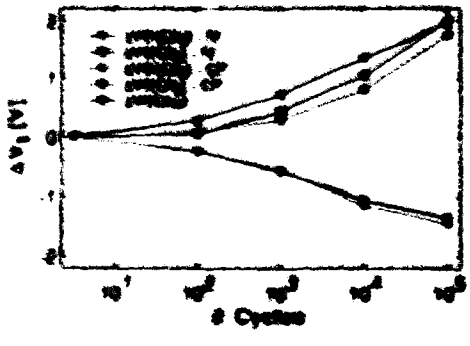


Fig. 6 Contributions to the degradation of floating gate charge and split point charge

Fig. 6. As can be seen, $\Delta V_{th} (Q_{fg})$ obtained from both techniques is identical, while the values of $\Delta V_{tl} (Q_{fg})$ agree to within 10%. However, as far as $\Delta V_t (Q_{sp})$ is concerned, only the value obtained from IV measurements is relevant when correlating $\Delta V_t (Q_{fg})$ and $\Delta V_t (Q_{sp})$ to the endurance characteristic. This is because $\Delta V_t (Q_{sp})$ obtained from IV-measurements has to be considered at 1µA in order to correlate with the endurance measurement. At this current level, $\Delta V_t (Q_{sp})$ is caused by both fixed oxide charge and interface traps. However, $\Delta V_t (Q_{sp})$ obtained from CP represents only the influence of fixed oxide charge and is thus smaller.

As Fig. 6 shows, in the low state $\Delta V_{tl} (Q_{fg})$ and $\Delta V_t (Q_{sp})$ are both positive, as both effects accumulate causing the rapid increase of V_{tl} with cycling. During the erase operation, negative charges are trapped above the drain. These charges give rise to a reduction of the injection current during erasure and thus decrease the amount of negative charge removed from the FG. This causes the positive shift of $\Delta V_{tl} (Q_{fg})$.

During the write operation, trapped negative charges near the split point cause a decrease of the injection current and therefore a decrease of the charge per cycle injected onto the FG. This results in a negative contribution of $\Delta V_{th} (Q_{fg})$. As can be seen in Fig. 6, in the high V_t-state, the series effect caused by charges at the split point almost exactly compensates for the decrease of the injection current due to fewer charges.

As these effects remain completely hidden when only the V_t-endurance characteristic is considered. The above results indicate that the classical endurance characteristic (Fig. 2) is an unreliable monitor for the degradation of SSI devices, since it provides erroneous information on the nature and the magnitude of the degradation.

3.5 Small area FGA cell

Based on the analysis made in the previous discussion, we can now make some conclusions on the choice of the V_t-window for optimal endurance behaviour. For this optimization a FGA cell with an area of 16.5 μm^2 was subjected to a W/E-degradation study, including several different initial V_t-windows varying both in magnitude and position.

a) For circuit applications, the current in the read-out conditions is more relevant than V_t. Fig. 7 shows the read-out currents corresponding to the different windows. The read-out conditions are also indicated: V_g = 2V, V_{dd} = 3V.

These read-out conditions yield an initial cell current of about 300 µA, depending on the particular value of V_{tl}. As the figure demonstrates, the degradation of the read-out current only depends on the initial value of V_{tl} and

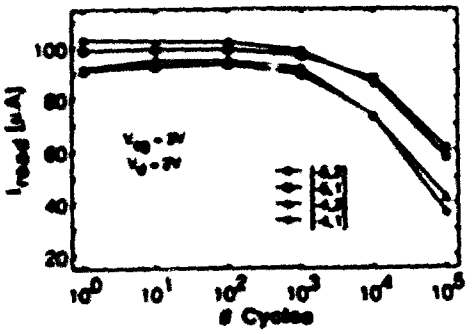


Fig. 7 Read-out current versus number of WE-cycles for a small VGA cell and different initial threshold voltage windows.

is very sensitive to the magnitude of the V_t window. For an initial V_g = - 5V, the current decreases from 100 µA to 60 µA after 10^5 cycles, which is still sufficient for proper operation. Note that up to 1,000 cycles, the read-out current remains almost constant. Due to the presence of a series select gate, the drain current saturates at PG voltages well above V_t. Therefore, an increase of the low V_t does not result in a corresponding decrease of the read-out current. Fig. 7 shows that after 10^5 cycles, the read-out current only decreases by 40%, while the margin between V_g and the read-out level (V_{PG}=0V) decreases by 75%. This situation is very different from that in stacked-gate devices where an increase of V_g is directly proportional to a decrease of the read-out current.

b) When defining the position of V_{th}, two effects should be taken into account. Firstly, the subthreshold slope degradation has to be considered in order to prevent leakage through written cells. Secondly, the charges trapped above the channel should be accounted for, because it has been observed experimentally that the damage at the split point can be completely annealed out by a short high temperature bake. However, in practice the impact of trapped split point charge will be less severe than in a degradation measurement, because annealing will occur simultaneously with cycling.

When taking all effects into account, a high threshold voltage of 2V at the read-out conditions was found to be adequate for this particular cell to guarantee 10^5 WE-cycles.

4. DISTURB EFFECTS

When considering the reliability issues of Flash EEPROMs, a study of the disturb mechanisms is also of great importance, since they cause unwanted changes in the threshold voltage of a particular cell. The main disturb mechanisms, that are active in the case of the HIMOS cell, are discussed in the next sections.

4.1 Program gate disturb

During programming of the HIMOS cell, the PG is pulsed to 12V. In an array, any erased cells on the same wordline as the cell which is being programmed see a large electric field across the tunnel oxide and thus exhibit a soft-write effect due to FN-tunneling from the substrate towards the FG. Fig. 8 shows the V_t-shift as a function of time for an initial cell V_t of - 5V. Also shown is the forward programming characteristic. It is estimated from this figure that this soft programming is more than 7 orders of magnitude slower than the fast SSI programming. This is a consequence of the small PG coupling factor (50%) which suffices for efficient programming. Because the negative-gate-bias-erase scheme implies a wordline-oriented sector approach, the maximum disturb time is the forward programming time multiplied by the number of cells on the same wordline, which is in the order of 10ms. As can be seen on Fig. 8, this leaves enough read-out margin with respect to the erased threshold voltage.

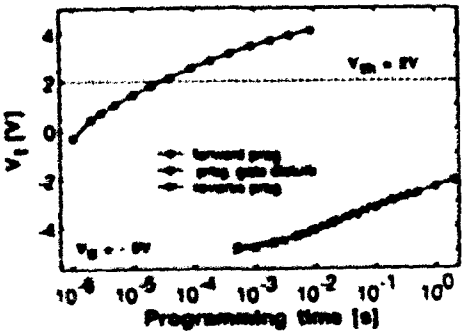


Fig. 8 Forward/reverse programming characteristics (V_g=5V) and program gate disturb characteristic (V_g=0V).

4.2 Reverse programming

Because of the VGA configuration, the bit lines are shared between neighbouring cells. Therefore, the reverse programming of a cell needs to be considered as well. In other words, if a particular cell is being programmed, the neighbouring cell sharing the same bit line will also draw current, but with the current flowing from the source of the split-gate cell to the drain. This reverse programming characteristic is also shown in Fig. 8, and is found to coincide with the gate disturb data. Reverse programming is thus entirely due to the

same mechanism (FN-tunneling from substrate to FG) and can again be neglected in a large array.

4.3 Drain disturb

During programming and read-out, any written cells sharing the same bit line as the cell being programmed or read are subject to an electric field at the FG-to-drain overlap. This can result in charge loss due to FN-tunneling from the FG towards the drain. Fig. 9 presents data for a cell in the high V_t-state and shows the stress time required to induce a V_t-shift of 100mV as a function of the inverse of the drain voltage. A threshold voltage of 2V has to be considered in order to meet the specifications induced by the endurance, as explained in a previous section. The worst case condition occurs for a cell that is not selected but which shares its bit line with a cell being read or programmed. Such unselected cells have a grounded SG, and thus no voltage is coupled from the SG to the FG. Under these worst-case conditions, extrapolation towards 10 years indicates that the drain read-out voltage can be as high as 3.2V. This higher immunity to drain disturb as compared to stacked gate cells is due to the much smaller excess negative charge on the FG in the written state.

When a cell is on the same bit line as a cell being programmed, it sees 5V at its drain. Due to the sector erase feature, the drain disturb time during 5V-programming is defined as the product of programming time, number of cycles, and number of wordlines, which is typically in the order of 2000s. This was found to correspond to a V_t-shift of 500mV, which is still reasonably small.

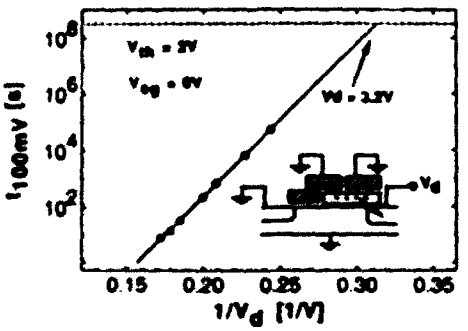


Fig. 9 Worst case drain disturb characteristic for a written cell.

4.4 Select gate disturb

Due to the small coupling ratio between SG and FG, a large electric field is established in the oxide between SG and FG during read-out conditions. This causes a soft-erase effect on a written cell via FN-tunneling through the interlevel polyoxide between FG and SG. Fig. 10 shows the SG disturb characteristic for a high threshold voltage of 2V. The data predicts that a SG voltage as high as 4.3V is necessary to induce a V_t-shift of only 100mV after 10 years for a high threshold state of 2V. Therefore, because of the soft-write immunity and the small drain disturbance (Fig. 9), a drain voltage of 2V can be used during read-out, while the SG voltage only needs to be 3V in order to ensure a read-out current of 100 µA. A large margin with respect to the drain and SG disturb effects is thus obtained.

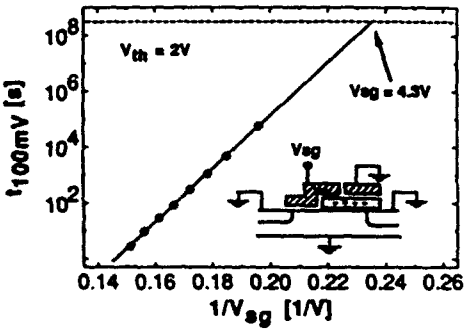


Fig. 10: Worst case select gate disturb characteristics for a written cell.

5. CONCLUSIONS

In this paper, it has been shown that the classical endurance characteristic (threshold voltage versus number of cycles) is not sufficient in order to characterize the degradation of source-side injection devices after write/erase-cycling. This is a consequence of: (1) the charge on the floating gate not being proportional to the threshold voltage due to the series effect of the localised degraded region; (2) the read-out current not being proportional to the difference between the gate read-out voltage and the erased threshold voltage due to the presence of a series select gate. Furthermore, the leakage current through written cells needs to be accounted for when choosing the high threshold state, by providing an additional margin for subthreshold degradation after write/erase-cycling. Gate and drain induced program disturb mechanisms in HIMOS devices are shown to be well controlled and have adequate margin. A novel disturb mechanism, soft-erase towards the select gate, has been found to be non-critical for normal device operation.

6. ACKNOWLEDGMENTS

This work has been carried out under the ESPRIT-APBB project(contracts ESPRIT-2039 and ESPRIT-7802).

The authors would like to thank L. Haspeslagh, F. Vleugels, M. Van Dievel, G. Van Horebeek and L. Deferm, all from Imec, for the processing of the memory cells.

7. REFERENCES

1. Naruke K., Yamada S., Obi E., Taguchi S. and Wada M. "A New Flash-Erase EEPROM Cell with a Sidewall Select-Gate on its Source Side", IEDM Tech. Dig., 1989, pp 603-606.

2. Yamauchi Y., Tanaka K., Shibayama H. and Miyake R. "A 5-V-Only Virtual Ground Flash Cell with an Auxiliary Gate for High Density and High Speed Application", IEDM Tech. Dig., 1991, pp 319-322.

3. Van Houdt J., Haspeslagh L., Wellekens D., Deferm L., Groeseneken G. and Maes H.E. "HIMOS - a High Efficiency Flash EEPROM Cell for Embedded Memory Applications", , IEEE Trans. Electron Devices, vol. ED-40, 1993, pp 2255-2263.

4. Van Houdt J., Wellekens D., Groeseneken G., Deferm L. and Maes H.E. "The High Injection MOS Cell: a Novel 5V-Only Flash EEPROM Concept with a 1μs Programming Time", Proc. ESSDERC, 1991, pp 617-620.

5. Van Houdt J., Heremans P., Deferm L., Groeseneken G. and Maes H.E. "Analysis of the Enhanced Hot-electron Injection in Split-Gate Transistors Useful for EEPROM Applications", IEEE Trans. Electron Devices, vol. ED-39, 1992, pp 1150-1156.

6. Van Houdt J., Wellekens D., Haspeslagh L., Deferm L., Groeseneken G and Maes H.E. "Optimization of a Submicron HIMOS FLash EEPROM Cell for Implementation in a Virtual Ground Array Configuration", Proc. ESSDERC, 1993, pp 381-384.

7. Heremans P., Witters J., Groeseneken G. and Maes H.E. "Analysis of the Charge Pumping Technique and its Application for the Evaluation of MOSFET Degradation", IEEE Trans. Electron Devices, vol. ED-36, 1989, pp 1318-1335.

8. Bergemont A., Haggag H., Anderson L., Shacham E. and Wolstenholme G. "Nor Virtual Ground (NVG) - A New Scaling Concept for Very High Density Flash EEPROM and its Implementation in a 0.5μm Process", IEDM Tech. Dig., 1993, pp 15-18.

9. Melanotte M., Bez R and Crisenza G. "Nonvolatile Memories - Status and Emerging Trends", Proc. ESSDERC, 1991, pp 603-612.

MODELLING OF THE THRESHOLD VOLTAGE - CURRENT GAIN RELATION IN A DEGRADED I^2L TEST CELL. APPLICATION FOR IMPROVING THE RELIABILITY OF THE TECHNOLOGICAL PROCESS

*Bonnaud Olivier, *Seidou MAIGA Amadou, *+Rolland Loïc, +TaurinMichel

*Groupe de Microélectronique et Visualisation, URA CNRS 1648
C.C.M.O., Université de Rennes I, Campus de Beaulieu 35042 RENNES Cedex
+SGS Thomson Microelectronics, Bd de Volclair, 35000 RENNES

ABSTRACT

After shown that the origin of the degradation of PNP lateral transistors of a I2L test cells is negative trapped charges in gate oxide induce by plasma etching, we have developed a modelling of the threshold voltage of the PMOS transistor associated to the PNP one which shows that the degradation can only be explained by taking in to account a location of trapped charges both close to the drain and to the source. In the present work, from a modelling of the relation between the current gain of the lateral PNP bipolar transistor and the threshold voltage of the associated PMOS, we have defined a criteria which allows to improve the reliabilty of the I2L process. Finally, we expose some solutions for the improving of technological process.

1- INTRODUCTION

To analyse the quality of a process involved in I^2L technology, it is usual to insert in the test cells lateral bipolar transistors the base of which is overlayed by thermal and deposited oxides and by the emitter metallization. This configuration creates a parasitic PMOSFET, as shown in figure 1, which allows to characterize the PNP lateral bipolar transistor.

The analysis of the electrical characteristics of some 500 bipolar transistors and their parasitic MOSFET have shown a correlation between the MOS threshold voltage (V_{TH}) and the bipolar current gain (H_{FE}). In previous works [1], the authors explained this variation by the effect of negative oxide charges trapped at deposited oxide/thermal oxide interface which induce inversion mode in the base layer (or MOSFET channel) and reduce its electrical width. They have also shown from electrical test during the process and from literature [2-4] that charges are introduced during the nitride plasma etching. From these results, they proposed a modelling of the threshold voltage to explain the degradation; they showed that only a symetrical location of oxide trapped charges close to the emitter and close to the collector allows to explain the behaviour of the structures [2]. In this case, the threshold voltage variation is expressed by the following formula (1):

$$\Delta V_{th}=\frac{q.N_{ss}}{C_{ox}}\left(1+\sqrt{1-\frac{\ln(R_2-L)/R_1}{\ln(R_2/R_1)}}-\sqrt{1-\frac{\ln(R_1+L)/R_1}{\ln(R_2/R_1)}}\right) \quad (1)$$

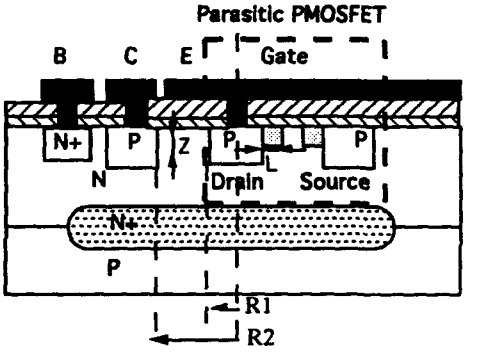


Figure 1. *Schematic cross-section of the structure. The geometry is cylindrical.*

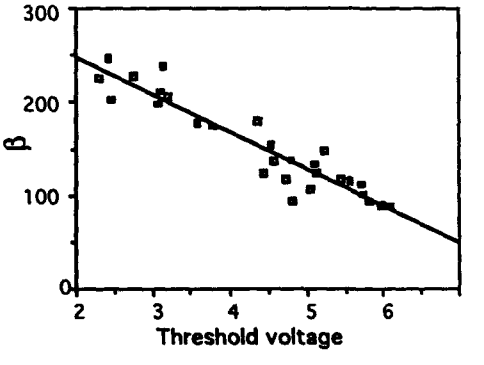


Figure 2. *Variation of H_{FE} versus V_{TH} of degraded structures. Note the linearity of this curve. The straight line corresponds both to the less mean square curve and to the modelling one.*

where R_1, R_2, and L are defined in figure 1; N_{SS} is the surface state density of trapped charges, C_{OX} the oxide capacitance.
In the present work, we develop a modelling of the current gain H_{FE} to explain the correlation with the threshold voltage V_{TH} experimentally observed, as shown in figure 2; the current gain varies linearly in function of the threshold voltage. By the control of this correlation we then propose solutions for improving the technological process.

2. MODELLING OF THE CORRELATION

The modelling of the current gain takes into account the circular geometry of the device, the low doping level of the base and the decrease of the effective base width due to the trapped charges in the overlayed oxides. In this case, the holes diffusion length is higher than the effective base width; since the differential equation for minority carrier distribution, p(r), in the base is:

$$r^2 \frac{d^2p(r)}{dr^2} + r\frac{dp(r)}{dr} = 0 \qquad (2)$$

this expression is a special form of the EULER differential equation, the solution of which is:

$$p(r) = A + B \ln(r) \qquad (3)$$

The constants A et B are determined with the boundary conditions :

$$p(R_2) = 0$$

and

$$p(R_1) = \frac{n_i^2}{N_d}\left[\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right] \qquad (4)$$

then,

$$p(r) = \frac{\ln(R_2/r)}{\ln(R_2/R_1)}\frac{n_i^2}{N_d}\left[\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right] \qquad (5)$$

N_D is the doping concentration of the epilayer (base of the transistor); V_{EB}, k,T, and n_i have the usual meaning.
An approximate expression of the current gain is obtained by using the various current components shown in figure 3, where,

- I_1 and I_2 are the lateral and vertical hole currents,
- I_3 the lateral electron current,
- I_4 the vertical one.

Neglecting the recombination currents and assumimg that I_1 is higther than I_2, an estimate of the current gain is given by :

$$\beta = \frac{I_1}{I_3 + I_4} \qquad (6)$$

The current component I_1 is determined by using p(r).; I_3 and I_4 are approximated by the currents calculated in a one-dimensional junction. Thus,

$$I_1 = \frac{2q\pi ZD_p}{\ln(R_2/R_1)}\frac{n_i^2}{N_d}\exp\left(\frac{qV_{EB}}{kT}\right)$$
$$I_3 = 2q\pi ZD_n\frac{n_i^2}{N_a}\exp\left(\frac{qV_{EB}}{kT}\right) \qquad (7)$$
$$I_4 = q\pi R_1^2 D_n\frac{n_i^2}{ZN_a}\exp\left(\frac{qV_{EB}}{kT}\right)$$

where the cross-sectional areas are chosen as $A_1 = 2\pi ZR_2$, $A_3 = 2\pi R_1 Z$ and $A_4 = \pi R_1^2$ for *current components I_1, I_3, and I_4 respectively. N_a is the* doping concentration of the p-type layer (collector and emitter). D_p and D_n are the diffusion coefficient of holes and electrons respectively.

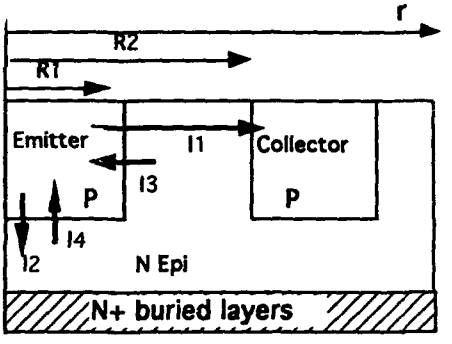


Figure 3: *schematical definitions of various current components in the PNP lateral bipolar transistor.*

The current gain is then expressed by:

$$\beta = \frac{1}{\ln(R_2/R_1)}\frac{N_a}{N_d}\frac{D_p}{D_n}\frac{1}{1 + \frac{R_1^2}{2Z^2}} \qquad (8)$$

Figure 4 shows the variations of β in function of $1/\ln(R_2/R_1)$; the measurements (dots) are performed in three different zones of the wafer. The curves corresponds to the modelling. One can observe a good agreement between experimental results and modelling.

When R_2/R_1 is much higher than 1, a finite series of the threshold voltage gives:

$$\Delta V_{TH} = \frac{A}{\ln(R_2/R_1)} \qquad (9)$$

Expressions (8) and (9) yields:

$$H_{FE} = C \times \Delta V_{TH} \qquad (10)$$

where

$$C = \frac{K \dfrac{1}{1 + (R_1^2/2Z^2)}}{\dfrac{qN_{SS}\ln(R_1+L)/R_1}{2C_{ox}}} \qquad (11)$$

In the formula (11), K is defined as follows:

$$K = \frac{N_a}{N_d}\frac{D_p}{D_n} \qquad (12)$$

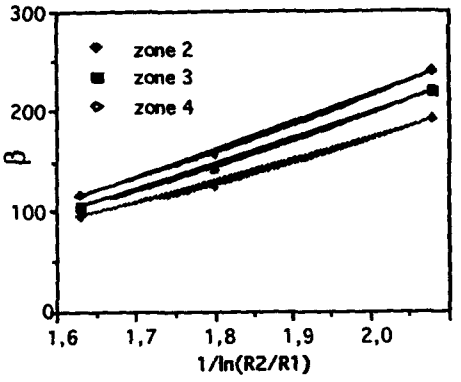


Figure 4. *Variation of H_{FE} versus $1/\ln(R_2/R_1)$ for several zones. One can observe a good agreement between experimental results and modelling.*

The formula (10) shows that the variation of the current gain can be linearly dependent of the variation of the threshold voltage.

3. CRITERIA OF RELIABILITY

A fit of an experimental curve is shown in figure 2. To this end and taking into account the technological parameters, the modelling parameters are chosen as follows: $K = 200$, $V_{THo} = 8{,}25$ V, and $N_{SS} = 3.5 \times 10^{11}\text{cm}^{-2}$ [5]. In this figure, the straight line corresponds both to the less mean square curve deduced from experimental dots and to the modelling one, that gives evidence again of the good agreement between experimental and modelling. We can note that as long as this type of degradation exists, we can observe such a variation, that means that our modelling, which leads to the relation:

$$H_{FE} = C \times \Delta V_{TH} \ , \qquad (10)$$

becomes a criteria to analyse the reliability of the process and the effects of process modifications.

4. IMPROVEMENT OF THE PROCESS

From this criteria and thus from the analysis of the evolution of the variation of the current gain, we tried to improve the process by the use of several technological steps.

The first treatment we performed is a thermal annealing in forming gas; this step did not exist in the initial process. This annealing induces an increasing of V_{TH} of degraded transistors by decreasing the concentration of trapped charges, as shown in figure 5, but does not modify the geometry of degraded zones. This step induces a translation of the curve V_{TH} (H_{FE}) as shown in figure 6.

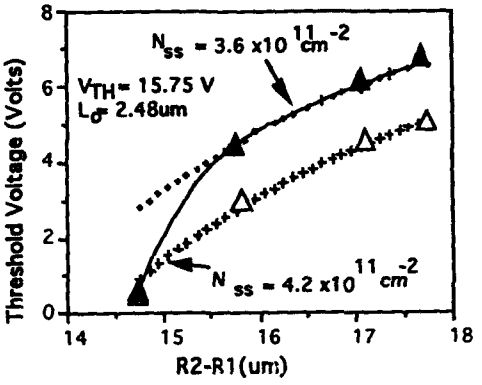


Figure 5. *Fit of experimental curves of threshold voltage versus the base width. Thermal annealing decreases trapped charge concentration.*

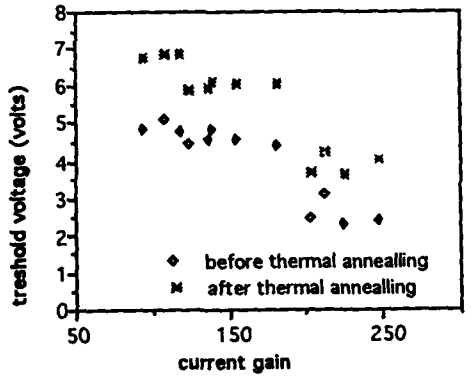


Figure 6. *Variations of V_{TH} versus H_{FE} before and after thermal annealling in forming gas. After annealing, the curve is globally shifted towards higher values of the threshold voltages.*

An other way is ionic implantation of phosphorus in the low doped base. It implies an increase of the effective concentration in the base, which makes more difficult the inversion in this zone and thus prevents the creation

of a hole channel. The dose of implantation was calculated to obtain a threshold voltage in the range of 20 Volts. The devices are characterized as previously in two zones of the wafer, first at the center (zone 1) and second in a lateral test cell (zone 5).

		5µm	6µm	7µm	8µm	9µm
zone1	V_{th}	19	22.6	23.2	23.6	24
	β	26	17	12	10.5	9
zone5	V_{th}	20.2	24.5	24.5	24.7	24.5
	β	27	17.5	14	11	9

Table 1: *Effect of ionic implantation of the base on the threshold voltage and the current gain. The correlation of this two electrical parameters disappears, Vth is in the range of 20 Volts and β in the range of 15*

Table 1 gathers the electrical results for the several transistor sizes; it gives evidence of the shift of the threshold voltages and more especially the important values of V_{TH} and its non dependence with the base width, but H_{FE} takes lower values. This can be explained by the increasing of base doping concentration because the additional ionic charges compensate for trapped charges and the linear variation of the V_{TH} in function of H_{FE} does exist no longer. On the other hand, the current gain is more dependent of the base width, the carrier lifetime (of the electrons) being affected by the implantation and the diffusion length becomes lower than the base width. Although this step allowed to decrease the effective role of the trapped charges in the characterisitics of the device, because the current gain is very low, this step is not actually interesting for applications.

The first technological modification proposal has appeared easier and was applied to increase the yield of fabrication process and its reliability.

5. CONCLUSION

Our modeling allows to well explain the correlation between the bipolar transistor current gain and the PMOS FET threshold voltage of the I2L test cell when trapped charges are located in the insulator gate on both sides of this electrode. Using this modelling, we have defined a criteria of the quality of the process which allows to improve this process and its reliability. Applying this technique, we have found the step of the process which occurs on this behaviour and a technological solution consisting all simply in an additional annealing in forming gas after plasma etching.

REFERENCES

[1] S. GUILLOU, A. SEIDOU, O. BONNAUD and M. TAURIN, "Analysis and modelling of the degradation of a I2L PNP bipolar transistor in relation with the threshold voltage variation of a PMOSFET associated in a test cell", Proc. ESREF 91, vol. 2, 1991, pp. 569-575.

[2] A. SEIDOU, O. BONNAUD, L.ROLLAND and M. TAURIN, "Origin of the degradation of the threshold voltage of a PMOSFET associated to a lateral PNP bipolar transistor in a I2L test cell", Proc. ESREF 92, vol. 1, 1992, pp. 443-445.

[3] P. SINGER, "Charge damage caused by nonuniform plasmas", Semiconductor International, Juillet 1992, p. 34.

[4] S. FANG, J.P. McVITTIE, "Thin oxide damage from gate charging during plasma processing", IEEE Elec. Dev. Lett., vol.. 13, (5), May 1992, p. 288.

[5] A. SEIDOU "Caractérisation électrique et analyse de composants intégrés dans une technologie bipolaire I2L en vue de l'amélioration du procédé de fabrication", thesis, N°926, Université de Rennes 1, France, 1993

FLOATING GATE P-MOSFET FOR THE MOBILE ION CONTAMINATION TEST.

Jozef C. Mitros, John B. Finn
National Semiconductor Co.
3333 W. 9000 S., West Jordan, UT 84088
801-562-7000 Fax: 801-562-7500

1. ABSTRACT

A floating gate p-channel MOSFET is proposed for end-of-line positive mobile ion contamination testing. The device can be fabricated in any CMOS or MOS capable process. The device allows for a quantitative measure of the total mobile ion charge collected during the measurement. Because the device attracts positive ions isotropically, it is sensitive to contamination above the gate, unlike capacitor or normal transistor based tests. The isotropic nature of the response also allows the device to detect cracks in overlying passivation, which provide transport paths for contamination from the passivation surface.

2. INTRODUCTION

Electrically Programmable Read Only Memory, EPROM, cells have negatively charged floating gates while programmed. Positive mobile ions, if present, are attracted to the floating gates and neutralize the charge, resulting in deprogramming. Because of that effect, EPROM cells are extremely sensitive to mobile ion contamination, which poses a serious reliability problem. However, that very sensitivity makes the EPROM cell an ideal candidate for use as a mobile ion sensor.

The purpose of this paper is to propose a floating gate p-channel MOSFET, similar to the EPROM cell, for use in an end-of-line positive mobile ion contamination test. The device can be fabricated in any CMOS or PMOS process, with poly or metal gate and with any gate oxide thickness.

3. DEVICE OPERATION

The proposed test structure is a p-channel transistor with floating poly or metal gate, i.e. not connected to any pad or to another structure. A cross-section and top view of the structure are shown on Figure 1. The device is programmed (turned on) by applying sufficient drain voltage to inject hot electrons into the gate. In the presence of mobile ions, the gate charge, and hence voltage is reduced after annealing through charge neutralization. If the shift in gate voltage can be determined, the amount of charge due to mobile ions can be calculated directly.

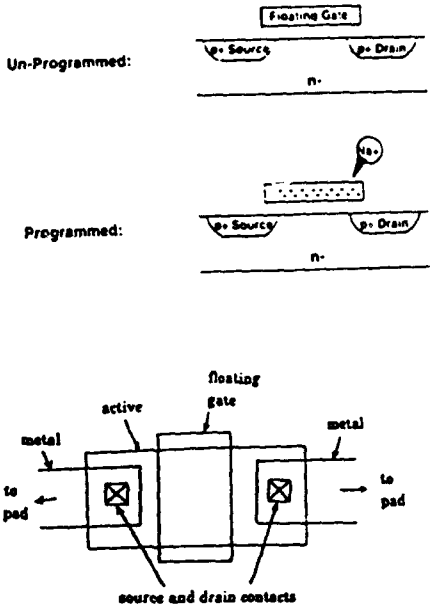


Fig. 1 Crossections and top view (layout) of single poly p-channel MOS transistor.

Figure 2 shows an example of device programming. In curve a), the drain voltage of the proposed device is swept from 0 to 25 V. At roughly 12:5 V, hot carrier injection starts to deposit negative charge on the gate, turning on the device. This can be seen from the increase in current above what is expected. At roughly 17.5 V, the device is fully programmed. Curve b) is the I-V characteristic of the fully programmed device. Curve c) is the I-V characteristics of a similar, normal p-channel device, swept Vgs = 0.2 Vds. The curve is very similar to curve a) for low voltages indicating that the coupling coefficient between the floating gate and the drain is 0.2. The similarity in curves is also important in justifying the use of similar p-channel transistor I-V characteristics to extract the floating gate voltage, as discussed in the next section.

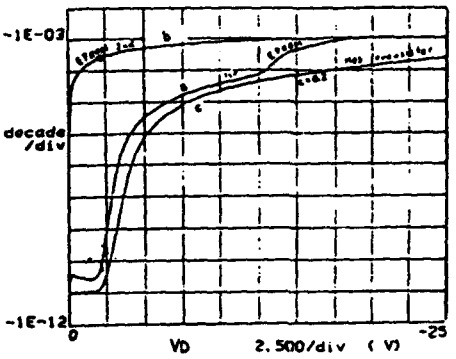


Fig. 2 Drain current versus drain voltage:
a - virgin MOS transistor,
b - programmed MOS transistor,
c - standard MOS transistor.

A programmed p-channel transistor has negative charge on the floating gate. At high temperature (200 - 400°C) the negatively charged floating gate attracts positive ions from the surrounding oxide. As the ions are attracted to, and neutralize, the magnitude of the gate voltage drops which results in reduced drain current when the device is biased at room temperature. Thus, the structure works as a monitor of positive mobile ion contamination. Because the charge on the gate attracts ions isotropically, it is sensitive to mobile charge above the gate. This is in contrast to techniques such as the bias-temperature capacitance-voltage method, which is only sensitive to contamination in the gate oxide. The ability to sense mobile ions in the passivation is important because in actual device operation those ions can diffuse to active areas causing degradation in electrical parameters. Also, if the passivation is cracked, mobile ions from the ambient environment can also diffuse towards active areas. The proposed floating gate p-channel MOSFET covers all of those problems.

While the device is compatible with all CMOS and PMOS processes, the channel length should be about 50% longer than the minimum length allowed in the technology design rules. This is because a floating gate MOS transistor is more sensitive to short channel effects than its standard transistor counterpart.

4. TEST PROCEDURE

1. Program the proposed device by applying Vds = BVdss - 0 5 V: e.g. if BVdss = -12 V then apply Vds = -12.5 V. A lower in (magnitude) Vds can be used with increased programming time.

2. Measure the drain current at low drain bias (-1 V for example).

3. Interpolate the floating gate voltage from the Ids vs Vgs characteristic of a similar (gate size and oxide thickness) normal p-channel transistor biased at the same Vds (see Figure 3).

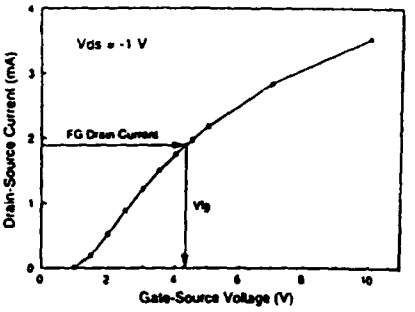


Fig. 3 Method of determining the floating gate voltage.

4. Bake wafer(s) for 30 minutes at 390°C (or longer for lower temperatures).

5. Remeasure the drain current at same low drain bias.

6. Interpolate the floating gate voltage from the Ids vs Vgs characteristics of a similar normal p-channel transistor biased at the same Vds (see Figure 3).

7. Calculate the effective ionic charge collected from the floating gate voltage shift ΔV_{fg}

$$Q_{ion} = C_{ox} \Delta V_{fg}$$

where C_{ox} is the gate oxide capacitance.

Figure 4 shows a plot of floating gate voltage as a function of bake time using the proposed test structure and the test method described above. The passivated wafers were contaminated with $NaHCO_3$ to provide a source of positive mobile ions. The voltage shift saturates as the mobile ion concentration above and below the gate is depleted.

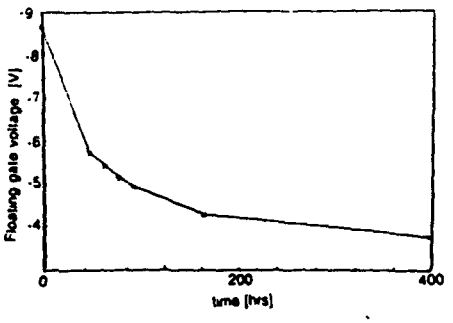


Fig. 4 Floating gate voltage versus the bake time at 250°C.

5. SUMMARY

The proposed device and test method detects and measures positive mobile ions from both gate and field oxides. It is also sensitive to defects in the passivation. If the passivation is cracked, ionic contamination can penetrate those cracks. Later, the ions can diffuse to active device areas causing threshold voltage shifts in MOSFETs or increased leakage in bipolar transistors.

Conventional CV tests detect mobile ions between gate and substrate only. Even if finger field oxide capacitors or MOS transistors are used, mobile ions outside the gate area are repelled when the gate is positively biased to drive the ions to the oxide-semiconductor surface. The ability to isotropically detect mobile ion contamination is a key benefit of the device based test structure..

6. ACKNOWLEDGMENT

The authors are greatful for the critical comments and wish to thank Donald Pierce of Sandia.

COMPARISON OF INTER-POLYSILICON OXIDE LIFETIME USING RAMPED AND CONSTANT VOLTAGE MEASUREMENTS

Andreas Martin, Paula O'Sullivan, Alan Mathewson
National Microelectronics Research Centre (NMRC) University College Cork,
Lee Maltings, Prospect Row Ireland
Tel: + 353 21 276871 ext 4080, Fax + 353 21 270271

1 Abstract

This paper reports the use of ramped voltage and constant voltage stress measurements for the assessment of reliability data of "thermally grown oxides on heavily doped polysilicon" (inter-poly oxide). When these measurements are performed for the characterisation of the oxide it is observed that the use of a ramped voltage stress results in an enhanced oxide lifetime compared to the constant voltage case. By plotting the evolution of the current as a function of time for both measurements charge trapping in the oxide is found to be the principal cause of longer lifetimes measured with a ramped voltage stress. Measurements on pre-stressed oxides confirm this. This work shows that in order to relate the results from constant voltage and ramped voltage stress charge trapping must be taken into consideration.

2 Introduction

Reliability predictions for dielectrics are essential to provide customers with yield analysis data of a product. For applications which require poly-poly capacitors such as switched-capacitor filters or A/D and D/A-converters etc. it is essential to characterise the inter-poly oxide in terms of lifetime and failure probability under operating conditions.

A prediction method has been proposed for inter-poly oxides in [1] which is based on results from constant voltage stress (CVS) measurements. This methodology uses the 1/E-Model (see equation 1) which has been developed by Hu et al. [2]. In [1] only the intrinsic properties of the oxide were investigated. However, defect related failures exist which lead to malfunction of devices and cause field failures of the product. Therefore, the defect related properties are important for the evaluation of inter-poly oxide reliability. The monitoring of defect related breakdowns with CVS involves long measurement times which are impractical for industrial applications. A ramped voltage stress (RVS) measurement, however, can be carried out faster than a CVS and monitors the defect related as well as the intrinsic breakdowns. Therefore, a RVS is preferable to record early and intrinsic breakdowns in a reasonable time. A correlation between RVS and CVS data for MOS gate oxides is described in [2]. The possibility of a similar approach for inter-poly oxides is examined in this work.

This paper shows that results of RVS measurements with slow ramp rates overestimate the lifetimes of inter-poly oxides when compared to those predicted by CVS. This overestimation is crucial when the RVS data is used for yield predictions. These longer RVS lifetimes are related to charge trapping which is observed in inter-poly oxides for low Fowler-Nordheim (FN) currents [3]. Experiments presented in this work on pre-stressed oxides and the I-t characteristics of CVS and RVS confirm the this. Therefore, it is important to account for charge trapping when long term

reliability data is predicted from RVS data.

The aim of this work was the quantitative determination of charge trapping for lifetime predictions from RVS and the correlation of RVS to CVS results by using the 1/E Model. By taking the charge trapping into account reliability predictions are possible from RVS data. The impact of the charge trapping and the application of the obtained results for the characterisation of long term inter poly oxide reliability are discussed.

3 Experimental Procedure

All measurements were performed on wafer level with a HP4062 Parametric Test System. Test capacitors from two different processes with the cross section of figure 1 were included in the measurements.

The lower polysilicon layer was implanted and the top polysilicon layer $POCl_3$ doped with doping concentrations $> 10^{20}$ cm^{-3} for both layers. The inter poly oxides consisted of thermally grown SiO_2 with thicknesses of 22.0 and 30.5 nm. In this work the smallest structure size ($1.5*10^{-3}mm^2$) of the test chip was tested to ensure intrinsic distributions.

The inter poly capacitors were all stressed with a positive potential on the top polysilicon plate as this showed the lowest breakdown fields. CVS and RVS measurements were then carried out with five or six different bias conditions for both processes. Ramp rates varied from 0.1Volt/0.15sec to 0.1Volt/7sec. All ramp rates used in the RVS consisted of the same voltage step of 0.1Volt. Sample sizes were around 40 capacitors for each measurement.

4 Analysis Of Measurement Results

For the prediction of lifetimes and failure probabilities at operating conditions the 1/E-Model (see equation 1) is applied to the experimental data.

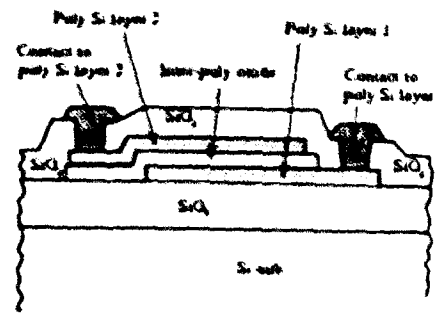


Figure 1 Cross section of an inter poly test capacitor

$$t_{bd} = \tau_0 * exp\left(\frac{G * T_{ox}}{V_{ox}}\right) \qquad (1)$$

Note, that for the evaluation of intrinsic lifetimes as described in [1] only CVS results (see figure 2) are necessary. For the determination of intrinsic lifetimes the log of the lifetime to 50% failure (t_{bd50}) of a CVS is usually plotted as a function of the reciprocal of the bias voltage (or field) [1, 2].

Since in this paper the main interests are to account for defect related properties of a dielectric and the correlation of RVS and CVS results the method which is described in [1] is expanded to RVS data. This leads to a direct comparison of lifetimes of both measurements. The RVS measurements performed in this work consisted all of a staircase with voltage steps of 0.1Volt and because such a RVS measurement is in effect a sequence of short constant voltage stresses with increasing magnitude, the time to breakdown under RVS is estimated by the time spent at the final step of the ramp prior to the step in which catastrophic breakdown occured. The stress of earlier steps of the ramp is not taken into account for the RVS lifetimes. Therefore, the log of the timestep of the ramp (t_{step}) is plotted versus the reciprocal of the breakdown voltage to 50% failure (V_{bd50}) in figure 3 and 4 for both processes [7].

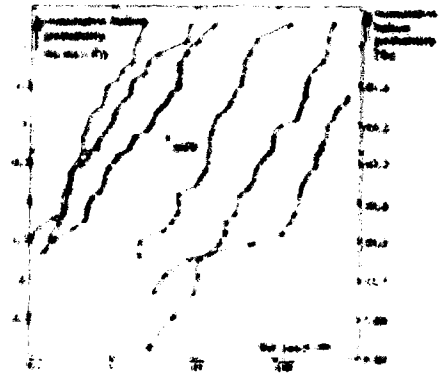


Figure 3 Lifetime distributions of the 22 nm inter poly oxide at six different ramp voltages

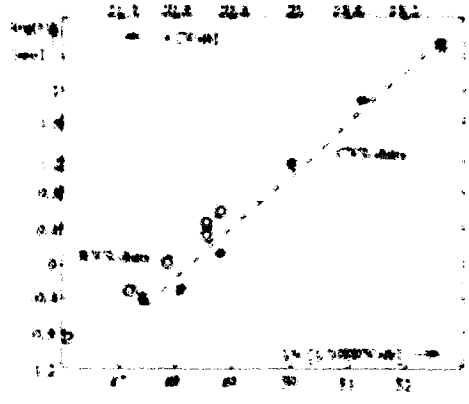


Figure 4 RVS and CVS data in a "ln(t) versus 1/E" plot for small area capacitors of the 22nm inter poly oxide

It would be expected that the stress of earlier steps of the ramp age the oxide and therefore reduce the lifetime at breakdown voltage level. However, the regions in figures 3 and 4 in which the line fits show an overlap of the reciprocal voltages give a direct comparison of lifetimes under CVS with lifetimes under RVS. It is clear from this region that the oxides last longer at the final step of the ramp than they would last if a CVS was applied at the same voltage level.

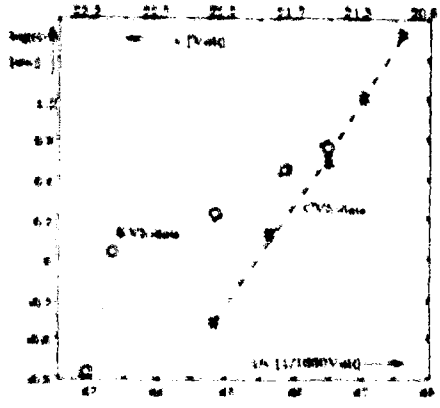


Figure 6 RVS and CVS data in a "ln(t) versus 1/E" plot for small area capacitors of the 30 nm inter poly oxide

5 Discussion Of RVS Lifetime Overestimation And Charge Trapping

Charge trapping due to electron tunneling during Fowler Nordheim conduction is reported in the literature for inter poly oxide [2, 4]. Electrons are trapped near the lower Poly Si Si_3O_4 interface [5]. Therefore the field across the oxide is no longer constant and the tunneling distance of the electrons through the potential barrier of the oxide is increased which results in smaller tunnel currents.

I-t characteristics (see figure 5) of RVS and CVS provide evidence of the charge trapping. Lower currents are always observed for the RVS compared with CVS. In figure 5 the average currents at 19.5 Volt for six RVS with different ramp rates are compared to the average currents of the CVS at 19.5 Volt for the 22nm inter poly oxide. The times which are

related to the average RVS currents are equal to t_{ramp} of the ramp rate of the specific RVS

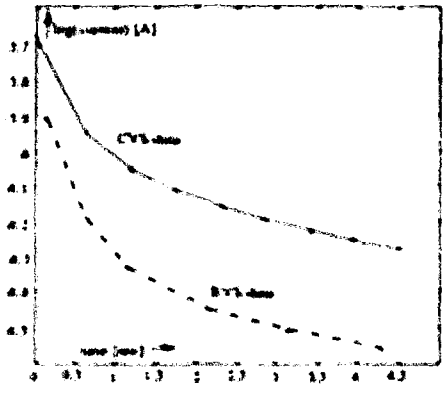


Figure 3 Measured I-t characteristics of CVS and RVS at the stress bias 19.5 Volt for the 22nm inter poly oxide

Additional measurements were performed on pre stressed inter poly oxides to confirm the charge trapping and the resulting longer RVS lifetimes. These measurements for the 30.5nm inter poly oxide consisted of an initial ramp from 11.25-21.25 Volt with a fast ramp rate of 0.5Volt/sec and a following CVS measurement at 21.25 Volt until catastrophic breakdown occurred. In figure 6 the lifetimes of the pre stressed oxides are compared with the lifetimes of the CVS at 21.25 Volt which had been performed earlier. Note, that the time of the ramped pre stress was not taken into account. However, it can be clearly seen that the pre stress increases lifetimes.

In other measurements of pre stressed inter poly oxides, when the ramp of the pre stress is started at a level of 6.25 Volt (see figure 6) no additional increase in lifetimes was observed. It can be concluded that for the 30.5nm inter poly oxide no "significant" charge trapping occurs at voltage levels less than 11.25 Volt

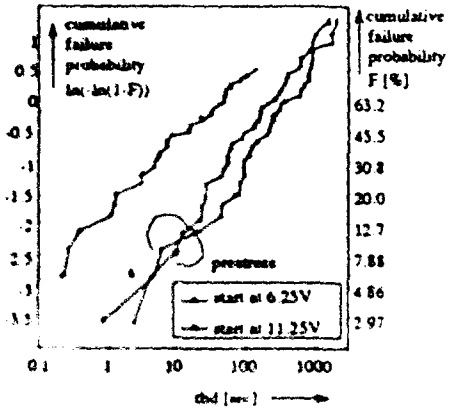


Figure 6 Comparison of inital and pre stressed 30.5nm inter poly oxides at 21.25 Volt

It is reported in [5] that the charge trapping in an inter poly oxide does not saturate and it is rapid initially because of the presence of traps with large cross sections which are soon completely filled. Therefore, when a slower ramp rate is chosen the quantity of trapped charges increases in comparison to an oxide which was stressed with a fast ramp rate. This effect can be observed from the evolution of the RVS current in figure 5. It can be concluded that with fast ramp rates less charge trapping occurs.

The observations which were made for both inter poly oxides of this work, namely, that lifetimes increase due to a pre stress are in contradiction to the literature. In general, a pre stress would be expected to shorten the lifetime of a capacitor. In [2] is reported that for gate oxides a pre stress such as a burn-in decreases lifetimes.

This contradiction can be explained by the influence of temperature and the time spent during a burn in. In [4, 6] it is shown that at high temperatures electrons are detrapped and for the same bias greater tunnel currents degrade the oxide faster. Usually, a burn-in step consists of a CVS at high temperature

and lasts longer than a RVS. Therefore, with a long stress time and higher initial tunnel currents than in a RVS the oxide degrades faster during the burn-in than in a RVS.

6 Practical Applications Of The Observed Results

Correlation of RVS and CVS Data : For the correlation of the RVS results to the CVS data charge trapping has to be considered. Some indicators of charge trapping are

- the difference between the RVS currents and the CVS currents at the same bias point (see figure 5),
- the difference of the lifetimes of RVS and CVS which can be observed from figure 3 and 4.

In a practical approach for the consideration of charge trapping the difference between the RVS and CVS line fits from figure 3 and 4 are used to correlate the lifetimes for each ramp rate. For the RVS 22nm oxide and the ramp rate of 0.1Volt/2sec the time difference is 1.25sec.

Figure 7 and 8 show predicted lifetimes from RVS data which take charge trapping into account. These are in good agreement with the measured lifetimes. For the prediction from RVS to CVS equation 2 used is based on the 1/E-Model and described in [7].

$$ln(t_{bd_{use}}) = G*\left(\frac{1}{V_{use}} + \frac{1}{V_{bd}}\right) + ln(t_{step}) - \Delta\, ln(t) \qquad (2)$$

V_{use} is the voltage at which the capacitors will operate and $\Delta\, ln(t)$ considers the charge trapping of the RVS and can be determined from the difference between the RVS data and the CVS line fit of figure 3.

Failure Probability Predictions : The prediction of failure probability as a function of lifetime requires measurements on large and small area capacitors. With several CVS and a RVS measurement on small area capacitors parameter G of equation 1 and $\Delta\, ln(t)$ of equation 2 are determined.

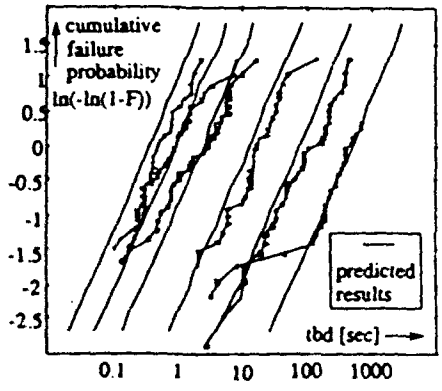


Figure 7: 22nm oxide: comparison between measured and predicted lifetimes from the RVS with the ramp rate of 0.1V/1sec, G=624MV/cm and $\Delta\, ln(t)$=0.69.

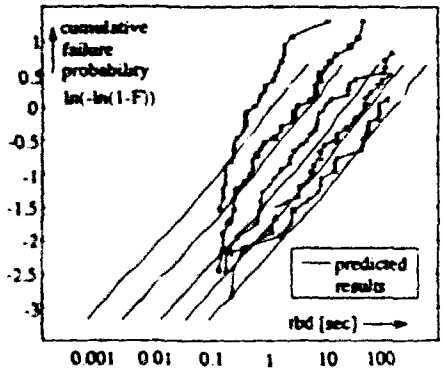


Figure 8: 30.5nm oxide: comparison between measured and predicted lifetimes from the RVS with the ramp rate of 0.1V/2sec, G=585MV/cm and $\Delta\, ln(t)$=1.96.

With RVS measurements on large area capacitors the defect related properties are recorded. Note, that the same ramp rate must be used for all RVS and t_{step} must lie in the region of the measured t_{bd50} which is important for the evaluation of $\Delta ln(t)$. Lifetimes and failure probabilities can be projected from the RVS data of the large area capacitors by using equation 2 [7].

Oxide Quality Assessment : In this work RVS and CVS measurements are performed on small area capacitors for reliability evaluations. Besides predicting lifetimes and failure probabilities from the measurement results the CVS/RVS current ratio and the CVS-RVS lifetime difference give a measure for the quality of the oxide for different processes in terms of charge trapping. In this paper the 30.5nm inter-poly oxide shows more charge trapping than the 20nm oxide for the fast and intermediate ramp rates.

7 Conclusions

It was found by direct comparison of the lifetimes in a "log(t) versus 1/E"-plot that RVS results overestimate the lifetime of inter-poly oxides with thicknesses in the range of 22 - 30.5nm. A comparison of RVS with CVS currents at the same bias conditions give evidence that charge trapping is the reason for the longer lifetimes. This was confirmed with measurements on pre-stressed oxides.

The charge trapping is dependent on the starting bias level and the ramp rate of the RVS. No additional "significant" charge trapping was observed below 11 Volt for the 30.5nm inter-poly oxide. Therefore, it is important to take the trapped charge into account when lifetimes at operating voltage are predicted from RVS data.

It has been shown that by relating RVS to CVS lifetimes accounting for the charge trapping, lifetimes and failure probabilities can be predicted from RVS results. Experimental and predicted results show good agreement. With the observations and results of this work RVS measurements can be used for lifetime predictions and in comparison to CVS results for the characterisation of the oxide quality in terms of charge trapping.

Acknowledgements

The authors wish to acknowledge the partial funding of this work by the JESSI Joint Logic Project under a subcontract from GEC Plessey Semiconductors in Roborough, UK.

References

[1] S.-J.Wang, I.C.Chen, H.L.Tigelaar, "TDDB on Poly-Gate Single Doping Type Capacitors", *IEEE International Reliability Physics Symposium*, pp.54-57, 1992.

[2] R.Moazzami, C.Hu, "Projecting Gate Oxide Reliability and Optimizing Reliability Screens", *IEEE Transactions on Electron Devices*, vol.37, no.7, pp.1643-1650, July 1990.

[3] M.-S.Liang, N. Radiy, W.Cox, S Cagnina, "Charge Trapping in Dielectrics Grown on Polycrystalline Silicon", *Journal of Electrochemical Society*, vol.136, no.12, pp.3786-3790, 1989.

[4] J.Bisschop, E.J.Korma, E.F.F.Botta, J.F.Verwey, "A Model for the Electrical Conduction in Polysilicon Oxide", *IEEE Transactions on Electron Devices*, vol.33, no.11, pp.1809-1815, Nov.1986.

[5] C.Hu, D.Y.Joh, Y.Shum, T.Klein, "Electron Trapping in Oxide Grown from Polycristalline Silicon", *International Electron Devices Meeting*, pp.229-232, 1979.

[6] S.Suyama, A.Okamoto, T.Serikawa, "Electrical Conduction Mechanism and Breakdown Properties in Sputter-Deposited Silicon Dioxide Films on Polycrystalline Silicon", *Journal of Applied Physics*, vol.65, no.1, pp.210-214, Jan.1989.

[7] A.Martin, P.O'Sullivan, A.Mathewson, "Evaluation of Lifetime and Failure Probability for Inter-Poly Oxides from RVS-Measurements", *Microelectronics Journal*, vol.25, to be published in 1994.

FREQUENCY DEPENDENCE OF DEGRADATION AND BREAKDOWN OF THIN SiO_2 FILMS

M. Nafría, D. Yélamos, J. Suñé, and X. Aymerich
Departament de Física (Electrònica)
Universitat Autònoma de Barcelona, 08193-Bellaterra (SPAIN)

I. ABSTRACT

Thin oxide MOS capacitors are subjected to bipolar voltage stresses of different amplitudes and frequencies. According to a previously proposed breakdown model, the evolution of the current with the stress time has been considered to be due to the degradation of the oxide, i.e. to the generation and partial occupation of electron traps. When *Log(J)* is represented versus the stress time, the slope of the plot and the magnitude of the current (which tends to decrease during constant-voltage tests) are taken as indicators of the oxide degradation rate and degradation level, respectively. Our results suggest lower degradation rates, and consequently lower degradation levels for the same stress times, at high frequencies. This is consistent with the increase of time-to-breakdown with stress frequency observed by other authors, and confirms that, also for dynamic stresses, the relation between degradation and breakdown is fundamental to understand the physics of these phenomena. The slower degradation rates confirm the improvement of oxide reliability under dynamic AC stress conditions.

2. INTRODUCTION

The ever increasing complexity of integrated circuits is pushing the reliability community towards a very broad change in the techniques for reliability assesment. Within the "building-in" reliability arena, two are the main directions of action: reliability through process control; and reliability through circuit simulation. Reliability is becoming one of the key specifications that have to be considered from the very beginning of the design process. Reliability simulation tools help to optimize the reliability by design through the evaluation of the susceptibility of ICs to the various device failure mechanisms, and through comparison of the different design and fabrication alternatives. This represents a very important change in reliability assessment because, traditionally, reliability has only been improved by making accelerated stress experiments on test structures, and by the analysis of the failures.

For the simulation of the degradation and final failure of the devices, models for the different failure mechanisms have to be implemented into the reliability simulators. The accuracy of the simulations is directly related to the suitability of these models. One of the mechanisms which limit the reliability of MOS circuits is SiO_2 dielectric breakdown. Oxide reliability has been usually analyzed under static stresses (constant-voltage or constant-current) or ramps. The extrapolation of time-to-breakdown data to operation conditions is generally performed under the assumption that there is not a strong dependence on the type of stress. However, in most circuits, the operation of the devices will be dynamic, and the experimental results of other authors indicate that the oxide reliability significantly improves under these conditions, especially if the stress is AC (Ref.1).

Much work has been dedicated to the study of the breakdown of thin silicon dioxide films under static conditions. Even though, many questions concerning the physics of this failure mechanism remain still open. In any case, however, there are several points of general agreement which are based on direct experimental observations:

i) There are at least two types of breakdown, defect-related and intrinsic, the later becoming more important for advanced circuits (Ref. 2). In this paper we only deal with intrinsic breakdown.

ii) Breakdown is a local phenomenon.

iii) Time to breakdown and any other breakdown variable (breakdown field, for example) are not deterministic. A statistical distribution of results is always found (Refs. 3,4).

iv) Breakdown is related to some sort of previous degradation of the oxide structure which takes place during stress or operation (generation of electron or hole traps, interface states, slow states,...).

This work deals with the degradation and breakdown of thin oxide MOS structures under bipolar AC conditions. Our general project has two main objectives:

a) the development of a degradation and breakdown model suitable for reliability simulators; and b) the analysis of dynamic data to get additional information about the physics of the breakdown. In this work, our particular goal is the evaluation of the relation between degradation and breakdown under AC stress conditions. We want to determine if the correlation between degradation and breakdown, which is widely recognized for DC conditions, is also valid for AC stresses.

3. DEGRADATION-BREAKDOWN MODEL

Although we do not want to discuss our degradation and breakdown model (Ref. 5) in this paper, we will dedicate this section to review it because this is convenient for the analysis of the AC breakdown data.

In stress experiments, after a very small heating distance, the kinetic energy that the electrons gain from the electric field in the oxide conduction band is dissipated by interaction with the SiO_2 lattice (Joule law). The principal assumption of our model is that a very small fraction of this energy is used to damage the oxide structure through the generation of defects which can act as electron traps. The generated traps become partially occupied by electrons (the measured cross sections correspond to neutral traps), and the resulting negative charge modifies the electric field in the oxide. There is much published experimental evidence to support the generation of electron traps and negative charge, and also their relationship with the dielectric breakdown (Ref. 6). Even under dynamic conditions the role of negative charges seems to be dominant (Refs. 7,8), and this paper adds more evidence in this direction.

If the degradation is related to dissipation of energy, the key equation for a degradation model should be Joule law:

$$\frac{dE}{dt} = J(t)F(t) \qquad (1)$$

where E is the dissipated energy per unit of volume, $F(t)$ and $J(t)$, the oxide electric field and the current density, respectively. If E_{dis} is the energy required to generate one defect, and β the efficiency of defect generation, the rate of trap generation is given by

$$\frac{dN}{dt} = \frac{\beta}{E_{dis}} J(t)\ F(t) \qquad (2)$$

$N(t)$ being the density of neutral electron traps.

In previous works it was shown that this simple picture is adequate to model the degradation of SiO_2 under DC electrical stresses. Assuming that the current is given by the Fowler-Nordheim expression, considering that the charge distribution is uniform, and making other simplifying assumptions, it was shown that this model reasonably explains the evolution of the DC stress conditions (current in constant-voltage tests, voltage in constant-current stresses, etc.) (Ref. 5).

In particular, in a constant-voltage stress, the generation and partial occupation of electron trap centres causes the reduction of the current with stress time. Thus, the analysis of the evolution of the current is a useful tool to evaluate the oxide degradation, at least under DC conditions. For this type of stress, our degradation model predicts a current evolution given by:

$$J(t) = \frac{J_o}{(1+\frac{t}{\tau})} \qquad (3)$$

where J_o is the initial current, and τ is a characteristic time which is a function of β and E_{dis}. If we represent *Log(J)* versus time this results in almost a straight line (after an initial transient which is due to trapping in native electron traps (Refs. 5-7)).

$$Log(J) = Log(J_o) - \frac{t}{\tau} \qquad (4)$$

According to this model, in the next section we will analyze the AC breakdown data assuming that the current level (for a particular value of applied voltage) is a measure of the degradation, and the slope of *Log(J)* versus t, a measure of the degradation rate.

As far as the breakdown is concerned, we consider it to be triggered by the generation of a critical number of defects n_{BD} in small portion S_o of the total active oxide area. The local nature of the breakdown and the statistical distribution of events is modeled in such a simple way, at the same time that the relation between degradation and breakdown is considered (Refs. 3,4).

4. EXPERIMENTAL RESULTS

The analyzed samples are polysilicon gate MOS capacitors with oxide thickness of 13.5 nm, area of $3.14\ 10^{-4}\ cm^2$ and n-type substrate ($N_D = 10^{18}\ cm^{-3}$).

Bipolar square voltage waveforms of different amplitudes (15V, 16V and 17V) and frequencies (10Hz, 100Hz, 1KHz, and 10KHz) have been applied to the capacitors. For each of these stress conditions, the current level at the end of the positive semiperiod has been measured after different stress times. The evolution of the current with time has been represented in a *Log(I)* versus *t* plot. In accordance with our degradation model, the current level *I(t)* is taken as an indicator of the degradation of the oxide at time *t*, and the slope as a measure of the degradation rate.

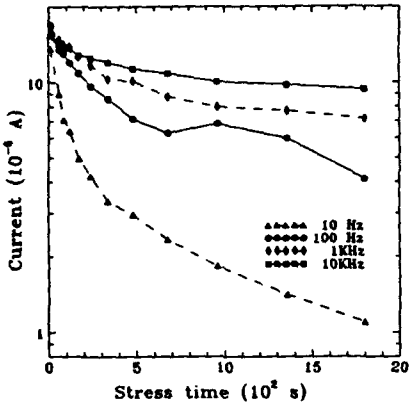


Figure 1.- Current evolution recorded on MOS capacitors subjected to ± 15V bipolar stresses of different frequencies.

Figures 1 to 3 show the evolution of the current with stress time for three different voltage amplitudes, with the frequency as a parameter (each data point corresponds to the average of 5 samples). In each figure one can distinguish two different groups of curves which are clearly separated, the curves corresponding to high frequencies having smaller slopes (smaller degradation rates) and, consequently, larger current densities (smaller degradation levels) for equal stress times. We distinguish two different regimes: a *low-frequency regime*, where the degradation rate is large and the evolution curves are quite similar (if not coincident) to those obtained under equivalent DC conditions (not shown in the figures); and a *high-frequency regime*, where the evolution of the degradation is much slower. In both regimes, the frequency dependence is very weak. A quite abrupt change between one regime and the other occurs at a threshold frequency which depends on the oxide electric field. For V_{appl}=15V (figure 1), only one curve (10 Hz) is in the *low-frequency regime*, and the others (100Hz, 1KHz, and 10KHz) can already be considered to belong to the *high-frequency regime*. For V_{appl}=16V (figure 2), two frequencies (10Hz, 100Hz) are *low* and the other two (1KHz, and 10KHz) are *high*. Finally, for V_{appl}=17V (figure 3), only one stress frequency (10KHz) corresponds to the *high-frequency regime*.

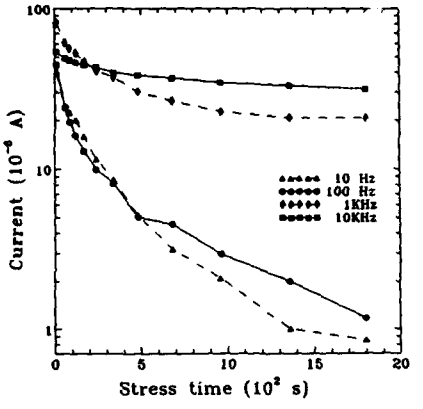


Figure 2.- Current evolution recorded on MOS capacitors subjected to ± 16V bipolar stresses of different frequencies.

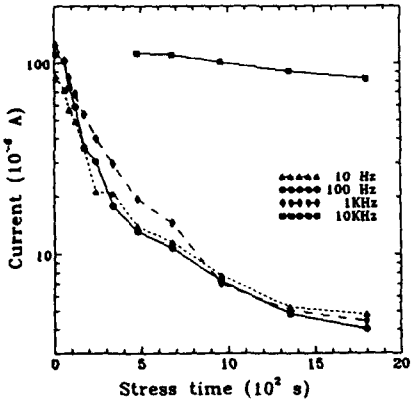


Figure 3.- Current evolution recorded on MOS capacitors subjected to ± 17V bipolar stresses of different frequencies.

Figures 4 and 5 show the evolution of the current with stress time for two fixed frequencies (100Hz

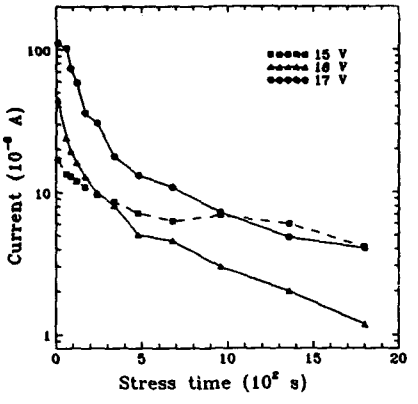


Figure 4.- Current evolution recorded on MOS capacitors subjected to 100Hz bipolar stresses of different amplitudes.

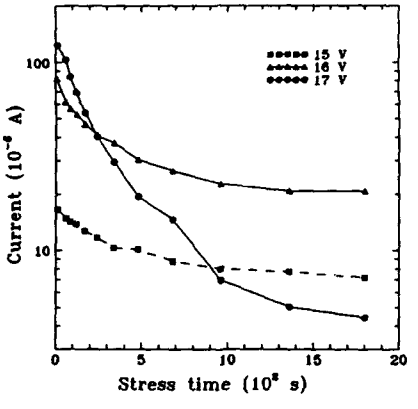


Figure 5.- Current evolution recorded on MOS capacitors subjected to 1KHz bipolar stresses of different amplitudes.

and 1KHz, respectively), and the voltage amplitude as a parameter. In figure 4, we see that the curves which correspond to 16V and 17V have similar degradation rates (slopes), while it is much smaller in the case of the stress waveform with 15V of amplitude. The fact that higher voltages (fields) lead to larger degradation rates should not be surprising within the degradation model of the previous section. In our case, however, the variation of electric field from one stress to another is less than 10%. Nevertheless, in the Fowler-Nordheim regime this small change of field causes an exponential change in the current (see the values of current at the beginning of the stress, t=10s). Thus, based on Joule law, one should expect the degradation rate to be controlled by the current density rather than by the applied voltage. This is perfectly consistent with the curves corresponding to 16V and 17V: at the beginning of the stress, the degradation rate is larger for V_{appl}=17V because the current is about twice the value corresponding to V_{appl}=16V; for longer stress times, the degradation rate (slope) is almost the same for equal values of current density. The same can be said of the curves corresponding to 15V and 16V in figure 5, though in this case the degradation rate is much smaller. The question that remains refers to what happens with the curves which cross over their two partners in these figures (the curve of 15V in figure 4, and the one of 17V in figure 5). To understand these crossovers, it is necessary to make reference to the results shown of figures 1-3. In figure 4, the 100 Hz stress can only be considered to belong to the *high-frequency regime* for V_{appl}=15V, so that in this case the degradation rate is much smaller than for 16V and 17V, which correspond to the *low-frequency regime*. This explains why for V_{appl}=15V the degradation rate is much smaller even at the end of the stress, when the current density is larger than for 16V or 17V. Just the opposite is true in figure 5, where only the 17V curve belongs to the *low-frequency regime*, having a larger degradation rate. If the same kind of representation is used for the 10 KHz data, the three curves (15V, 16V and 17V) are almost parallel, and there is not a crossover because the three stresses belong to the *high-frequency regime* (this figure is not included).

5. DISCUSSION

In previous works (Ref. 8,9), the same samples were used for studying the DC degradation and breakdown characteristics, and the transient current evolution during the application of pulsed voltages (unipolar and bipolar). Also according to our degradation model, these current transients and their evolution under accelerated dynamic stresses were qualitatively explained by assuming trapping and detrapping of electrons near the oxide interfaces. The role of the interfaces in dynamic experiments is of uppermost importance because only traps located near the interfaces can alter their occupation state when the polarization changes during a short time (this is particularly true for dynamic unipolar stresses).

The experimental results of the previous section confirm the idea that the reliability of the oxide improves under AC conditions, as shown by breakdown data (Ref. 1). Moreover, at least qualitatively, the degradation concepts developped for DC stresses have been shown to be valid for AC conditions within both the *high* and *low-frequency regimes*. The reason for the change from one regime to the other requires a more detailed model for the dynamics of generation, i.e. a model for the frequency dependence of the defect generation efficiency, β. The presentation of such a model is out of the scope of this paper but, the basic ideas developped by Rosenbaum et al. (Ref.1), which involve transport of holes to the cathode interface (or of other type of charged species liberated at the anode, i.e. H^+), sound rather convincing if this transport has the generation of electron traps (near the injecting interface) as the main consequence. The guidelines of such a degradation model would be: 1) liberation of species at the anode interface by the energy dissipated in the oxide; 2) transport of positively charged species through the oxide; 3) generation of electron traps in the bulk and near the injecting interface.

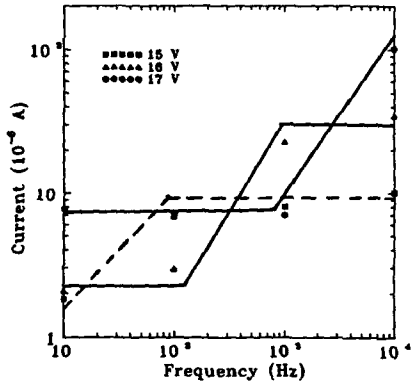


Figure 6.- Current recorded after 960s of stress as a function of frequency for each value of the applied voltage.

As far as the relation between degradation and breakdown under AC conditions, we can compare the results of figures 1-3 and the breakdown data of Rosenbaum et al. (Ref.1). This can be done by plotting the current measured after a fixed stress time as a function of frequency for each value of the applied voltage. Such a representation of the results

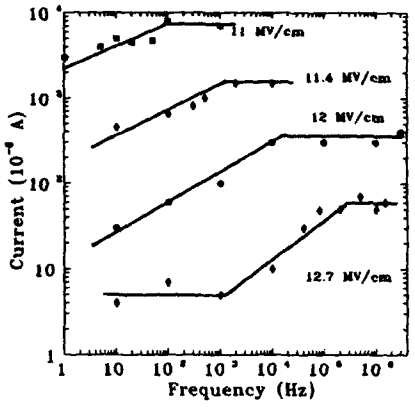


Figure 7.- Time-to-breakdown as a function of frequency for different oxide fields (after Rosenbaum et al., Ref. 1).

is shown in Figure 6 for a stress time of approximately 10^3 s. In Figure 7, the breakdown data of (Ref. 1) are reproduced for comparison. This comparison demonstrates with any doubt that there is a strong correlation between degradation and breakdown under AC conditions. High and low frequency regimes are distinguished in both sets of data, showing weak frequency dependences at both low and high frequencies. A second correlation comes from the fact that the change between one regime and the other occurs for critical frequencies which increase with the electric field (voltage amplitude of the stress) also in both cases. Being different sets of samples, these are rather conclusive results, which demonstrate that:

a) The degradation and the breakdown are strongly correlated under AC dynamic stress conditions.

b) As in DC stresses, the degradation and breakdown are related to electron trapping in the oxide.

This second point may be controversial because we have to accept that the positive charge trapping model of Chenming Hu and co-workers is quite widely accepted in the reliability community. However, our results confirm the importance of negative charge generation in the degradation and breakdown of thin oxide films (Refs. 5-8).

6. REFERENCES

1. Rosenbaum E., Liu Z., and Hu C. "Silicon Dioxide breakdown lifetime enhancement under bipolar bias conditions", IEEE TED-40, No. 12, 1993, pp. 2287-2295.

2. Crook D.L. "Evolution of VLSI reliability engineering", Proc. ESREF, 1991, pp. 293-312.

3. Suñé J., Placencia I., Barniol N., Farrés E., Martín F., and Aymerich X. "On the breakdown statistics of very thin SiO_2 films", Thin Solid Films, vol. 185, 1990, pp. 347-362.

4. Farrés E., Nafría M., Suñé J., and Aymerich X. "Breakdown statistics of thin SiO_2 layers in wearout tests", Proc. ESREF, 1991, pp. 285-292.

5. Suñé J., Placencia I., Barniol N., Farrés E., Aymerich X. "Degradation and breakdown of gate oxides in VLSI devices", Phys. Stat. Sol. (a), 111, 1989, pp. 675-685.

6. Nissan-Cohen Y., Shappir J., Frohman-Bentchkowsky D., "Trap generation and occupation dynamics in SiO_2 under charge injection stress", J. Appl. Phys., vol. 60, 1986, pp. 2024-2035.

7. Haddad S. and Liang M.S., "The nature of charge trapping responsible for thin oxide breakdown under dynamic field stress", IEEE EDL-8, 1987, pp. 524-527.

8. Nafría M., Suñé J., Yélamos D., and Aymerich X., "Behavior of thin oxides under dynamic high field stresses", Proc. ESREF, 1993, pp. 219-222.

9. Nafría M., Suñé J., Aymerich X., "Characterization of SiO_2 dielectric breakdown for reliability simulation", IEEE TED-40, 1993, pp. 1662-1668.

GATE OXIDE SHORTS IN *n*MOS TRANSISTORS: ELECTRICAL PROPERTIES AND LIFETIME PREDICTION METHOD.

Samar A. Dababneh and Charles F. Hawkins
The University of New Mexico
Electrical and Computer Engineering Dept.
Albuquerque, New Mexico 87131

Jerry M. Soden
Sandia National Laboratories
Dept. 2275
P. O. Box 5800
Albuquerque, New Mexico 87185-1081

ABSTRACT

Degradation in *n*MOS transistors from gate oxide shorts is dependent upon oxide trapping and interface state generation. Three distinct damage mechanisms were identified, including generation of : 1) electron traps in the bulk oxide by the injected holes, $N_{ox,h}$, 2) electron traps in the bulk oxide by the injected electrons, $N_{ox,e}$, and 3) interface states, N_{ss}. The three damage mechanisms are incorporated into a device lifetime prediction method.

1. INTRODUCTION

A gate oxide short is an unintended electrical connection through the gate oxide of an MOS transistor. Such a defect can occur between the gate and the source, drain, or channel of the transistor. Gate oxide shorts are a major type of fabrication defect and, in some CMOS processes, the dominant defect (Ref. 1). Detailed accounts of the physical origins of gate oxide shorts have been presented by Soden and Hawkins (Ref. 2) and Yamabe and Taniguche (Ref. 3). A gate oxide short can cause diverse changes in the electrical properties of a device and, therefore, can be responsible for IC failures. Previous experimental (Refs. 1-4) and theoretical (Refs. 5-7) studies show that the phenomena caused by these defects are not straightforward. A single gate oxide short can be responsible for a number of distinct electrical failures in VLSI IC's. The models traditionally used in this area, such as transistor gate-to-source and gate-to-drain shorts (Refs. 8-13), over-simplify what is taking place. Since the physical phenomena associated with a gate oxide short can lead to a gradual degradation of the transistor characteristics and eventually to circuit failure, there is a need to better understand these mechanisms.

Due to the complex nature of the gate oxide short defect, only gate oxide shorts located in the channel region of the transistor will be considered. This paper covers three major areas.

The first area addresses electrical properties affected by a gate oxide short in *n*MOS test chip transistors. The second area of this paper employs a hot carrier induced degradation stress method to examine the damage mechanisms at low gate voltages, medium gate voltages, and high gate voltages. Finally a physical method for predicting the lifetime of a device containing a gate oxide short within the channel is presented.

2. EXPERIMENTAL PROCEDURE

2.1 The Test Transistor

The test transistors used in this work are polysilicon gate technology, with 4 μm metal line widths and 3 μm polysilicon widths (4/3 technology) (Ref. 1). This technology has a nominal 0.6 μm thick, phosphorus doped polysilicon gate with a resistance of 15-25 ohms per square. The gate silicon dioxide (SiO_2) was formed in dry O_2 at 1000 °C for thirty minutes. The *n*-doping concentration of the polysilicon gate is approximately 10^{17} cm^{-3} at the upper surface and decreases to 10^{14} at the SiO_2 surface. The gate oxide thickness is approximately 450 Å. The *n*-MOS transistors are formed in a 6 μm *n*-epitaxial layer, grown on a 25 mil n^+ substrate. The *p*-wells are 5 μm deep and were formed by boron implantation. All source and drain diffusion depths are about 0.5 μm and the operational range for V_{DD} is 5 to 10 volts. The n^+ source/drain to *p*-well junction avalanche breakdown occurs at about 17 volts. The *n*MOS test chip transistors had a drawn gate width of 16 μm and drawn

gate lengths of 2 μm, 3 μm, and 4 μm. All four contacts of the defective transistors were isolated.

2.2 Damage Technique

Since naturally occurring gate oxide shorts can be difficult to detect and analyze in complex VLSI ICs, the work presented here uses intentionally induced gate oxide shorts. The gate oxide short defects were created using a xenon laser cutting instrument (Florod Corp.). The laser pulse was passed through a small rectangular aperture (< 1 μm minimum dimension) and focused onto the polysilicon gate. The laser damage method was chosen because it produces gate oxide short defects electrically similar to manufacturing defects (Ref. 1).

2.3 Test Procedure

A block diagram of the test procedure is shown in Figure 1.

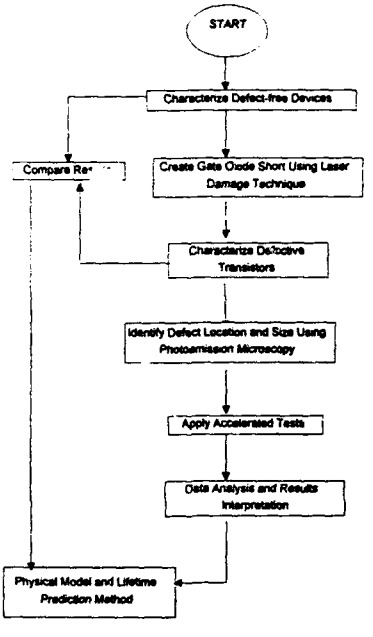


Figure 1. Block diagram of the test procedure.

3. ELECTRICAL PROPERTIES OF GATE OXIDE SHORTS IN nMOS TRANSISTORS

To establish a basis for identifying the damage mechanisms present in an nMOS device containing a gate oxide short, the electrical properties associated with this defect type were first examined. A comparison of transistor parameters before and after the rupture of the gate showed that substrate current and gate current were the dominant indicators of the presence of a gate oxide short. Although other measured parameters changed significantly, these changes were inconsistent.

3.1 Substrate Current (I_B)

When measuring the substrate current, I_B, the source and substrate terminals were fixed at zero volts, while the drain was fixed at 5 V. Gate voltages ranged from -2 V to 5 V. A potential drop of 5 V between the source and drain ensured a maximum substrate current while the gate voltage was varied. Figure 2 shows the test condition schematic used in substrate current (I_B) measurements.

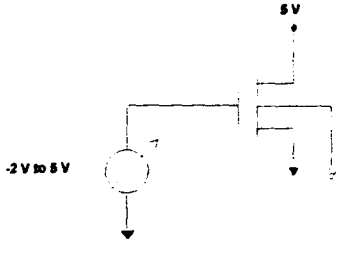


Figure 2. Test condition schematic for substrate current measurements.

A transistor cross section showing the substrate current is shown in Figure 3.

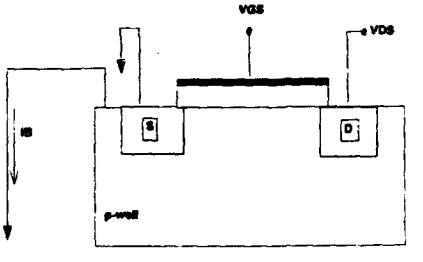


Figure 3. Cross section of a transistor showing the substrate current (I_B).

When a gate oxide short is present in the channel region, the resulting substrate current displayed a diode-like (rectifying) characteristic, corresponding to the *pn* junction that is formed between the *p*-well and the *n*-doped polysilicon gate. A typical I_B vs V_G curve is shown in Figure 4. For negative gate voltages lower than -0.6 to -0.7 V, the junction is strongly forward-biased and high current is observed. The substrate current contribution for a positive gate voltage is not due to the gate oxide short and was also observed in non-defective transistors. Changes in the substrate current, although consistent, proved to be significant only at negative gate voltages, which are not typically used in normal CMOS *IC* operation.

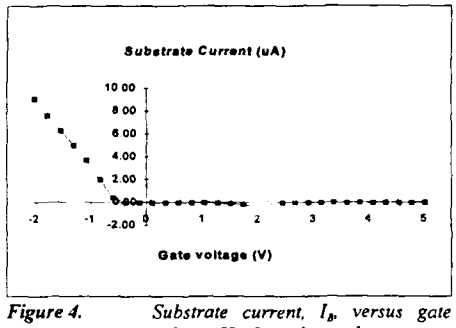


Figure 4. *Substrate current, I_B, versus gate voltage, V_G for a device that contains a GOS.*

3.2 Gate Leakage Current (I_G)

Gate current (I_G) is an important indicator of potential oxide defect mechanisms. The gate current characteristic in a defective device proved to be not only a reliable indicator of the presence of a gate oxide short but also identified the defect's general location. Measurements of I_G in the defective transistors resulted in five distinct types of I_G versus V_G defect characteristics. However, only defects that produced non-linear gate current (connection between the *n*-doped polysilicon gate and the *p*-well) will be discussed in this paper. A typical I_G-V_G characteristic for such defect is shown in Figure 5.

Quadrant III in Figure 5. shows a sharp increase in I_G when V_G is negative. This increase occurs between -0.5 and -0.7 V, which indicates that the creation of a gate oxide short caused a connection between the *n*-doped polysilicon gate and the *p*-well. Thus, the increase in I_G with negative gate voltages is a result of the forward biased diode current. Quadrant I in Figure 5 shows the reverse bias state of the "defect junction" with a soft breakdown current beginning at V_G = 1.8 V. This breakdown is well below the avalanche breakdown voltage (17 V) of normal *pn* junctions in this technology. The 100 *u*A current limit was the programmed value used for the HP 4145A.

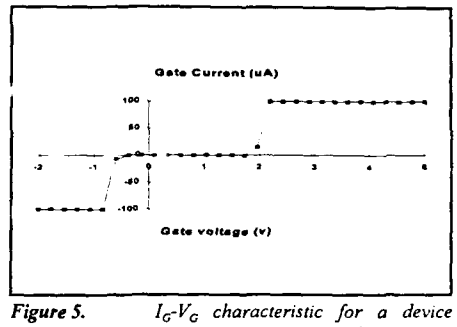


Figure 5. *I_G-V_G characteristic for a device containing a gate oxide short.*

4. EVALUATION OF HOT CARRIER INDUCED DEGRADATION IN *n*MOS DEVICES CONTAINING GATE OXIDE SHORTS.

Hot-carrier-induced transistor degradation arises from the high energy acquired by channel carriers, either electrons or holes, as they move from the source to the drain (Ref. 14). Hot electrons are emitted in *n*MOS transistors from either the silicon substrate or the surface channel into the gate oxide when channel field strength is sufficiently large. Some of the carriers may acquire enough energy (3.7 eV) to overcome the Si/SiO_2 barrier and pass into the gate dielectric. Subsequent trapping of the carriers injected into the oxide can cause instabilities in the form of transconductance degradation and threshold voltage drift with time.

it is proposed that *n*MOS transistor degradation due to the presence of a gate oxide short is analogous to the hot-carrier charging phenomenon occurring at the defect site. When a gate oxide short forms, the regular nature of the insulator is disrupted, thus creating defects within the oxide. These defects generate discrete electronic energy states, or traps, within the forbidden energy band of the oxide. These allowed energy states can act as dynamic recombination centers, capturing both electrons and holes. Therefore, the presence of a gate oxide short increases the number

of traps present in the oxide, as well as the electric field through which the carriers move. This phenomenon increases the energy of the channel carriers, thus enabling them to enter the dielectric. The filling and emptying of the traps can change the degree of degradation in the transistor characteristics and can accelerate the failure mechanism.

In order to characterize the transconductance and threshold shift, in a device containing a gate oxide short, as a function of the gate voltage, several devices containing a gate oxide short were stressed at different gate voltages (V_G = 0.5, 1, 1.5, 2, 3, and 4 V) and constant drain voltage, V_D = 8 V, for 500 seconds. This was followed by an electron injection phase at V_G = V_D = 8 V for 30 seconds. The transconductance change is shown in Figure 6 where each point on the curve represents one of the test transistors. Curve A represents the transconductance degradation after the initial stress and curve B represents the transconductance degradation after the electron injection phase (V_G = V_D = 8 V).

Figure 6 shows that the maximum damage (transconductance shift) after the first stress occurred at V_G = 2 V. This condition corresponds to the maximum substrate current, I_{Bmax}, and the maximum hot electron generation. After the electron injection phase, the maximum transconductance shift occurred at a lower gate voltage (around V_G= 1.0 V). Also, the magnitude of degradation was almost four times greater than that before electron injection. The difference between the two curves is due to created oxide traps.

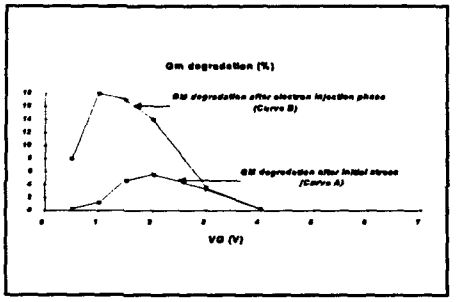


Figure 6. *% Transconductance change, ΔG_M versus gate voltage, V_G*

Figure 7 shows the threshold voltage, V_T, behavior as a function of applied gate voltage, V_G.

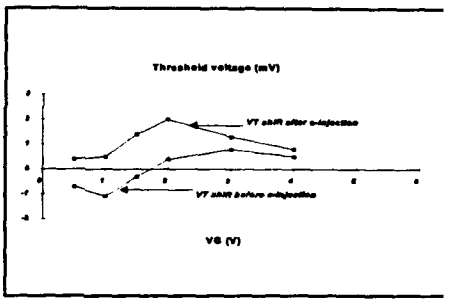


Figure 7. *Threshold voltage shift, ΔV_T, as a function of gate voltage, V_G*

It is clear, from Figure 7, that the threshold voltage was strongly affected by the electron injection phase. During the initial stress, a negative shift in the threshold voltage was measured for V_G < 2 V which indicates a positive trapped charge in the oxide. For $V_G \geq 2$ V, no positive charge was trapped, as indicated by the measured positive threshold voltage shift. After the electron injection phase (V_G = V_D = 8 V), the positive charge that was trapped during the first stress was compensated by injected electrons for V_G < 2 V.

Next, floating gate measurements (Ref. 15) were performed to correlate the observed damage in threshold voltages and transconductances to the gate current for a device biased at high drain voltage. This method uses the charge injected into the oxide to charge or discharge the gate, which is not biased (floated). From the rate of charging or discharging, the gate current and its sign can be determined. Figure shows the gate current, I_G, as a function of gate voltage, V_G, for a fixed drain voltage, V_D = 8 V.

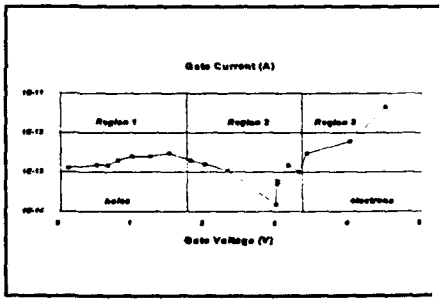


Figure 8. *Floating gate current as a function gate voltage.*

There are three major regions of the gate voltage range (Refs. 16-18) in which markedly different damage mechanisms occur. These three regions are distinguished by the charge injected into the oxide during hot-carrier stress.

Region 1: The low gate voltage region, within which holes are the predominant component of the gate current.

Region 2: The medium gate voltage region, where both electrons and holes are injected in approximately equal numbers.

Region 3: The high gate voltage region, where electrons are the main gate current species.

In Region 1, the predominant damage mechanism is due to generation of electron traps in the bulk oxide by the injected holes, N_{ot}. These electron traps are initially neutral, which precludes observation of their effects. However, if the stressed transistors (Curve A in Figures 6 and 7) were injected with electrons (brief stress with $V_G = V_D = 8$ V), the neutral traps became charged and contributed to the degradation (Curve B in Figure 6 and 7). The charged traps caused the threshold voltage shift values to change from negative to positive (Figure 7) and caused a significant increase in the transconductance (Figure 6).

Region 2 of Figure 8 represents the gate voltage condition for injection of both electrons and holes into the oxide. This condition is normally interpreted as corresponding to the creation of interface states, N_{it} (Refs. 19-23). Therefore, the dominant damage mechanism is due to the generation of interface states, N_{it}, at medium gate voltages ($V_G = V_D/2$). Interface traps alone would produce a positive V_T shift due to mobility degradation (Ref. 20). Therefore, as the number of interface states increases with higher gate voltages, the effect of the original trapped positive charge is masked, which leads to a positive threshold voltage shift.

The third type of stress damage occurs at high gate voltages, or Region 3 of Figure 8, where the dominant damage mechanism is the generation of electron traps in the oxide by the injected electrons, N_{ot}. In this region, a positive threshold voltage shift was observed, which is slightly affected by the electron injection phase (Figure 7). Furthermore, the transconductance degradation remained the same before and after the electron injection phase for high gate voltages (see Figure 6).

5. LIFETIME PREDICTION METHOD FOR DEVICES CONTAINING A GATE OXIDE SHORT

The previous section demonstrated that the presence of a gate oxide short in devices causes damage in the form of interface trap creation, N_{it}, and oxide traps, N_{ot}, and N_{ot}, at the defect site. These mechanisms can eventually lead to premature circuit failure by decreasing the transistor lifetime. In order to develop an accurate method for predicting the lifetime of a device containing a gate oxide short, the damage mechanism, which is highly dependent on the voltage of operation, must be considered. Therefore, the following procedure is divided into three parts: low gate voltage device stressing ($V_G = V_D/4$), medium gate voltage device stressing ($V_G = V_D/2$) and high gate voltage device stressing ($V_D/2 < V_G < V_D$). The lifetime of a device is defined as the time necessary for a transconductance shift (ΔG_m) of 10%.

5.1 Low Gate Voltages Stress

The main damage mechanisms at low gate voltages result from the injection of avalanche-generated hot holes into the oxide. These injected holes can create neutral electron traps and hole traps (Refs. 17, 18, 24). The hole traps do not significantly affect transistor lifetime estimates and will not be discussed. Initially, it was not evident that electron traps were being created at low gate voltage stress. Since the traps were neutral, their effects on the current versus voltage characteristics were not observed. However, when the stressed transistors were injected with electrons (brief stress with high V_G equal to V_D), the neutral traps became charged and contributed to the degradation, as seen in the previous section. Figure 9 shows the transistor lifetime (τ_{low}) for oxide trap damage created at low gate voltages as a function of the ratio of substrate current to drain current (I_B/I_D).

It can be seen that all devices lie approximately along a straight line that obeys

$$\frac{1}{\tau_{low}} = A \left(\frac{I_B}{I_D}\right)^n \qquad (1$$

where τ_{low} is the transistor lifetime for low gate voltage stress. I_B and I_D are the substrate and the drain currents during stress. The constants A and n wi

determined by fitting the experimental data, the value of n was calculated to be ~4.1

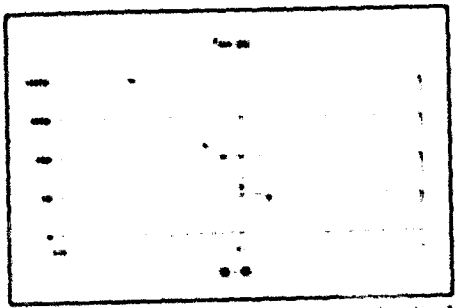


Figure 9. Lifetime τ_{max} versus ... for 3 devices

5.) Medium Gate Voltages Stress

The damage mechanism that occurs in the medium gate voltage region is mainly caused by interface state generation. In this region, a direct relationship exists between the degradation and the peak substrate current (Ref. 25). Experimental results showed that the closer the defect is to the drain region, the greater is its effect on the lifetime of the transistor.

The amount of degradation due to interface state generation peaks at the same gate voltage as the peak in substrate current. Thus, it is quite possible to predict the lifetime of the device due to the presence of interface states (τ_{max}) using the substrate current as an indicator. Figure 10 shows the relationship between τ_{max} and the substrate current for four devices.

The transistor lifetime under stress conditions that generate interface states can be experimentally expressed by

$$\frac{1}{\tau_{max}} = B\, I_p^{n} \qquad (2)$$

Where τ_{max} is the transistor lifetime for medium gate voltage stress, and I_p is the maximum substrate current. The constants B and n are established by fitting the experimental data.

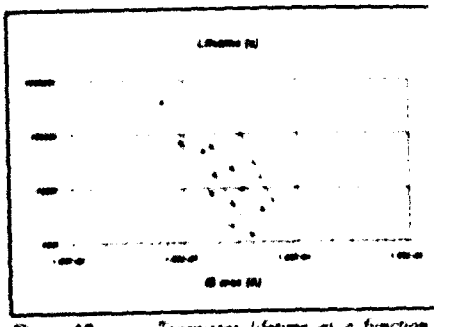


Figure 10. Transistor lifetime as a function of the peak substrate current I_{max}

6.) High Gate Voltages Stress

The lifetime prediction method developed here for hi gate voltage stress is based on the gate current, as it is a more meaningful quantity to use at high ga voltages than the substrate current, which was used to predict interface trap lifetimes (Refs. 16, 25). T reason for this is that the damage in the high-ga voltage region is due to electron traps created by electron injected into the oxide (Refs. 15, 26).

Figure 11 shows a plot of lifetime as a function of gate current (the lifetime criterion was taken to b 10% G_m change) for the two test devices. The str voltages have been varied from 6.5 V to 8 V, and ratio of (V_G:V_D) maintained at 1:0 (i.e., $V_G = V_D$) in cases.

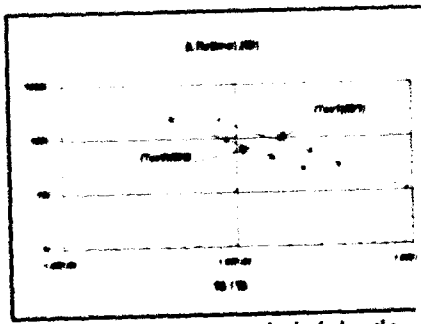


Figure 11. Lifetime multiplied by the current, $\tau_{max} I_p$, as a function of

The data in Figure 11 form straight lines for each of the two transistors, with approximately equal gradients. Thus, the transistor lifetime at high gate voltage stress can be experimentally expressed as

$$\frac{1}{(\tau_{\text{stress}})I_D} = C\left(\frac{I_G}{I_D}\right)^p \qquad (3)$$

Where τ_{stress} is the transistor lifetime under high gate voltage stress. I_D and I_G are the drain and gate currents during stress. The constants C and p are determined by fitting the experimental data.

It should be emphasized here that the above relationship only applies when $V_G = V_D$ (or rather, under conditions of constant gate field between gate and drain). Since I_G depends on both the field in the silicon and the field in the oxide, changing the gate-to-drain voltage ratio also changes the oxide field, and makes the barrier for electrons in the oxide higher. Also, the type of damage changes from oxide traps to *interface traps* as the ratio V_G/V_D approaches 0.5.

5.4 Transistor Lifetime Prediction Method

The contributions of the three damage mechanisms to the device's lifetime can be calculated by integrating equations 1 through 3 over one period, T. The resulting expressions for the lifetimes due to each of the three physical mechanisms are given by

$$\frac{1}{\tau_{\text{dynamic}}} = \frac{A}{T}\int_0^T \left(\frac{I_B}{I_D}\right)^n dt \qquad (4)$$

$$\frac{1}{\tau_{\text{Nox}}} = \frac{B}{T}\int_0^T I_B^m \, dt \qquad (5)$$

$$\frac{1}{\tau_{\text{Nox}\beta}} = \frac{C}{T}\int_0^T \left(\frac{I_G}{I_D}\right)^p I_D \, dt \qquad (6)$$

Treating $1/\tau$ as a damage function, the total dynamic stress damage can be expressed as

$$\frac{1}{\tau_{\text{dynamic}}} + \frac{1}{\tau_{\text{Nox}\beta}} + \frac{1}{\tau_{\text{Nox}}} + \frac{1}{\tau_{\text{Nox}\beta}} \qquad (7)$$

Note that equation 7 takes into account the actual stress conditions to which MOSFETs are subjected under normal circuit operation. It takes into account the three damage mechanisms described in the previous sections.

6. CONCLUSIONS

This study provided a better understanding of the damage mechanisms in a device containing a gate oxide short that lead to the gradual degradation of its characteristics and to eventual transistor failure. The results of this work provide several important insights regarding the impact of a gate oxide short defect on transistor performance. Degradation of the electrical properties in *n*MOS transistors from gate oxide shorts was found to be dependent upon both oxide trapping and interface state generation. The presence of a gate oxide short reduces the device's lifetime, which jeopardizes the reliability of the integrated circuit. The damage mechanisms associated with gate oxide shorts, as well as the stress conditions that generate these mechanisms, were identified.

Note:
The current address for Samar Dababneh is :
P. O. Box : 20095
Amman - Jordan
This work was done for the University of New Mexico as part of the requirements for the PhD program. Thesis title: Gate Oxide Shorts in *n*MOS Transistors: Electrical Properties and Lifetime Prediction Method.

REFERENCES

1. Hawkins, C.F. and J.M. Soden, "Electrical Characteristics and Testing Considerations for Gate Oxide Shorts in CMOS ICs," Proc. Int. Test Conf., 1985, pp. 544-555.
2. Soden, J.M. and C.F. Hawkins, "Test Consideration for Gate Oxide Shorts in CMOS ICs," IEEE Design and Test, August 1986, pp. 56-64.
3. Yamabe, K. and K. Taniguchi, "Time Dependent Dielectric Breakdown of Thin Thermally Grown SiO_2 Films," IEEE Trans. Electron Devices, Vol.

ED-32, February 1985, pp 423-428

4 Hawkins, C F and J M Soden, "Reliability and Electrical Properties of Gate Oxide Shorts in CMOS ICs," Proc Int Test Conf, September 1986, pp 300-309

5 DiStefano T H and M Schatzkes, "Impact Ionization Model for Dielectric Instability and Breakdown," App Phys Lett, Vol 25, December 1974, p 685

6 Holland, S, I C Chen, T P Ma and C Hu, "On Physical Models for Gate Oxide Breakdown," IEEE Electron Devices Lett, Vol EDL-5, August 1984, pp 302-305

7 Maly, W, F J Ferguson, and J P Shen, "Systematic Characterization of Physical Defects for Fault Analysis of MOS IC Cells," Int Test Conf Proc, 1984, October 1984, pp 390-399

8 Abraham, J A, Fuchs, W K, "Faults and Error Models for VLSI," Proc IEEE, Vol 74, No 5, May 1986, pp 639-654

9 Al-Arian, S A, Agrawal, D P, "Physical Failures and Fault Models of CMOS Circuits," IEEE Trans on Circuits and Systems, Vol CAS-34, No 3, March 1987, pp 269-279

10 Banerjee, P, Abraham, J A, "Generating Tests for Physical Failures in MOS Logic Circuits," Proc Int Test Conf 1983, pp 554-559

11 Favalli, M, Olivo, P, Damiani, M, Ricco, B, "Fault Simulation of Unconventional Faults in CMOS Circuits," IEEE Trans on Computer-Aided Design, Vol 10, No 5, May 1991, pp 677-682

12 Rajsuman, R, Malaya, Y K, Jayasumana, A P, "On Accuracy of Switch-Level Modeling of Bridging Faults in Complex Gates," Proc Design Automation Conf, 1987, pp 244-250

13 Storey, T M, Maly, W, "CMOS Bridging Fault Detection," Proc Int Test Conf 1990, pp 842-851

14 Ning, Osburn, Yu, "Effect of Electron Trapping in IGFET Characteristics," Journal on Electron Mater, Vol 6, 1977, pp 65-76

15 Farr, R B, Sun, R C, "Threshold-Voltage Instability in MOSFET's Due to Channel Hot-Hole Emission," IEEE Trans Electron Devices, Vol ED-28, 1981, pp 83-89

16 Hu, C, Tam, S C, Hsu, F, Ko, P, Chan, T, Terrill, K W, "Hot-Electron Induced MOSFET Degradation Model, Monitor, and Improvement," IEEE Trans Electron Devices, Vol ED-32, 1985, pp 375-385

17 Bellens, P, Heremans P, Groeseneken, G, and Maes, H E, "Hot Carrier Effects in n-Channel MOS Transistors Under Alternate Stress Conditions," IEEE Electron Device Lett, Vol EDL-9, 1988, pp 232-234

18 Doyle, B, Mistry, K, "The Generation and Characterization of Electron and Hole Trap Created by Hole Injection During Low Gate Voltage Hot Carrier Stressing in n-MOS Transistors," IEEE Trans Electron Devices, Vol ED-37, 1990, pp 1869-1879.

19 Weber, W, C Werner, and G Dorda "Degredation in n-MOS Transistors After Pulsed Stress," IEEE Electron Devices Lett, Vol EDL-5 1984, pp 518-520

20 Hsu, Tam, "Relationship Between MOSFET Degredation and Hot Electron Induced Interface State Generation," IEEE Electron Devices Lett Vol EDL-5, 1984, pp 50-52

21 Takeda, Shimizu, Hagiwara, "Role of Hot-Hole Injection in Hot-Carrier Effects and the Small Degraded Channel Regions in MOSFETS," IEEE Electron Devices Lett, Vol EDL-4, 1983, pp 329-331

22 Hsu, Grinolds, "Structure-Enhanced MOSFET Degredation Due to Hot Electron Injection," IEEE Electron Devices Lett, Vol EDL-5, 1984, pp 71-74

23 Hsu, Chiu, "Temperature Dependence of Hot Electron Induced Degredation in MOSFETs, IEEE Electron Devices Lett, Vol EDL-5, 1984 pp 148-150

24 Reply to "Comments on 'The Generation and Characterization of Electron and Hole Trap Created by Hole injection During Low Gate Voltage Hot Carrier Stressing in n-MOS Transistors'" by Doyle, B et al, IEEE Trans on Electron Devices, Vol ED-39, 1992, pp 460-464

25 Takeda, E, Suzuki, N, "An Empirical Model for Device Degradation Due to Hot-Carrier Injection, IEEE Electron Device Lett, Vol EDL-4, No 4, April 1983, pp 111-113

26 Tsuchiya, T, Frey, J, "Relationship Between Hot Electrons/Holes and Degradation of p- and n-Channel MOSFET's," IEEE Electron Device Lett Vol EDL-6, January 1985, pp 8-13

Reliability Issues in New Technology Implementation

R.J.Francey

GEC Plessey Semiconductors, Lincoln, England

No text available

THE 3-D TURN-OFF PROCESS OF GTO THYRISTORS. DESCRIPTION OF FAILURE USING MEASUREMENTS AND SIMULATION

H. Bleichner, K Nordgren[1], M. Rosling,
P. Jonsson and E. Nordlander[1]
Scanner Lab, Department of Technology, Uppsala University,
P.O. Box 534, S-751 21 Uppsala, Sweden
tel. +46 18 183146 fax. +46 18 555095

Abstract - Measurements based on a time-resolved free-carrier absorption (FCA) technique are used to map the local carrier densities two-dimensionally at any time of the switching cycles. Inductively loaded GTO:s under snubberless operation close to the safe operating area, SOA, are studied.

The results of the experimental research give visible evidence of the relationship between *e. g.* the turn-off gain, G, and an uneven development of the field region in the blocking junction of the device. The irregularity of the field region is denoted in this work as a quasi space-charge region (QSC). During the fall-time period of the turn-off, the QSC expands towards the anode emitter, and causes a local enhancement of the anode emitter injection *viz.* local dynamic punch-through in the n base. As a consequence of this punch-through mechanism, a high peak of excess holes is built up in the *p* base and reaches a maximum value at the beginning of the tail period. This charge depolarizes the cathode junction locally, and current filaments connecting the cathode and anode sides of the device is formed.

It is a well-know fact that the use of a highly doped *p* base leads to a reduction in current gain of the *n-p-n* transistor portion of the GTO. Thus, a reasonable conclusion is that a higher peak of excess holes in the *p* base is needed to re-trigger the device, and, consequently, the SOA range is enlarged.

1. INTRODUCTION

High-power gate-turn off (GTO) thyristors are becoming the commercial key to power control. Specific applications are *e. g.* electric trains and power transmission systems. The device characteristics of such switches are similar to those of a thyristor (SCR) except for the controlled turn-off feature by means of gate operation, thus making otherwise needed commutation circuits excessive.

Operating an inductively loaded semiconductor power device may be associated with dynamic breakdown phenomena. In the case of GTO thyristors, the dynamic range of turn-off operation is defined by the the safe-operating area (SOA). The SOA, however, is only an empirical measure of the ultimate conditions for allowable operation, but the breakdown mechanisms involved in these limits may not be well known. Questions to be raised in this aspect would certainly comprise what process steps or treatment could be performed in order to enlarge the SOA range.

[1] Gävle/Sandviken University College,
P.O. Box 504, S-811 25 Sandviken, Sweden.

In this paper, experimental results from free-carrier absorption (FCA) measurements paired with electrical measurements facilitate for the understanding of SOA limitations. The FCA technique is briefly described below, and a complete description is found in (Ref. 1). A 3.39-μm laser is scanning one of the free sample surfaces, *i. e.* a surface perpendicular to the anode and cathode surfaces. Due to the silicon transparency of light of this wavelength the laser beam is only probing free carriers injected into the structure, and not carriers present in thermal equilibrium. The photons of the laser beam are absorbed/scattered by the injected carriers, thus decreasing the light transmitted through the sample. Transmitted photons are detected by means of a photo diode, and the signal is processed in a digital oscilloscope before the data manipulation and presentation in a small computer. Fig. 1 is showing a typical output of the measurements, *i. e.* a carrier map, and for clarity the carrier map is placed on top of a schematic sample outline and a doping profile.

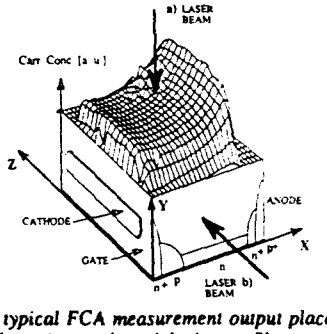


Fig. 1. A typical FCA measurement output placed on top of a schematic sample and doping profile.

Fig. 1 also contains two arrows showing the possibilities of measurement on an oblong sample like a GTO unit cell. The arrow angles are perpendicular, and the angle marked with an "a" represents a long-side measurement or a cathode-length measurement whereas the angle marked with a "b" represents a short-side measurement or a cathode-width measurement.

2. GTO TURN-OFF HAZARDS

The electrical characteristics of the GTO turn-off process is divided into three stages: *the storage period* – the interval between turn-off triggering and the beginning of the rapid fall of the anode current; *the fall-time period* – the rapid transition between high an low values of the anode current ending by definition when the cathode current changes sign (the turn-off point); and *the tail*

period – the interval succeeding the turn-off point. During these cycles different processes effect the internal carrier distribution of the device. Two phases of plasma squeezing occur (Refs. 2, 3). These phases are illustrated in Fig. 2.

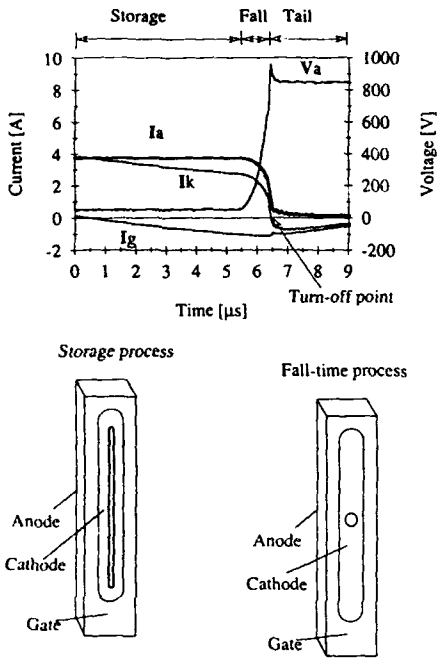


Fig. 2. Illustration of the 3-D plasma squeezing process.

The first phase of plasma squeezing is mainly observable in the cathode-width direction, and the speed of this process is controlled by the gate determining the length of the storage period. The second phase goes in the cathode-length direction, where the extension of the plasma is collapsing, thus making a current filament in the center region of the device. In previous publications, it was shown that this filament is modulating the electric field of the blocking junction of the device (Refs. 2–4). This area is denoted by us as the quasi space-charge region (QSC). As the QSC is approaching the anode junction, a local enhancement of the anode emitter injection is invoked. For increasing blocking voltages the QSC gets closer to the anode junction, thus increasingly connecting the anode to the *p* base. In this way, a local punch-through mechanism in the *n* base is acting as a first step of a chain of events promoting failure.

At the turn-off point, the electron supply of the cathode emitter is cancelled, and shortly hereafter the gate-cathode junction goes into avalanche. The number of holes entering the *p* base exceeds the number of holes extracted by action of the gate. Since the holes remain focused in the *p* base center region, a piling up of holes is initiated there. The piling up of holes in the *p* base is located to the center of the GTO cell, exactly at the place where the fall-time carriers previously were focused.

Now, if the hole concentration in the *p* base reaches the critical value for gate-cathode junction depolarization, the conditions are favorable for local re-triggering of the device. Since the excess hole charge is concentrated to a small region in the *p* base, device re-triggering will be of a local nature thus promoting current-filament formation, and this second event makes the turn-off failure of the GTO a fact.

3. EXPERIMENTAL RESULTS AND DISCUSSION

In order to verify the described plasma squeezing process, FCA measurements from the short side as well as from the long side of a GTO-thyristor unit cell have been performed. Fig. 3 is displaying a turn-off carrier map sequence of an anode-shorted GTO unit cell taken from the short side (*cf.* direction "b" of Fig. 1), and Fig. 4 is displaying the corresponding measurement taken from the long side (*cf.* direction "a" of Fig. 1).

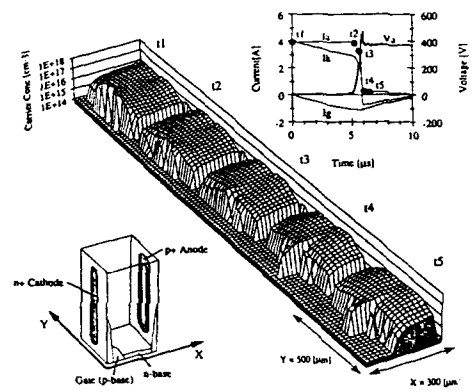


Fig. 3. Turn-off carrier-map sequence of an anode-shorted sample taken from the short side.

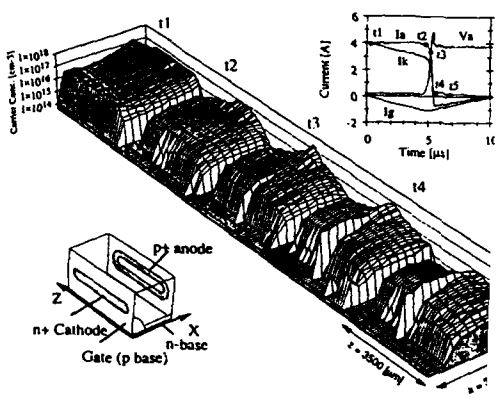


Fig. 4. Turn-off carrier-map sequence taken from a long side of the sample of Fig. 3.

Both Figs. 3 and 4 present measured sequences of the carrier distribution at different moments of the turn-off cycle, which also are marked by dots in the electrical characteristics of each figure. The QSC in these cases is represented by a ridge of carriers (filament) which is a consequence of the anode design. Anode design and orientation of the sample are also presented in Figs. 3 and 4. As can be seen in Figs. 3 and 4, the filament has the same shape independently of measurement direction. This observation indicates that the filament becomes cylindrical already at the end of the fall-time period, and it is active far in the tail period.

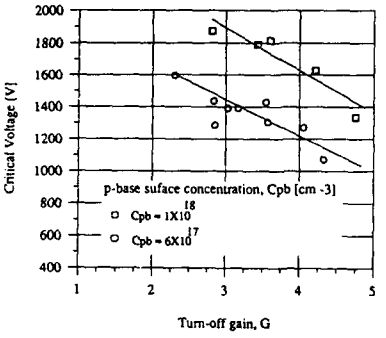


Fig. 5. SOA measurements on GTO samples of different p-base doping.

The QSC may be represented by a ridge of carriers in the carrier maps or by the opposite, *i. e.* an incision, depending of the anode design (Ref. 6). This has to do with the different anode injection efficiencies of the different anode designs. According to (Ref. 5), a relationship between the turn-off gain, G, and the QSC electric-field value is proposed. This relationship has also been experimentally observed and discussed (Refs. 2, 3). Recently, a careful investigation of this relationship was performed regarding both experiments and simulation (Ref. 6). In this work it is shown that the QSC is absent if the G value is close to unity. For higher G values, although the same blocking voltage is applied, the QSC expands deeper in the *n* base. Thus, the SOA limit becomes lower for higher G values than for the case of G close to unity. This effect is shown in Fig. 5 where the SOA limit is negatively sloping for increasing turn-off gain. Two sets of samples are presented here, each of different *p*-base doping. One of the sets is associated with a surface concentration of *p*-base dopants of $6 \cdot 10^{17}$ cm^{-3}, and the other of $1 \cdot 10^{18}$ cm^{-3}.

However, samples where the QSC is represented by an incision give a more accessible information about the *p*-base behaviour prior to a failure, see Fig. 6. In Fig. 6, an example of a non-shorted sample with an n^+ buffer layer adjacent to the p^+ anode was used. In this sequence, a *p*-base carrier peak is clearly visible, and is still increasing a few hundred nanoseconds in the tail period. The closer the blocking voltage comes to the critical value, the higher becomes the peak. It has been shown experimentally that the peak of carriers is almost unipolar, thus made up out of holes (Refs. 2, 7).

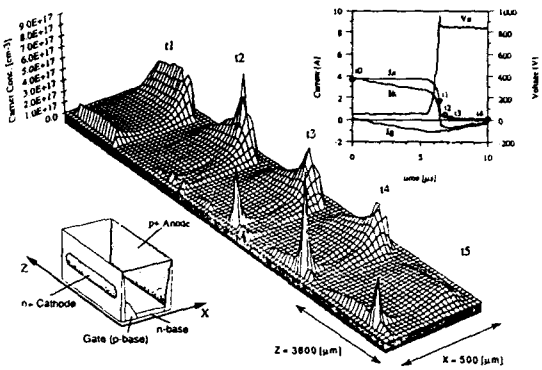


Fig. 6. Measured build-up of holes in the p-base of a non-shorted GTO sample operated close to the SOA limit.

In the case of a highly doped *p* base, the required concentration to depolarize the gate-cathode junction must be higher due to the reduction in current gain of the inherent *n-p-n* transistor portion of the GTO thyristor.

4. SIMULATION

Three-dimensional transient simulations were performed to illustrate the experimentally observed 3-D effects of the GTO turn-off process. A quarter of a non-shorted, n^+ buffered GTO unit cell was simulated using the 3-D device simulator *daVinci* by TMA. The results of the simulations of the device for a turn-off gain of 3, approximately, are shown as current-density contours in Fig. 7. The 3-D transient simulations confirm the formation of a high-current filament as discussed in previous sections.

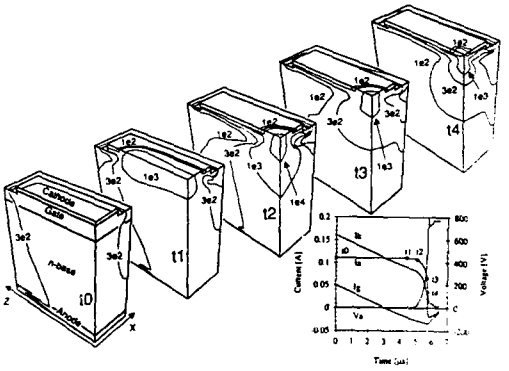


Fig. 7. 3-D turn-off simulation of a non-shorted GTO sample showing the current-density contours in A/cm^2.

In Fig. 7, the conducting area is collapsing into a cylindrical current filament during the transition between the storage period and the fall-time period, *cf.* the maps marked t1 and t2. Moreover, the simulations predict the survival of the high drift-current filament even after the

turn-off point (not shown in Fig. 7).
Although 3–D simulations were used to produce qualitative evidence for the current filamentation, and, hence, the device proportions were not outlined correctly, a far better agreement with experiments were obtained than using 2–D simulations. It should be clearly stated that 2–D simulations are inadequate for unprejudiced predictions of the GTO turn-off process.

5. CONCLUSIONS

The turn-off failure of GTO thyristors is dependent of the turn-off gain, G. If the G value is well exceeding unity a dynamic punch-through failure may occur. The failure mechanism is composed by a chain of events, which is summarized in the following:
When entering the turn-off fall time period an unevenly distributed electric field is developing in the blocking junction. If the blocking voltage is close enough to the critical value of the SOA, the field region approaches the anode junction. Thus, an enhanced anode injection takes place locally, and holes are supplied to the p base of the device. These holes are piled up and kept focused by action of the gate. If the amount of holes is large enough to depolarize the gate-cathode junction, local re-triggering will occur and a drift-current filament is formed. The current density of the filament is of such a magnitude that localized heating finally makes the failure destructive. Hence, monitoring the p-base peak of holes facilitates for failure studies.
If the p-base doping is increased, the ratio between the p-base peak and the p-base doping will decrease for a certain blocking voltage. Thus, higher p-base doping will increase the SOA range. In other words, the degradation of the n-p-n current gain of the GTO due to an increased p-base doping reduces the influence of a dynamic punch-through as regards turn-off failure.
The FCA measurements clearly show the 3–D nature of the turn-off process of a GTO due to the oblong shape of the device. Hence, 2–D simulations are insufficient for predictions of turn-off effects without knowledge of the effects in beforehand. Using 2–D simulators for GTO turn-off investigations would require several tricks in order to achieve a highly increased current density due to plasma squeezing in a direction which is not accounted for.
Even if 3–D simulation facilities are available one should bear in mind that:
– Simulation alone might not give the sought information by definition, because realistic dimensions of the device (or at least true proportions) will generate a tremendous amount of grid points for the numerical solution of the transport equations. In addition, some transport parameters have to be experimentally determined or correlated to device processing. Another uncertainty factor in simulation work is if the physical models reliable or, rather, to what extent the models are applicable.
– 3–D simulations are extremely time consuming. The simulations presented in this paper had to use weeks in CPU time in a powerful work station.
Thus, in order to achieve reliable knowledge of fundamental physical phenomena in a transiently operated GTO thyristor, 3–D simulations should be used together with 3–D experimental results from *e. g.* FCA measurements.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Swedish and Swiss ABB companies for their concern in this study. Dr. M. Bakowski is especially acknowledged for supplying the samples and for valuable discussion.
This work was partly financed by the Swedish Board for Technical Development (NUTEK), the Swedish programme for industrial research, and a co-operation programme with Apple Computer AB.

REFERENCES

1. Bleichner H., Nordlander E., Rosling M., and Berg S. "A Time-Resolved Optical System for Spatial Characterization of the Carrier Distribution in a Gate Turn-Off Thyristor (GTO)", IEEE Trans. Instrum. Meas., Vol. IM-39, No. 3, 1990, pp. 473-478.

2. Bleichner H., Bakowski M., Rosling M., Nordlander E., Vobecky J., Lundqvist M., and Berg S. "A Study of Turn-Off Limitations and Failure Mechanisms in GTO Thyristors by Means of 2-D Time-Resolved Optical Measurements", Sol.-St. Electron., Vol. 35, No. 11, 1992, pp. 1683-1695.

3. Bleichner H., Rosling M., Vobecky J., Bakowski M., and Nordlander E. "Measurements of Failure Phenomena in Inductively Loaded Multi-Cathode GTO Thyristors", IEEE Trans. Electr. Devices, Vol. ED-41, No. 2, 1994, pp. 251-257.

4. Vobecky J., Bleichner H., Rosling M., and Nordlander E. "A Time-Dependent Two-Dimensional Analysis of the Turn-Off Process in a Gate Turn-Off Thyristor (GTO)", IEEE Trans. Electr. Devices, Vol. ED-40, No. 12, 1993, pp. 2352-2358.

5. Nakagawa A. "A Time- and Temperature-Dependent Two-Dimensional Simulation of GTO Turnoff Process II-Inductive Load Case", Sol.-St. Electron., Vol. 28, 1985, pp. 677-687.

6. Bleichner H., Nordgren K., Rosling M., Bakowski M., and Nordlander E. "The Effect of Emitter Shortings on Turn-Off Limitations and Device Failure in GTO Thyristors under Snubberless Operation", subm. IEEE Trans. Electr. Devices.

7. Bleichner H., Rosling M., Vobecky J., Lundqvist M., and Nordlander E. "A Comparative Study of the Carrier Distributions in Dynamically Operating GTO:s by Means of Two Optically Probed Measurement Methods", Proc. ISPSD, 1990, pp. 246-251.

A CONTACTLESS ANALYSIS SCENARIO FOR THE INVESTIGATION OF THE DYNAMIC BEHAVIOUR OF POWER SEMICONDUCTORS: INTERNAL AND EXTERNAL LASER PROBING IN COMPARISON WITH COMPUTER SIMULATIONS

G. Deboy, H. Brunner, G.Sölkner
Siemens AG, Corporate Research and Development, Otto-Hahn-Ring 6, D-81739 München

W. Claeys, S.Dilhaire, V. Quintard
Université de Bordeaux I, IXL, 351, Cours de la Libération, F-33405 Talence Cedex

F. Koch
Technische Universität München, Physics Department E16, James-Franck-Str., D-85748 Garching

ABSTRACT

A novel laser technique for the internal probing of carrier and temperature gradients is presented, which allows for the first time the measurement of absolute temperatures inside the low-doped region of power semiconductors. The absolute carrier concentration is measured by absorption, and the dilatation of the device by external laser interferometry, respectively. Experimental results are compared with computer simulations.

1. INTRODUCTION

Recent trends for power semiconductors towards a shrinking of the device size, with a subsequent increase in current density as well as the demand for improved switching characteristics require a detailed comprehension of the internal electrical and thermal transport phenomena within the device. The numerical solution of the Poisson equation and the semiconductor transport equations is a well established procedure. Nevertheless only a few experimental techniques have been proposed to verify the results achieved by the device simulation. Both the measurement of free carrier absorption (Refs. 1-2), and the electron-hole recombination radiation (Ref. 3) yield information on the carrier concentration. However, dynamic effects which are important for the device functioning, like carrier density gradients that occur during switching operation, are difficult to observe. The local temperature, being a keypoint for the design of power semiconductors, cannot be studied.
This paper presents a novel approach to detect gradients of both carrier concentration and temperature with high temporal and spatial resolution within power semiconductors during transient operation. Transport phenomena under strong non-equilibrium conditions in semiconductors are therefore accessible for experimental studies.
This technique is accompanied by free-carrier absorption measurements to determine the absolute carrier concentration, and by measurements of the thermal expansion with a highly sensitive laser interferometer. The experimentally determined dynamic behaviour of the device is subsequently compared with the computer simulation.

2. INSTRUMENTATION

Carrier densities and temperatures within a semiconductor material are probed with an infra-red laser beam incident on the side face and transilluminating the bulk of the device. The detection principle is based on the fact that the local index of refraction of a semiconductor is dependent on the free carrier concentration, as well as on the temperature. Gradients in both carrier density and temperature will therefore lead to a gradient in the refractive index, which in turn will deflect the transmitted laser beam. The principle is outlined in Figure 1. A position sensitive photodetector is used to measure the probe light deflection, thus allowing a quantitative determination of charge and temperature gradients within an operating power semiconductor device.

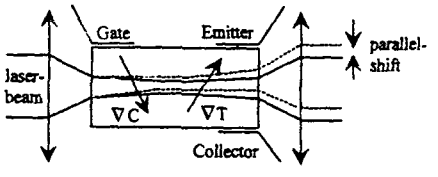


Figure 1: The measurement principle used for internal laser probing.

Furthermore, the free carrier absorption can be measured simultaneously with the same probe, yielding results on the carrier density.
It is worth mentioning that the detected charge and temperature gradients are due to the electrical operation of the device under test; in contrast to hitherto published work (Refs. 4-5) an optical excitation is not required.

The measurement result represents a line integral of carrier and temperature gradients along the beam path, rather than a value at a specific location. By scanning the beam, the local electronic and thermal transients can be mapped. The spatial resolution is given by the effective probe beam diameter. Typically a 3 mm diameter device allows for an average beam width of 30 µm.
In the present measurement set-up the laser probe wavelength was chosen to be 1320 nm, where the absorption spectrum of silicon has a minimum for a wide range of dopant concentration (Ref. 6). The laser source is operated in a continuous mode so that the frequency bandwidth of the measurement is given by the detector response time, which at present is around 1 µs. This is adequate to resolve fast transients in carrier concentration for most power devices. In principle, time resolution can be enhanced by sampling with a pulsed light source. However, for reasons of noise reduction, this would require fast synchronous switching of the power device, which is hard to achieve in practice.

The dynamic thermal expansion of the device during an applied current pulse is measured with a highly sensitive Michelson interferometer (Refs. 7-9). The lateral resolution depends on the one hand on the diameter of the focus, which is in the order of 1 µm, and on the other hand on the thermal range of the generated heat. Surface displacements can be measured with a resolution of 10^{-6} nm/√Hz and a typical time response of 200 ns. The laser beam can be positioned either on the lateral or on the top side of the power device. A wavelength of 632.8 nm is used so that light penetration into the semiconductor is negligible.
The local thermal dilatation itself is related to a geometry dependent integral of local temperatures and thermal expansion coefficients within the device. Together with thermal gradient measurements by device-internal probe beam deflection, it forms a complementary set of experimental techniques. These allow the observation of time-dependent thermal effects and also the accurate assignment of absolute local temperature values, which so far were hardly accessible by any other experimental method.

The numerical modelling of transient current and heat transport within the power semiconductor was performed using the device simulator MEDICI, which solves the fully coupled system of Poisson, carrier transport, continuity and heat flow equations. Technological parameters like doping profiles and carrier life times are implemented in the simulation. The mobility in the low doped region is described by the model of Dorkel and Leturcq (Ref. 10), taking into account effects of lattice, ionized impurity, and carrier-carrier scattering. The heat generation in the device is determined by the Joule and the recombination term. The thermal boundary condition at the collector electrode is represented by a thermal resistance to a heat sink kept at 300 K. The external electric circuit is described by lumped resistors and inductance elements. The simulator yields transient solutions for the local carrier concentration and the temperature.

3. EXPERIMENTAL RESULTS

As a typical representative of the new generation of power semiconductors, a 1200 V-IGBT is being investigated. A current load of 70 A/cm² is applied during 70 µs. Figure 2 represents a cross section of a single device cell with the probing positions of the internal measurements. The electron current is injected from a surface MOS channel structure and is driven vertically through the n^--layer towards the collector. An equivalent number of holes is injected from a thin backside p^+-layer in order to enhance the conductivity in the on-state.

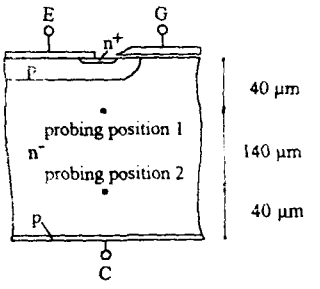


Figure 2: Cross section through a cell of the investigated 1200 V-IGBT, indicating the positions for internal laser probing.

In Figure 3 the absolute carrier concentration, measured by free-carrier absorption, is compared to computer simulation.

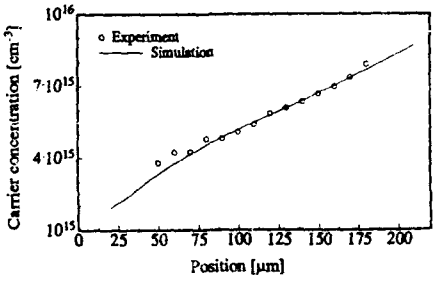
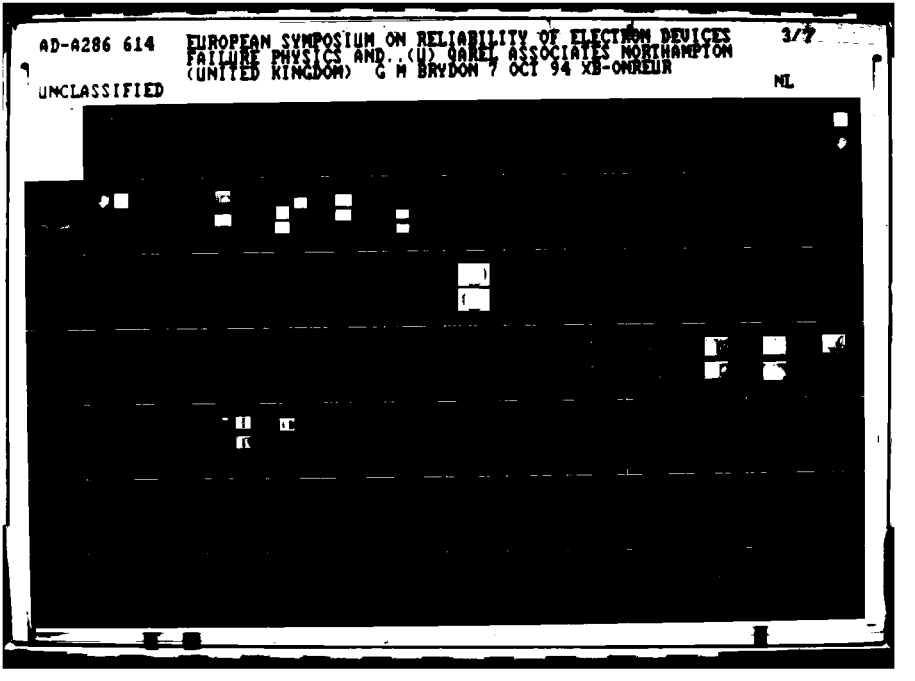
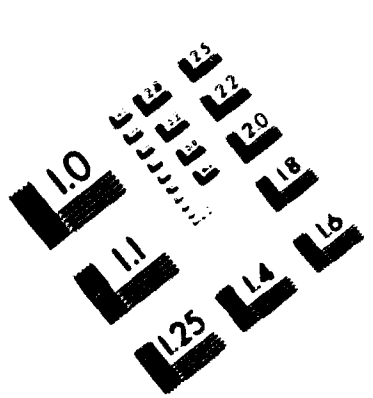


Figure 3: Comparison of absolute carrier concentration, measured by absorption, to computer simulation.

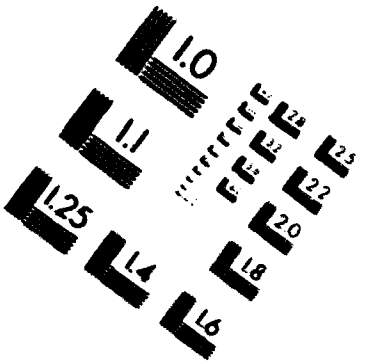
AD-A286 614 EUROPEAN SYMPOSIUM ON RELIABILITY OF ELECTRON DEVICES FAILURE PHYSICS AND..(U) QAREL ASSOCIATES NORTHAMPTON (UNITED KINGDOM) G M BRYDON 7 OCT 94 XB-ONREUR 3/7

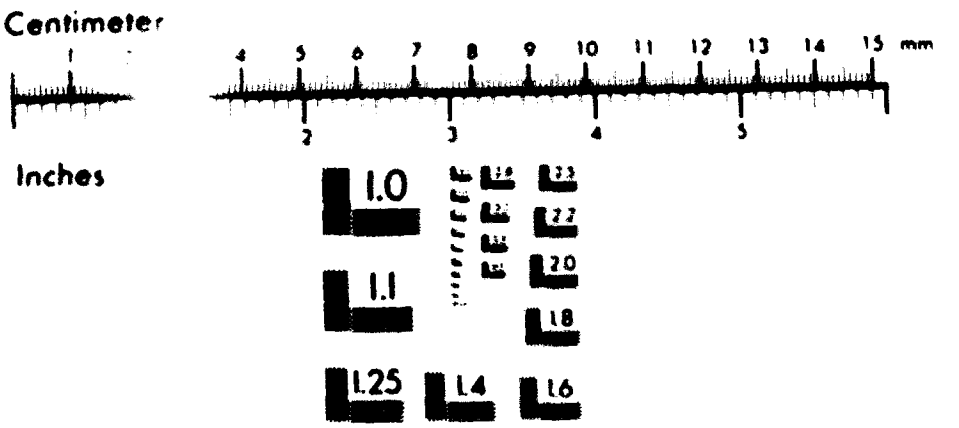
UNCLASSIFIED NL

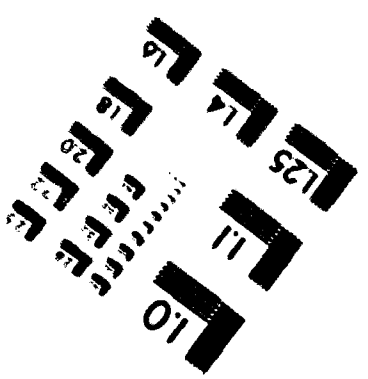




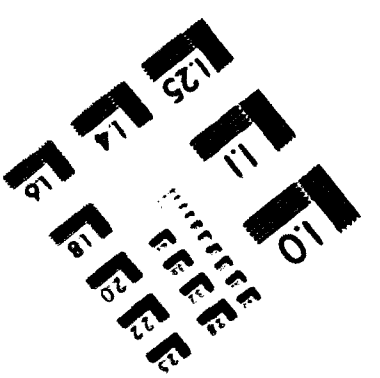
Association for Information and Image Management







MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.



The experimental data is transformed into absolute carrier concentrations on the basis of the Drude theory (Ref. 1). The crystal potential is referred to by effective masses of both electron and holes, and the carrier mobility is described by the same model, as used in MEDICI (Ref. 10). The variation of the carrier concentration along the light path, especially at the border of the active region of the device, is taken into account.

Figure 4a shows time resolved deflection signals obtained at the two typical probing positions indicated in Figure 2. At probe position 1, a strong thermal gradient is observed due to the large amount of Joule heating within the channel region. The steep rise of the signal at turn-on and the peak at turn-off are due to the additional amount of heat dissipated during switching. At position 2, where the thermal gradient has only weak influence, steep carrier gradients are observed during turn-on and turn-off at time t=0 µs and t=70 µs, respectively. With the carrier distribution reaching its equilibrium the carrier gradient slightly decreases. During turn-off the direction of the gradient is reversed, and the deflection signal changes its sign accordingly.

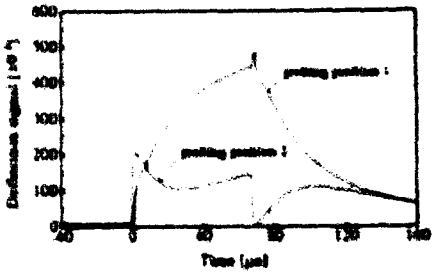


Figure 4a: Time resolved deflection signals measured by internal laser probing at two representative positions.

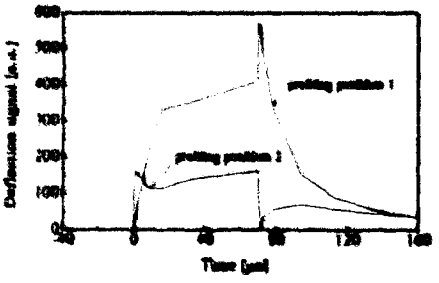


Figure 4b: Calculation of deflection signals on the basis of the carrier concentration and the temperature distribution yielded by computer simulations.

The calculation of the deflection signal, based on the local temperature and carrier density as extracted from MEDICI, represents all prominent features of the experimental measurements, see Figure 4b. The good qualitative agreement between experiment and simulation indicates that the major device-internal electronic and thermal effects are well understood, and are correctly implemented within the theoretical models.

The local carrier density as derived from absorption measurements and theoretical models, allows the evaluation of local gradients of the index of refraction, and hence a prediction of the beam deflection signal. This in turn can be used to calibrate the experiment. Further work towards a quantitative analysis of the deflection signal is in progress.

If boundary effects, e.g. reflectivity changes at the side faces, are negligible, the absorption signal is independent of temperature. Thermal effects are therefore detected by the deflection signal only. As can be seen from Figure 4a, electronic contributions to the deflection signal are in general much faster, thus enabling a separation of those two effects by time resolution.

Figure 5 shows a plot of the thermal contribution of the deflection signal versus the device thickness. Data points were taken 70 µs after turn-on, thus representing a "snap shot" of the heat transport at this time. On the one hand, an internal temperature gradient can be calculated for each data point, yielding the local temperature by a simple spatial integration procedure. In the present example a surface temperature rise of 5.1 K has been obtained.

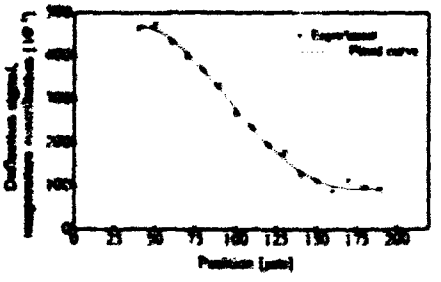


Figure 5: Thermal contribution of the deflection signal, achieved by internal laser probing, plotted versus the device thickness.

On the other hand, time resolved thermal expansion measurements by laser interferometry serve as the required cross-reference for the assignment of local temperature values. In Figure 6 the vertical dilatation across the top surface of an IGBT is shown 70 µs after turn-on. On this time scale, heat

transfer to external heat sinks is still negligible, so that the dissipated power is stored in the bulk of the device. The thermal expansion, however, is modulated by thermomechanical stress, occurring especially at the border of the active region. We therefore observe a curved distribution of the dilatation along the device diameter, in contrast to a more homogenous temperature gradient, achieved by scanning the internal probe along the surface.

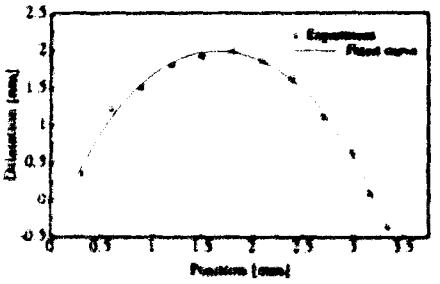


Figure 6: Absolute surface displacement of the IGBT measured by laser interferometry

Calculating a thermal expansion from the local internal temperature gradients (see Fig. 5), we obtain a value of 0.96 nm which is in excellent agreement with the measured average dilatation of 1.1 nm, but differs substantially from the value achieved by computer simulation. Further investigations towards an improved modelling of the heat generation is in progress.

4 CONCLUSION

A novel laser probing for the detection of gradients of the carrier concentration and the temperature inside operating power semiconductors has been introduced. In comparison with computer simulation, the observed effects, which yield information on the device internal behaviour, can be well understood.
For the first time, absolute values of the local temperature inside the device can be measured. Excellent agreement with external thermal expansion measurements by laser interferometry is achieved.

ACKNOWLEDGEMENTS

The continuous support and encouragement by E. Wolfgang is gratefully acknowledged. Thanks are also given to H. Mahrer for the device mounting, J. Otto for experimental advice and help, and J. Muggeridge for revising the manuscript.
A part of this work is supported by a NATO grant.

REFERENCES

1 C.M. Horwitz, and R.M. Swanson, "The optical (free-carrier) absorption of a hole-electron plasma in silicon", Sol St Electr., No 23, 1980, pp. 1191-1194

2 D.E. Houston, M.S. Adler, and E.D. Wolley, "Measurement and analysis of carrier distribution and lifetime in fast switching power rectifiers", IEEE Trans. Electr. Dev. ED-27, No. 7, 1980, pp. 1217-1222

3 J. Krausse, "Messung der Ladungsträger-konzentrationsverteilung im Mittelgebiet eines legierten Silizium-pin-Gleichrichters bei Belastung in Durchlassrichtung", Sol St Electr., No. 15, 1972, pp. 841-848

4 W.B. Jackson, N.M. Amer, A.C. Boccara and, D. Fournier, "Photothermal deflection spectroscopy and detection", Appl. Optics, No. 20(8), 1981, pp. 1333-1344

5 D. Fournier, C. Boccara, A. Skumanich and, N.M. Amer, "Photothermal investigation of transport in semiconductors: Theory and experiment", J. Appl. Phys., No. 59(1), 1986, pp. 787-7951

6 R.A. Soref, and B.R. Bennett, "Electrooptical effects in silicon", IEEE J. Quant. Elec., No. QE-23, 1987, pp. 123-129

7 W. Claeys, S. Dilhaire, L. Sarger and, Ch. Zardini, "Joule displacement and thermoelastic optical test probe for electronic components", Proc. ESREF 91, 1991, pp. 819ff

8 W. Claeys, V. Quintard, S. Dilhaire, D. Lewis, and Y. Danto, "Testing of solder joint quality through the analysis of their thermal behaviour with an interferometric laser probe", QRE Internat., 1994, accepted for publication

9 W. Claeys, S. Dilhaire, and V. Quintard, "Laser probing of thermal behaviour of electronic components and its application in quality and reliability testing", Microelectron. Eng., No. 24, 1994, pp. 411-420

10 J.M. Dorkel, and Ph. Leturcq, "Carrier mobilities in silicon semi-empirically related to temperature, doping and injection level", Sol St Elec., No. 24, 1981, pp. 821-825

CONTRIBUTION TO FAILURE PREDICTION INDUCED BY AGEING DUE TO THERMAL FATIGUE FOR POWER ELECTRONIC DEVICES

M.KHELIF; G.ROJAT; G.CLERC
CEGELY (URA CNRS 829)
Ecole Centrale de Lyon
36, Avenue Guy De Collongue
69131 Ecully cedex, France

Abstract

Ageing due to thermal fatigue is the main problem for long term reliability of power electronic devices. Failure mechanisms under the assumption of repeated thermal stresses are investigated and experimentally validated. Then, an off line related failure prediction procedure is built up.

1 - INTRODUCTION

Repeated temperature cycling due either to fluctuations in power dissipation or to changes in ambient temperature, are at the root of ageing problems for power electronic devices. Starting from the usual multilayered thermal circuit (fig -1), it is shown quantitatively that shear stresses are developed at the joining surfaces of the different layers (Refs.2,4,7). This is essentially due to the differences in the thermal expansion coefficient of the various materials. In the general case where the silicon chip is mounted on the copper base sink by means of a soft solder (Pb-Sn usually), most of the stresses will be concentrated in this joint (shock absorber effect). The magnitude of these stresses generally exceeds the elastic limit of the material. A memory effect takes then place (plastic effect), and the strain energy builds up with the repeated stresses. Hence, after a given number of cycles, the fracture strain limit is reached. This initiates material structure dislocation and induces micro cracks. Holding the stresses on, the voids area will increase and may become a significant proportion of the total chip area. Thermal exchanges with the ambient are then reduced, and the major part of the dissipated power will contribute to the junction temperature elevation. Hence, a thermal breakdown will occur when the maximum allowable temperature is reached.

According to the analysis made above, it is clear that junction to case thermal impedance (Zth) should be correlated with the ageing threshold due to thermal fatigue. An intensive experimental work has been carried out to validate this assumption. At this stage of the work, three devices have been investigated:

- Bipolar transistors
- MOSFET transistors
- IGBT

After a brief survey of the experimental procedure set up to validate the assumptions made, the application of thermal impedance Zth as ageing threshold indicator is discussed. Because of the inherent limitations of applying this method in line and its restricted reliability, a second approach is proposed. The latter is based on a rational identification of the Foster thermal circuit parameters. The off line implementation of this technique and its theoretical basis are presented along with some results.

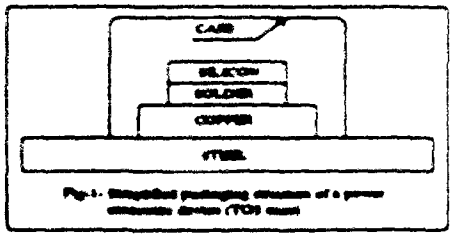


Fig-1- Simplified packaging structure of a power semiconductor device (TO3 case)

2 - EXPERIMENTAL PROCEDURE

In order to make all the expected phenomena observable over a reasonable range of time, several accelerating test benches have been designed. These are aimed to amplify thermal fatigue stress by submitting the device under test to a large temperature excursion. This may be achieved in two different ways:

1- Heating and cooling cycle is fixed by external means. Then, temperature excursion is the same as the ambient one. The latter may be changed by means of regulated temperature enclosures.

2 - The linear characteristics of all the investigated devices may be turned to account. In that case, while cooling is still fixed by external means, heating phase is achieved by a fixed level of power dissipation in the silicon chip. Standard techniques for measuring thermal impedance may then be applied (Ref.1). The latter argument is very attractive because it allows thermal impedance monitoring while ageing test is running. The actual measurement procedure of Zth, is performed during the cooling phase (fig-2). Junction temperature is evaluated through sensitive electrical parameters : Vbe for bipolar transistors, Vd or Vgs for the MOS, and Vge for the IGBT. These characteristics [Vbe(Tj) , Vd(Tj) , Vgs(Tj) , Vge(Tj)] are established experimentally for standard operating conditions (Refs.1,7). Moreover, the thermal fatigue process in that case is nearer to the real one and is simple to implement in practice. For all these reasons, this way of stimulating ageing has been adopted to design the experimental benches.

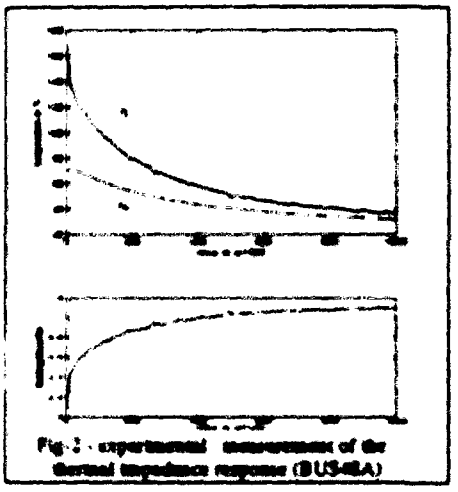


Fig-2 - experimental measurement of the thermal impedance response (BUS48A)

Depending essentially on the control type (current for bipolar, voltage for MOS and IGBT), two specific benches have been designed (fig-3 and fig-4)

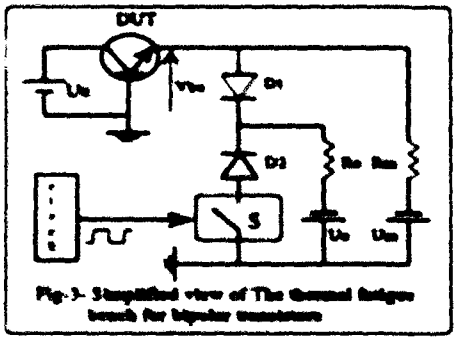


Fig-3- Simplified view of The thermal fatigue bench for bipolar transistors

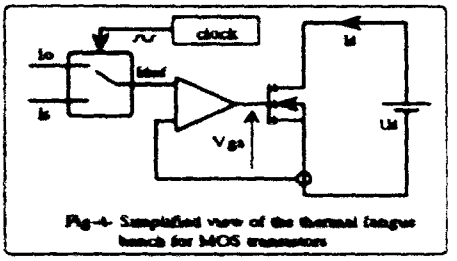


Fig-4- Simplified view of the thermal fatigue bench for MOS transistors

Transistor type	Reference	Packaging	Manufacturer	Samples
Bipolar	BUS 48A	TO3	Motorola	40
MOS	MTP[illegible]	TO220	Motorola	8
	MTP[illegible]	TO220 ISOLATED		8
IGBT	GT25Q101	TO3P7 AC	Toshiba	8

Tab-1- Tested devices

3- ACCELERATED AGEING PROCESS

Several tests were carried out with different series of devices (Tab-1). In order to ensure that the thermal cycling stimulates the only thermomechanical fatigue phenomenon, several precautions have to be taken. These are usually easy to verify with the support of the general technical data given by the manufacturer. Without going into details of the overall experimental results, they can be resumed as follows.

The monitoring of thermal impedance of all the tested devices along with the ageing process has revealed three major steps (fig-5):

1- A long steady phase where the thermal circuit does not show any changes.

2- The stage after is characterised by a small decrease of the thermal impedance steady state value. This may be explained by the activation of the diffusion phenomena under the testing conditions. The result is relatively better thermal contacts between the different layers, hence the observed decrease on the overall thermal resistance.

3- Finally at the beginning of the last stage, an increasing rate is observed on either transient and steady state thermal impedance. The latter speeds up at the end before lethal thermal defects occur. this last step is well explained by the occurrence and the development of micro cracks in the joint. However, the small inherent shifts on the thermal characteristics of the different materials under the cycling stresses seem a better explanation of the slow rate in the beginning.

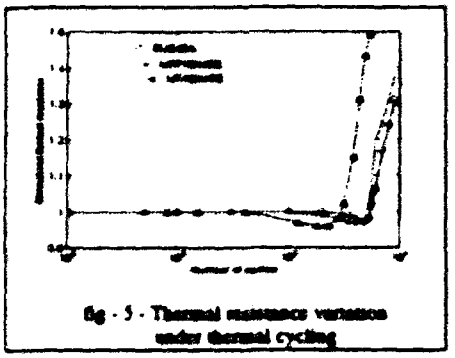


fig - 5 - Thermal resistance variation under thermal cycling

As clearly shown in fig-5, thermal resistance Rth seems to be a good indicator of the onset of ageing failure. Nevertheless, the application of this approach with the real device environment may lead to a high rate of false alarm. In fact, the registered increase of Rth may also originate from the inherent shifts of the external thermal circuit characteristics. Moreover, the junction temperature measurement needs highly standardized conditions. Calibrations are lengthy and very sensitive procedures. Hence, to discriminate the ageing effects, and enhance the robustness of the prediction, a method based upon the thermal circuit model is elaborated. The inner thermal modes (inside the case) being clearly different from those outside, a more reliable discriminating approach can be built up. The identification of the thermal model along with the running ageing process allows detection of any mode shift. A complete localisation of the contributing layer to the increase of thermal impedance is achieved. If the latter corresponds to an inner one, the registered change is related to ageing, else it indicates an external shift. Beside that, no more calibrations are needed. The junction temperature is replaced as input of the system by the dissipated power which is much easier to measure.

4 - THEORETICAL MODEL

Making use of the existing analogy between electrical and heat conduction, the global thermal circuit of a power device associated with its heat sink can be represented by a RC network (fig-6)

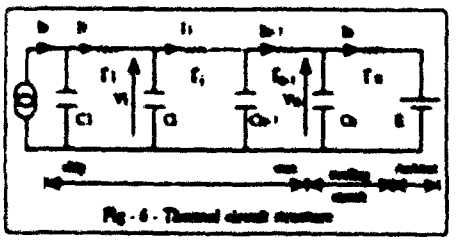


Fig - 6 - Thermal circuit structure

The circuit of fig-6 can be modelled by means of a state space equations system as follows:

$$\dot{X} = AX + BU \qquad (1)$$

Where:

$$A = \begin{bmatrix} -\frac{1}{\lambda_1} & \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ \frac{1}{\tau_1} & -\left(\frac{1}{\lambda_2}+\frac{1}{\gamma_1}\right) & \frac{1}{\lambda_2} & & \\ 0 & \frac{1}{\gamma_1} & -\left(\frac{1}{\lambda_2}+\frac{1}{\gamma_2}\right) & \ddots & \\ & & \ddots & & 0 \\ 0 & \cdots & & \frac{1}{\tau_n} & -\frac{1}{\lambda_n} \end{bmatrix}$$

$$X = \left[I_1 - - - I_i - - - I_n\right]^T$$

$$B = \left[1/\tau_1\, 0 - - - - - - 0\right]$$

$$U = \left[I_g\, 0 - - - - - - 0\right]^T$$

$$\lambda_i = r_i C_i \cdot \lambda_n = r_n (C_n C_{n+1})/(C_n + C_{n+1})$$

$$\tau_1 = r_1 C_1 \cdot \gamma_n = r_n C_{n+1}$$

$$V_i = E + \sum_{j=1}^{n} r_j I_j \qquad (2)$$

5 - SIMULATION

In order to apprehend in details the effects of the observed defects, and the way to monitor them through observable parameters, a flexible tool like simulation is very useful. However, to be effective one has to deal with the real scale of parameters. In this way, and concerning the intrinsic device thermal circuit, particular contribution of the manufacturer is desirable. In default of this, the usual available data, especially thermal response is of great interest (ref.3). The external thermal circuit characteristics may be evaluated with an acceptable precision through some measurements.

The simulation support used in this application is the bipolar transistor BUS48A of MOTOROLA (sub-1). The latter has been the main experimental device, and details about its thermal circuit were supplied by the manufacturer.

To complete the effectiveness of the simulation procedure, it is important to reproduce the real limits on the measurements and their acquisition. In this field, there are two main restrictions :

1- case temperature sensor bandwidth

2- Sampling frequency and samples number

The global structure adopted for simulation purposes, is composed by six RC cells. Four of them are used to represent the circuit inside the package. The discrete form of the corresponding transfer function is given by (3). The latter has the dissipated power Pe as the input, and case temperature Tc as the output.

$$\frac{Tc}{Pe}(z) = \frac{b_5 z^5 + \ldots\ldots + b_1 z + b_0}{z^6 + a_5 z^5 + \ldots\ldots + a_1 z + a_0} \qquad (3)$$

Relying on the real magnitude of the experimentally observed shifts on thermal impedance, these are simulated by incremental changes imposed upon the RC cell corresponding to the solder (fig-6). Furthermore, to assess the ability of discriminating ageing effects, parallel shifts were made upon the case -ambient thermal resistance

Assuming an ideal measurement quality, and a complete absence of unknown perturbations, the sensitivity of the different transfer function parameters to the imposed incremental changes is investigated. Through that hypothetical case, the possibility of detection and localisation of the ageing defects is established (fig-7). However, according to the best registered sensitivity, the task is likely to be difficult in the real world, especially when dealing with the identification problems. A thorough assessment of these effects is experienced in the next step of simulation.

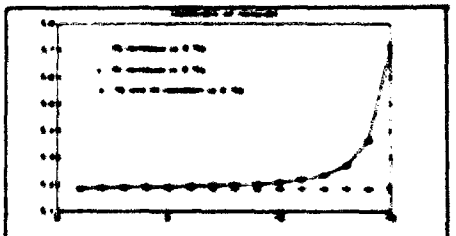


Fig-7- Theoretical variation of a_0 parameter of the transfer function with ageing

Using the black boxes approach, the aim is to build up a sufficiently accurate thermal model, from input-output measurements (fig-8). In the present case, and for simulation purposes, different sequences were generated by imposing a sufficiently large band input to the above known structures. Two sets of simulated measurement sequences were generated. In the first one, corrupting noise and unknown inputs are omitted, while a realistic signal to noise ratio has been introduced in the second set.

The application of the appropriate parameter identification techniques (ref.5) to the first set of simulated measurements, has perfectly reproduced the original structures. However, as soon as the measurements were corrupted by some noise, even with optimistic signal to noise rano, a rapid degradation of detection and discriminating ability is observed. No more clear trends related to the ageing threshold are available through the identified parameters. This explanation of this resides in the fact that the large band information describing the layer of interest, is packed up after a drastic attenuation through much larger layers (inner copper heat sink, and the steel case).

6-EXPERIMENTAL IDENTIFICATION

6-1 Experimental environment

From the previous analysis, it is clear that the widely spread time constants of the system is a major problem. This situation needs special attention because of its prominent influence on the identification effectiveness as a whole. It is not feasible for either practical (optimum sampling rate) or numerical reasons (ill conditioned system..) to handle all these time scales at once and with a single model. As the interest goes to the thermal circuit inside the device package, the identification environment is defined consequently. A model reduction may then be achieved around the frequencies of interest (ref.6). Hence, instead of a single ill conditioned system, several reduced but well conditioned models are constructed. However, the extraction of a reliable reduced model, still need the effective presence in the measurements of a sufficiently related energy in the investigated frequency range. From the practical implementation point of view, and in the context of the tested devices, this is not clearly achieved. Admitting that the most useful measurable output (closer to the investigated region) is the case temperature Tc, it is evident that the joint dynamic information should be drastically filtered out when travelling through heavier layers. A typical joint thermal time constant is of the order of 0.2 ms, while those related to the inner copper heat sink and the steel case (fig- 1) are of the order of 0.2 s. Furthermore, one has to deal with the limited bandwidth of temperature sensors (0.16s for the used one). In such a context, any joint structural changes detection should generate a high rate of false alarm in presence of inherent noise measurement. This issue is well validated through simulation results. Nevertheless, all the quoted limitations are closely related to the physical scale of the considered component. A cost effective implementation of such approach may well be successful for larger devices. To show that, the experimental procedure has been implemented with the same measurement environment, where defects

are artificially amplified to some reasonable levels. These are created by inserting thin layers of different heat conducting materials between the case base and the heat sink. The platinum resistance used as output sensor is located at the interface between the newly inserted layer and the heat sink. By doing so, the identification procedure is intended to detect the presence and evaluate the occurring changes when the added layer is replaced by another one of different thermal characteristics. Besides the case where no added layer is used, two other types of situations were tested with:

1- a thin film of aluminium

2- the classical electrical insulation washer

The input u (t) which corresponds to the dissipated power in the silicon chip, has been chosen as a pseudo-random binary sequence (PRBS). The latter replaces the clock in the designed ageing benches (fig-3 and fig-4) when input-output data is needed for identification purposes (fig-8). To ensure an informative experiment, the PRBS should have a sufficiently large band to excite all the system modes of interest. The designed benches offer a good flexibility in this way. However, to ascertain the statement of feasibility made above, all the measurements to be presented were made in a standard configuration defined as follows:

PRBS order	7
PRBS clock period	200 ms
PRBS levels	0 - 100 w
Sampling period	30 ms
Observation time	30 s
Measurement sequence	1024 points/per variable

6.2 Model structure

An overall ARMAX model structure has been adopted for all the tested cases (4) (ref 5).

$$y(t) = \frac{B(q)}{A(q)} u(t - nk) + \frac{C(q)}{A(q)} e(t) \quad (4)$$

q^{-1} is the delay operator, nk the system delay, u the known input, and e is assumed to be white noise. A(q) B(q), and C(q) are polynomials in the delay operator q^{-1} of respective orders na , nb , and nc.

The choice of such structure is justified by two major arguments:

1- The unknown input dynamics in this case are likely to be correlated with those of the identified system. This of course is a good situation for auto regressive (AR) structures.

2- The mean adjusted (MA) part of the perturbation model is introduced to deal with the large band unknown inputs.

To converge toward the optimum model in the set of candidates defined by (4), a spectral analysis of the input-output data (fig-8) is first made. In the same way, prior information about the identified intrinsic system bandwidth is used. In fact, the latter is a good indicator of how to share the observed dynamics in the step before, between the system model (B(q)/A(q)) and the perturbation one (C(q)/A(q)). Hence, the iterative search, according to a given criterion fit, for the optimum structure (Ref. 5) fixed by the different polynomial orders (and the delay nk), is reduced. In the present case, the optimum is reached for the following combination :

na = 4 ; nb = 4 ; nc = 4 ; nk = 1

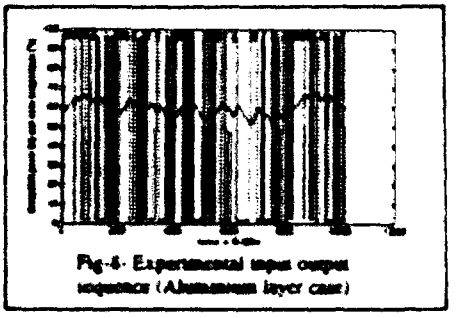


Fig-8- Experimental input output sequence (Aluminium layer case)

6.3 Identification

Table II resumes the different identified system model parameters and the corresponding standard deviations (between brackets). Loss function (Ref.5) is also given in the three experimental situations introduced above.

	case -1- (no added layer)	case-2 (aluminium layer)	case-3 (isolation washer)
a_0			
a_1			
a_2			
a_3			
b_0			
b_1			
b_2			
b_3			
loss function	5.630 e-3	6.126 e-3	5.923 e-3

Table II different model identified parameters

Loss function given in table-II allows a good assessment on the validity of the identified models. However, a more expressive way to do so is to compare the model an the real system responses when they are excited by the same input (fig-9).

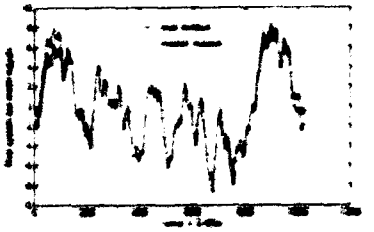


Fig-9- Comparison between the simulated model output and measured output

Fig-9 shows that the model is quite efficient in representing the true system dynamics. The next analysis step has to deal with the quantification of the observed shifts. These have to be linked to the operated changes, in order to achieve better discrimination of the operated changes contribution. To achieve that, a simple idea consists to convert the identified (continuous) transfer function to a state space representation similar to (1). Provided that the identified structure allows that, it is possible to realize this conversion in a direct way by means of a model reduction technique. The latter makes use of Padé approximations when developing the continuous transfer function around s = 0 (named Cauer form) [Ref 6]. A shift from the standard point (s = 0), to the frequency of interest ($f = \omega/2\pi$) is simply achieved by replacing s by (p+ω) in the originally identified transfer function. The standard technique gives then a reduced model around ω , which admits a direct state representation similar to (1). The RC network calculation is then evident. The main arising problem with such approach is the lack of direct physical relation between the reduced model parameters and those of the actual thermal circuit. However, it remains always possible to make reference to the homologous parameters identified in the same way and in a standard situation (when the device is practically new for example). Hence, a physical basis is available for the observed shift characterization in terms of thermal resistances, thermal capacitances or both simultaneously (Ref. 7).

7 - CONCLUSION

An overview of the concept of thermal fatigue applied to the electronic power devices has been presented. The substantial experimental work achieved in this area, has revealed that the main consequence of this phenomena, is an increase of the junction to case thermal impedance. Under sustained stresses, the latter continues to increase until thermal breakdown occurs. Starting from this observation, a prediction procedure of the long term ageing failure is built up. The latter relies on the identification of the thermal circuit characteristics. Investigations about the feasibility of this approach, are first made through simulation work. To make this analysis effective, realistic parameter scales of the tested devices and the "natural" limits on the measurements are introduced. In this situation, the results have shown a small sensitivity of the observable parameters to the dynamics of interest, used for discrimination purposes. On the other hand, this limitation being related to the device scale, the feasibility of the approach has been clearly established for large devices. The latter statements in turn, has been validated through various experimental works.

REFERENCES

1- Frank F O , David L B and Sherwin R "Thermal Characterization of Power Transistors", IEEE Transactions on Electron Devices, vol ed-23, No 8, August 1976, pp 831-838

2- Zommer D , Feucht D L , and Heckel R W "Reliability and Thermal impedance studies on soft soldered power transistors", IEEE Transactions on Electron Devices vol ED-23, No 8 August 1976, pp 843 - 850

3- E. Lu and Bronner J "Computer - Aided Methods of Determining Thyristor Thermal Transients", IEEE Transactions on Energy Conversion, Vol 5,No 4 , December 1990, pp 616-623

4- Alcan P A "Thermomechanical Degradation and Thermal Fatigue", ESREF 93 Bordeaux France October 1993

5-Ljung L "System Identification Theory For the User", Prentice-Hall, 1987

6-Berna J , Dauphin-Tanguy G , Richard J P , Rancilla F , and Zambettakis I "Modelisation et Identification des Processus" Ed Technip 1992

7- Khalil M "Contribution a la prédiction des délais de vieillissement par fatigue thermique des composants électroniques de puissance", Thèse de doctorat , E.C.L. Lyon , France, to be presented

IDENTIFICATION OF PERFORMANCE CRITICAL STEPS IN POWER SEMICONDUCTOR PROCESSING

C Mark Johnson, Department of Electrical and Electronic Engineering, University of Newcastle upon Tyne, Newcastle upon Tyne, NE1 7RU, UK

Patrick R Palmer and David A Hinchley, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, UK

ABSTRACT

Non-invasive infra-red imaging and magnetic field measurement techniques are employed in combination with wafer probing and computer numerical simulation to identify process steps which dictate limiting safe turn-off conditions in large area GTO thyristors.

1. INTRODUCTION

Experience has shown that the maximum controllable current of GTO thyristors does not increase in proportion to their active area [1,2,3]. Redistribution of the anode current during turn-off gives rise to one or more sites of greatly increased current density. Attempting to increase the average turn-off current will eventually cause failure at one of these sites. The device rating must, therefore, be reduced from that expected to ensure that those areas which turn off last do not exceed their maximum controllable current. This non-uniform behaviour may be explained in part by topological constraints of the design such as variations in gate access resistance. Other non-uniformity can only be explained by inhomogeneity in the starting silicon and the technological limitations of the manufacturing process when applied over large areas.

The phenomenon of current redistribution was first noted by New *et al.* [4] and has subsequently been investigated by various techniques including infra-red imaging [5,6] and magnetic field measurement [2,8]. Further experiments have shown the gate and anode circuit conditions to have considerable influence over the degree of current redistribution occurring during turn-off [5,7,8]. In particular, a high peak anode voltage attained before the current fall has been shown to increase the degree of redistribution. When combined with the high level of local power dissipation this can initiate a thermally driven positive feedback process and cause failure of the unit. It is for this reason that capacitive snubbers must be employed in all practical GTO applications requiring hard turn-off operation. Inductive strays associated with such snubber circuits must be kept as low as possible to ensure rapid transfer of current from the GTO and maintain a low anode spike voltage.

From the foregoing it is clear that the non-uniform turn-off behaviour of the GTO has profound implications for both its manufacture and application. To reduce the total losses associated with a particular GTO application users would naturally prefer to use small values of snubber capacitance and high turn-off gains. Such devices can only be satisfied if the uniformity of processing and topology are high enough to reduce the effects of current redistribution to acceptable levels. Manufacturers typically resolve such issues by means of considerable experimentation with production processing apparatus and sample testing of the devices so produced. This is a time consuming and costly business (a single batch of GTOs may cost in excess of £20,000 to process). In contrast, the techniques presented in this paper allow evaluation of features of the manufacturing process by permitting direct, non-invasive and non-destructive measurement of GTO turn-off performance. The discussion will concentrate on the application of these techniques to the identification of a critical process step in the manufacture of practical GTO thyristors.

2. CHARACTERISATION TECHNIQUES

The GTOs studied in the following were rated at 4.5kV 3000A with a total silicon area of roughly 36cm^2. The cathode was divided into more than 1200 segments arranged in 7 concentric rings whilst the gate contact was made available at the centre of the device and between cathode rings 4 and 5.

2.1 Thermal Imaging

Redistribution of current in the device at turn-off will be accompanied by an increase in losses in the region to turn off last and a reduction in the losses elsewhere. If these turn-off losses can be made to dominate turn-on and on-state losses, the temperature variations across the device may be interpreted as representing the degree of current redistribution. Where there is a means for viewing the device, variations in temperature can be measured using an infra-red thermal imaging system.

Early work on the thermal imaging of complete GTO thyristors encountered wire bonding and hence disruption of the device cathode metallisation [5]. An improved technique was developed allowing the

imaging of pressure contacted GTOs from the cathode side through a specially designed silicon window [6]. Figure 1 shows a thermal image of a 4.5kV, 3000A GTO (device CUED1), 75mm. in diameter, switching 1250A into a voltage clamped, inductive load at a clamp voltage of 400V. The device was processed using a standard 75mm device technology. A clear hot-spot is visible at the top of the image, indicating a site of current concentration in the outermost rings of segments at an azimuth of 0° relative to the mask alignment marks (azimuth is measured anti-clockwise from a datum at the top of the image).

2.2 Magnetic field measurement

Changes in the current distribution within the bulk silicon of a GTO thyristor during turn-off will alter the electromagnetic field outside the device. Small search coils and an integrating amplifier may be employed to determine the radial and tangential components of magnetic field as a function of time and azimuthal position around the device [7]. The tangential component of field is strongly related to the device anode current whilst the radial component of field will be relatively small provided the current distribution is nearly uniform. Any changes in either component of magnetic field occurring before the onset of the current fall are indicative of current redistribution [8].

Figure 2 shows a plot of the tangential magnetic field as a function of time and angular position for device CUED1. A clear peak, indicating a site of strong current crowding, is seen to develop at an azimuth of 0° during the latter stages of the turn-off storage period.

2.3 Segment probing

The device parameters for each cathode segment may be conveniently found by using a computer controlled probing system. Probed parameters typically include the gate cathode reverse breakdown voltage, V_{gk}, on-state voltage, V_t, gate triggering current, I_{gt}, anode current at latching, I_L, and the minimum quasi-static, turn-off gate current, I_{gm}. Plots of each of these parameters against angle and segment ring number can be used to make inferences about the uniformity of processing.

Figures 3 and 4 show, respectively, plots of V_{gk} and I_{gm} for device CUED1. Note that the 7 rings of segments contribute a total angle of 2520°. Both plots show a periodic pattern with clearly identifiable peaks at an azimuth, measured separately for each ring, of around 0°. Note also that the non-uniformity of the probed parameters is greatest in the outermost rings of the device.

2.4 Numerical simulation

Although pen and paper techniques can yield rapid solutions to many problems in device design, they are usually limited in accuracy and typically cannot predict the effect on all aspects of device operation simultaneously. Careful application of physical, numerical models can, however, yield such information, particularly when applied with realistic, circuit based boundary conditions.

Within the context of this paper, numerical modelling was used to study the effect of non-uniformity in selected process variables on the static and dynamic characteristics of GTO thyristors. By modelling a series of individual GTO segments, each with slightly different physical characteristics, deductions can be made about the influence of various inhomogeneities on, say, probed parameters and the degree of turn-off current redistribution.

Figure 5 shows the results of a MEDICI transient simulation involving 4 GTO segments connected in parallel with common anode and gate circuits. Each segment has a different gate mesa-etch depth and is represented in the simulation by a 2-D device physics model. The presented simulation covers a full turn-on and turn-off transient in which the total anode current is ramped up on a voltage clamped inductive load prior to turn-off. Clear differences are visible in both the on-state and turn-off behaviour. In particular, the segment with the deepest mesa-etch depth shows a marked increase in its share of the total anode current during the turn-off process. It should, therefore, be expected that the current in a practical GTO will redistribute to sites with a deep mesa-etch provided that no other process variation dominates.

3. DISCUSSION

The results shown in Figs. 1, 2, 3 and 4 refer to a single device, taken from a batch processed using a standard reference technology. Results obtained from other devices in the same batch were substantially the same.

Comparisons of Figs. 1 and 2 with Figs. 3 and 4 indicate a strong correlation between the current crowding site and the sites of peak V_{gk} and peak I_{gm}. No correlation was observed between the site of current crowding and any other probed parameter. It is, thus, reasonable to conclude that the non-uniformity giving rise to the observed current redistribution is also the source of non-uniformity in the probed parameters, V_{gk} and I_{gm}. Both parameters depend, to a large extent, on the physical characteristics of the gate-cathode junction of the GTO.

A definite correlation between the distributions of these parameters supports the hypothesis of non-uniformity in the vicinity of the gate-cathode junction.

In an attempt to isolate the performance critical step of the manufacturing process, results from the numerical modelling experiments were examined. From Fig. 5, it is clear that areas with a deep mesa-etch should be expected to display the greatest degree of turn-off current redistribution. On its own this result is of little consequence since variations in almost any process variable will lead to some degree of redistribution. To resolve the problem, further simulations were performed to determine the influence of the mesa-etch depth on the range of probed parameters. The results showed that increases in the mesa depth led to increases in both V_{gk} and I_{gm}. Furthermore, modelled variations in other process parameters, such as bulk lifetime, anode shorting density and doping level, did not yield simultaneous increases in both V_{gk} and I_{gm}.

Collectively, these studies indicate that areas with a deep mesa-etch will display high values of V_{gk} and I_{gm} and that they are also liable to become sites of turn-off current crowding. Since the experimental results presented in Figs. 1, 2, 3 and 4 display the same trends and correlations across the set of measured quantities, it is reasonable to conclude that the mesa-etching process is limiting the turn-off performance of the GTOs.

On the strength of the above evidence, an improved etching technique was devised and a single test-batch of devices produced. Probing results taken from devices in this batch show greater uniformity of both V_{gk} and I_{gm} when compared to devices processed using the standard technology. Fig. 6 shows the distribution of I_{gm} for a typical device (CUED2). When compared with the distribution for standard device CUED1 (Fig. 4), it is clear that the peaks in the distribution are less sharp and of smaller amplitude, indicating a more uniform mesa-etch.

The devices also display reduced levels of turn-off current redistribution. Figs. 7 and 8 show, respectively, the turn-off magnetic field profile and thermal image for device CUED2. The field profile indicates peaks in the azimuthal current distribution at angles of 140° and 320° whilst the thermal image shows a broad area of increased temperature extending over the range 90° to 320°. Differences in the two distributions may be attributable to the gate contact arrangements which necessitate the use of the centre gate contact, instead of the more usual ring gate, for the thermal imaging experiments.

It is of interest to note that neither the peaks of the magnetic field profile (Fig. 7) nor the hot spots of the thermal image (Fig. 8) correlate strongly with the probed parameters, V_{gk} and I_{gm} (Fig. 6). This indicates that the mesa-etching process is no longer the only process step limiting device performance. Also worthy of note is a change in the nature of the thermal and magnetic field distributions. The single, large peak, typical of the standard technology devices, has been replaced by several peaks of reduced amplitude. Such behaviour has previously been identified with devices displaying a particularly robust turn-off safe operating area [10]. In practical applications, this equates to reduced turn-off snubber requirements yielding significant gains in overall system efficiency.

4. CONCLUSIONS

- Non-invasive techniques based on infra-red thermal imaging and magnetic field measurement have been successfully applied to the location of turn-off current crowding sites in 4.5kV, 3000A GTOs.
- Segment probing and numerical modelling have been employed to determine the extent and effect of non-uniformity in selected process steps. Correlation between the results and the location of current crowding sites has enabled the identification of a critical process step whilst requiring the production of just one test batch.
- Results obtained from devices manufactured using an improved process exhibit an enhanced turn-off safe operating area and reduced turn-off snubber requirements.

5. ACKNOWLEDGEMENT

The authors wish to thank Mr D. Crees and Mr L. Coulbeck of GEC Plessey Semiconductors for their assistance with the segment probing. This work was funded by the EPSRC and DTI under the LINK PEDDS scheme.

6. REFERENCES

1) A. A. Jaecklin and B. Adam, "Scaling laws and limitations of power turn-off devices," in *Power Electronics Specialists' Conference*, (Milwaukee), pp. 337-342, 1989.

2) T. Yatsuo, S. Kimura and Y. Satou, "Design considerations for large current GTO," in *Power Electronics Specialists' Conference*, (Kyoto), pp. 895-902, 1988.

3) M. Ishiodoh, A. Tada, F. Tokunoh, M. Yamamoto and T. Nakagawa, "A new high power Asymmetrical GTO," *IEE PEVSD*, (London), pp. 424-428, July 1990.

4) T. C. New *et al.*, "High power gate controlled switch," *IEEE Trans. Electron Devices*, vol. ED-17, pp. 706-710, Sept. 1970.

5) P. R. Palmer and C. M. Johnson, "Characterising the turn-off performance of multi-cathode GTO thyristors using thermal imaging," *IEEE Trans. on Power Electronics*, vol. 5, pp. 357-362, July 1990.

6) P. R. Palmer and D. A. Hinchley, "Thermal imaging of large area pressure mounted power devices," in *International Symposium on Power Semiconductor Devices*, (Monterey), May 1993.

7) P. R. Palmer and C. M. Johnson, "Measurement of the redistribution of current in GTO thyristors during turn-off," in *European Power Electronics Conference*, (Aachen, Germany) pp. 1621-1625, October 1989.

8) C. M. Johnson and P. R. Palmer, "The influence of gate drive and anode circuit conditions on the turn-off performance of GTO thyristors," *IEE proceedings, Part B*, Vol. 139, No. 2., pp. 62-70, March 1992.

9) C. M. Johnson and P. R. Palmer, "Simulation of wafer-scale GTO thyristors in circuits," *IEEE Trans. on Power Electronics*, pp. 308-313, April 1991.

10) C. M. Johnson, A. A. Jaecklin, P. R. Palmer, P. Streit, "Correlation between local segment characteristics and dynamic current redistribution in GTO power thyristors," in *International Symposium on Power Semiconductor Devices*, (Baltimore), pp. 121-127, April 1991.

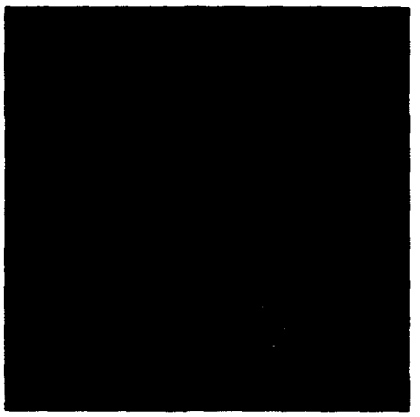


Figure 1: Thermal image of 75mm thyristor CUED1

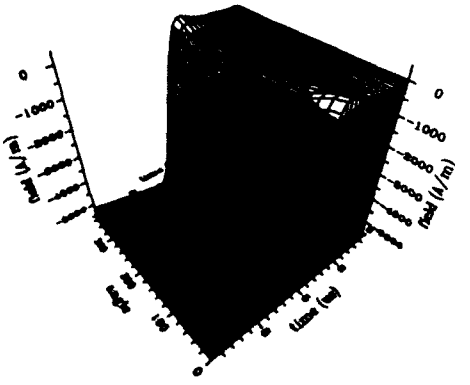


Figure 2: Tangential field as a function of time and azimuth for device CUED1.

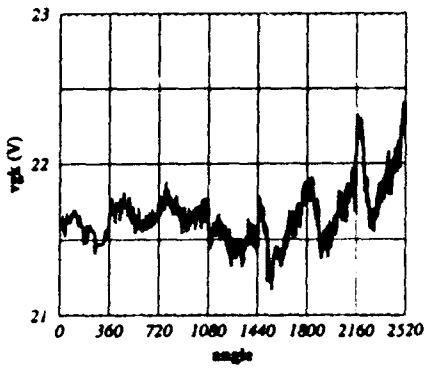


Figure 3: Gate-cathode reverse breakdown voltage for device CUED1.

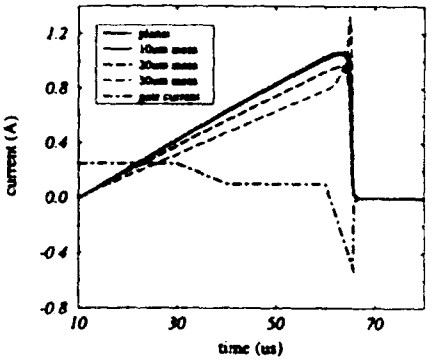


Figure 5 Simulation results showing individual segment currents for a device with non-uniform mesa-etch depth.

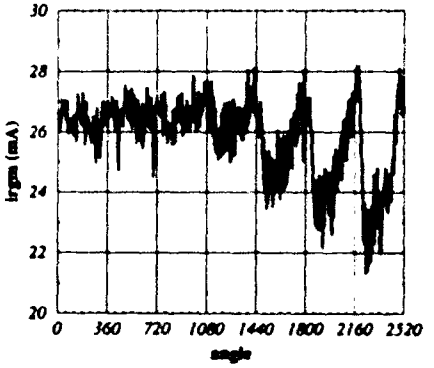


Figure 4: Quasi-static turn-off gate current for device 589-20.

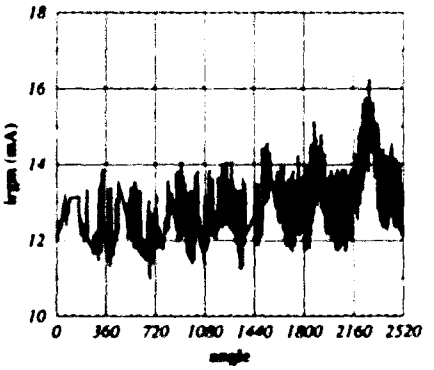


Figure 6: Quasi-static turn-off gate current for device CUED 2.

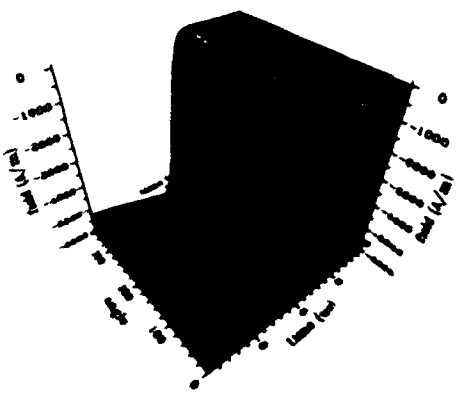


Figure 7 Tangential field as a function of time and azimuth for device CUED2.

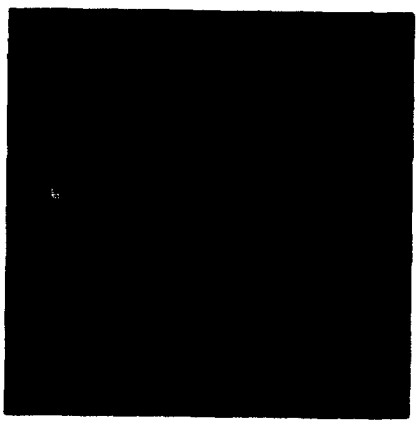


Figure 8 Thermal image for device CUED2.

FAILURE MECHANISMS IN SMART POWER IC's

L. Lonzi, M. Piccoli, F. Speroni

SGS THOMSON Microelectronics, via Tolomeo 1 , 20010 Cornaredo (MI), Italy

ABSTRACT

Field application in SMART POWER IC's strongly characterizes and distinguishes the reliability requirements for these devices. In this context a deep investigation on specific failure mechanisms is mandatory. This paper reviews the main failures mechanisms typical of SMART POWER IC's on the ground of reliability qualification and production monitoring results collected in our lab. during several years of investigation.

INTRODUCTION

The range of application of SMART POWER IC's is becoming broader and broader every day. This is made possible by new advanced mixed technologies (Ref.1) with high integration level which allow to realize power, small signal analogic and high speed structures at the same time, and so make possible to integrate function more and more complex on a single chip. The severe environment and the high performances requested by the application make necessary a deep analysis of the intrinsic failure mechanisms specific of these devices.

The typical failure mechanisms put in evidence during the extensive experimentation simulating high stress field conditions are:

- Metal degradation phenomena
- Bond wire degradation phenomena
- Thermomechanical stress effects
- Surface effects

All these phenomena were investigated in high stress conditions and long term tests to verify the reliability margin in the field (Ref.2).

METAL DEGRADATION PHENOMENA

Degradation phenomena affecting aluminum alloy interconnections are very important for the reliability of SMART POWER IC's due to the high power level present in this type of devices.

These phenomena are investigated through dedicated dynamic or static life test carried under high current/temperature conditions. These tests are generally carried-out during qualification exercise to validate new products and are able to simulate the field application, covering at least the guaranteed useful lie. Sometimes the tests are exasperated to investigate the reliability margin versus these wearout mechanisms.

Pure electromigration phenomena can be avoided by using adequate design rules and generally are not observed. On the other hand complex combined stresses can arise by internal power dissipation in the SMART POWER IC's under dynamic operating conditions, which can cause:

- high temperature gradient present on die surface due to high localized power dissipation (Ref.3);
- remarkable temperature excursion during on/off cycling.

These local stresses conditions can activate thermomigration and stressmigration phenomena (Ref.4) which combined with electromigration can give rise to well evident aluminum damages also in situation that, considering only the mean temperature and the current density, are under the electromigration threshold.

In particular conditions as above described are locally encountered in or near the power components (BJT or DMOS) present in the IC.

Fig. 1 shows an example of this type of degradation on an output bipolar transistor of an IC's submitted to high accelerated dynamic life test. The current density ($2 \div 3 \times 10^5 A/cm^2$) alone is not sufficient to explain the observed degradation which can be justified only taking into account the high temperature gradient present in the region.

FIG. 1: *Degradation of the emitter metal in a power BJT after OLT.*

On the basis of these consideration current density not always is the most meaningful stress parameter for metal degradation in the power structures. By the analysis extensively done on parts submitted to high accelerated life test, a stress parameter more correlated to the observed Al degradation was singled out to be the power density internally dissipated by the power structures. For power density over 10 W/mm² remarkable metal degradations appear (fig. 2).

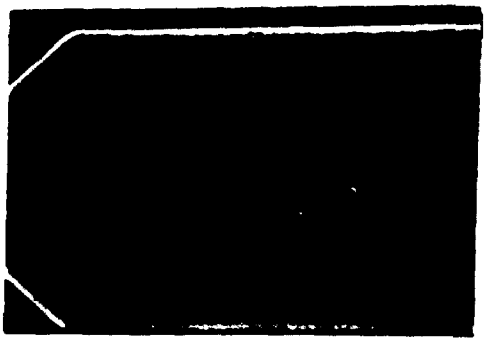


FIG. 2 *Metal degradation after OLT (Pd ≈ 12.5W/mm²)*

BONDING WIRE DEGRADATION

One of the most important failure mode in POWER IC's is wire bonding degradation due to the following three main causes:

- thermomechanical stress
- high temperature
- high current density

Thermomechanical stress is strictly related to the IC's application and to the environmental variation. To simulate and to evaluate the effects due to these kinds of stress, normally temperature cycling test is performed. Lastly some new tests, called power and temperature cycling, are implemented. Figs. 3,4 shown typical degradations of electrical resistance and mechanical strength of gold wire bonding due to thermal cycling.

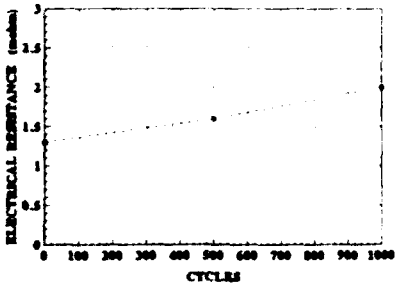


FIG. 3: *Ball bonding electrical resistance increase in thermal cycling test.*

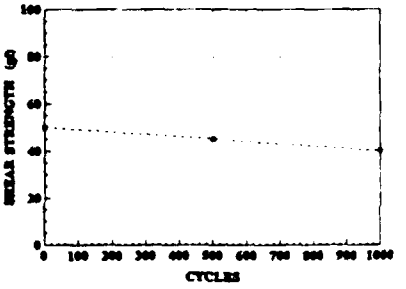


FIG. 4: *Ball bonding shear strength degradation in thermal cycling test.*

High temperature effect is evaluated through a storage test.

In figs. 5,6 typical ball bonding degradations of gold wire on aluminum due to the temperature is reported.

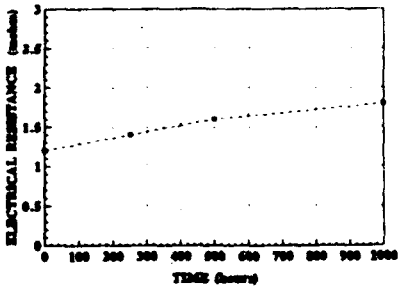


FIG. 5: *Ball bonding electrical resistance increase in storage test.*

Intermetallic compound (fig. 7) evolve during storage test giving rise to Kirkendall voids (fig. 8).

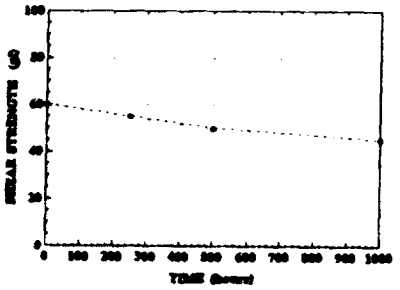


FIG. 6: *Ball bonding shear strength degradation in storage test.*

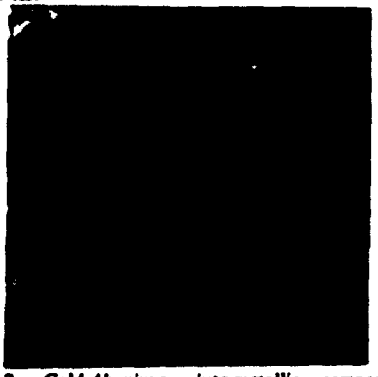


FIG. 7: *Gold-Aluminum intermetallic compound (Au_4Al_2).*

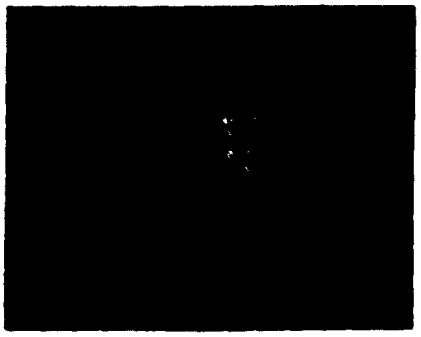


FIG. 8: *Intermetallic evolution after storage test (Kirkendal voids).*

High current density and thermomechanical interaction between resin and wire is able to modify the crystalline structure of the wire itself (fig. 9) causing an increase of electrical resistance as shown in fig. 10.

In high current density the critical structure is the wire itself and not the bonding regions. This is due to the high current density localized in the wire section, while in the two bonding regions, if the wire bonding is correctly done, the current density is lower.

FIG. 9: *Wire crystalline structure modification after conduction test.*

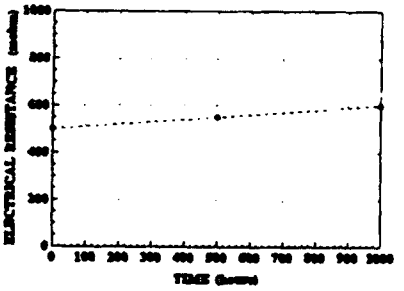


FIG. 10: *Ball bonding electrical resistance increase in conduction test at I=1A, T=150°C (wire length 1cm, wire diameter 25.4um).*

THERMOMECHANICAL EFFECTS

Field environment for SMART POWER IC's can be very severe especially in automotive application where remarkable temperature variation take place. This circumstance and the consideration that the main package involved for the SMART POWER IC's are asymmetric make the investigation of thermomechanical effects very important for the reliability of these devices.

The ruggedness of the IC's as concerns this type of stresses is assessed through temperature cycling and thermal shocks tests.

Typical degradation induced by these stresses on the

die are passivation cracks and metal displacement which can be very heavy (see fig. 11) for dice of large size with uneven surface and thick metal. Suitable package structures, resin compounds, passivation layer are mandatory to contain these effects.

FIG 11: *Metal displacement after 2000 T.S. for a devices in multiwatt package*

FIG 12: *Same device of fig. 11 after passivation thickness increase and package frame improvement.*

SURFACE EFFECTS

We are speaking of those "reversible" degradation mainly induced by contamination: ionic at wafer level or in terms of resin compounds.

The usual way to recover the degradation induced by surface effects is to bake at high temperature failed parts; typically after 2, 3 hours at Ta=150°C, the "virgin" status can be recovered.

The failure mechanisms inducing surface effect are activated by temperature that increases the mobility and by electric field on the direction of which the motion is forced.

These two factors, high temperature and high electric field, are both critical for SMART POWER IC's applications. High temperature condition can be reached by power dissipation and/or by the ambient temperature itself such as in the automotive field where Ta can be in the range of 80°C. High voltage, and hence high electric field, can be up to hundreds of Volt such as in the case of off-line environment.

These characteristics of SMART POWER IC's lead to the necessity of an in depth investigation on surface effects during the reliability qualification of process and products.

The reliability stress test to guarantee that SMART POWER IC's are free of surface effects are in fact performed in two steps: at process level on structures (DMOS, BJT and so on) kit parts, at product level directly on IC's.

In both cases H.T.R.B. (High Temperature Reverse Biasing) stress test are performed at Tj= 150°C and, as regard voltage stress, at the absolute maximum rating allowed by the structure or by the product respectively. These stress conditions are usually applied for 1000 hours or more for particular customer request. Degradation, in term of electrical parameter drift, due to surface effects can anyway be detected in the first 500 hours of HTRB stress tests.

On kit parts structures performing HTRB their design mainly in term of process architecture is checked. The degradation is detected by monitoring electrical parameters; the more sensitive are for sure junction leakage and junction breakdown anyway as function of the structure other parameter can be considered such as Ron for DMOS or more generally the threshold voltage for MOS structure (Refs.5,6).

These degradations can saturate or not as function of the amount of contamination and also as function of the structure sensitivity. Because the extent of degradation can not be a priori guaranteed, when problem due to surface effects are detected they must be removed modifying the structure layout and/or adopting filed plate protection.

These kind of improvements, thanks to the experience gained in twenty years of reliability activities, are formalized in consolidated Design Rules that allow to design robust process. Anyway the challenge of higher and higher performances in term of voltage, such in the case of our BCD process, requires a constant level of attention to these aspects in a continuous improvements approach.

During products qualification the HTRB stress test allow to verify the surface effects as regard: die package interaction, and also for the interaction among different structures as function of layout. In particular this second aspect is becoming more and more important for off line SMART POWER IC's application where on the same die are integrated structures able to sustain hundreds of

Volt and standard CMOS cells. The HTRB on products is mandatory to check these kind of phenomena because in no other way they could be simulated working for example at structure kit part level.

Such as for single structure also during products qualification no failure due to surface effects can be accepted. On products is not a priori evident which could be the more sensitive electrical parameter but for this reason a lot of attention is paid in the analysis of parameters drift collected during HTRB stress test. To this aim dedicated data base tools have been developed in our lab (Ref.7).

When surface effects is detected during products qualification, to allow corrective actions a failure analysis activity became absolutely fundamental, it plays the difficult job to localize the parasitic structure activated by contamination. In fig. 13 the first step of this activity, a pin to pin comparison between good and failed parts (parts number 62, 23), is reported. In fig. 14 the recover after one hour of baking at Ta = 150C is shown.

FIG. 13: *Pin to pin comparison of failed (parts n. 23, 62) after HTRB and good (unstressed) one.*

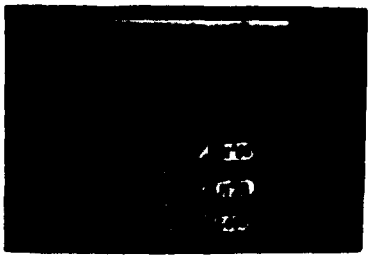


FIG. 14: *Recovery phenomena after one hour of baking at Ta = 150C.*

All the above consideration are related to the methodology adopted during qualification where the subject is to check the immunity of process/product to the "intrinsic" level of contamination.

A complete different approach has to be adopted when the aim is to guarantee/control the contamination level of production process both as concern wafer and assembly processes. In this case is not matter of qualification but more properly of monitoring, this is performed by means of an RTC (Real Time Control) activities on packaged parts, by means of SPC (Statistical Process Control) data analysis and also using WLR (Wafer Level Reliability) techniques recently developed.

CONCLUSION

At present, on the basis of all the experimental reliability results coming from: qualification, manufacturing area and field test data, the market requirements in terms of reliability performances for today SMART POWER IC's are completely guaranteed.

However, the fast rate of the market evolution, characterized by the demand of more and more power density and circuit complexity, leads to investigate structures, materials and new technological solutions at the extreme limits of their reliability performances, discovering in this way, sometimes, new failure mechanisms.

The constraint to anticipate as much as possible the market scenario needs a continuous improvement process and a concurrent engineering approach.

ACKNOWLEDGEMENT

The authors thank D. Galletti, responsible of the failure analysis lab, for the fundamental support given to this work.

REFERENCES

1. A. Andreini, C. Contiero and P. Galbiati "A New Integrated Gate Technology Combining Bipolar Linear, CMOS logic and DMOS Power Parts", IEEE Trans. Electron Devices, Vol. ED-33, December 1986.

2. F.R. Borri and L. Lonzi "Reliability Approach for SMART POWER IC's", ESREF 93.

3. F.R. Borri and G. D'Espinosa "Power Reliability", in "Power Integrated Circuit Physics Design and Application" - Chap. 8, PAOLO ANTOGNETTI Editor,

Mc Graw-Hill Book Company, 1986.

4. A. Scorzoni, B. Neri, F. Fantini "Electromigration in Thin-Film Interconnection Lines: Models, Methods and Results", in Material Science Reports - Vol. 7, N. 4,5, December 1991.

5. F.R. Borri, C. Cognetti, G. D'Espinosa, G. Fioravanti, E. Ghio and L. Lonzi " Reliability of Power Devices", in "The Reliability of Semiconductor Electronic Component", E. Pollino Editor, STET school pubblication, July 1987.

6. G. Barbottin and A. Vapaille Editors "Instabilities in Silicon Devices", North Holland.

7. P. Mauri "Integrated Tools for the Analysis of Reliability Data", Proceeding of ESREF, Bordeaux 1991.

HIGH POWER SEMICONDUCTOR RELIABILITY EVALUATION BASED ON TRACTION VEHICLE FAILURE ANALYSES

G. Bifolco (*), F. Fasce (**), M. Pasqualetti (**), M. Portesine (**), P. Zani (**)

(*) Ansaldo Trasporti, Quality Department, via Nuova delle Brecce 260, Napoli, Italy

(**) Ansaldo Trasporti, Semiconductor Department, Via N. Lorenzi 8, Genova, Italy

ABSTRACT

In the development of high power semiconductor devices it assumes an important role the determination of reliability data.

A system based on a collection of failure analysis data from traction vehicle applications (Italian railways, subways, etc) has been developed.

It allows a field feedback to the device design and prototype laboratory reliability tests.

1. INTRODUCTION

In the development, manufacturing and application of high power semiconductor devices it assumes an important role the determination of reliability data.

These data may be used as feedback for the manufacturer to improve device characteristics or to review critical design and fabrication steps (FIG. 1). Moreover they may be used by the customer to predict equipment reliability and to evaluate spare parts necessity.

To assure a high level predictive evaluation of reliability data different inputs are needed: market requirements, laboratory reliability tests, Failure Mode Effect Analysis (FMEA), failure data from field applications.

Reliability data for high power semiconductor devices are not as well established as for the signal and discrete devices whose reference data are widely reported in MIL-HDBK (1).

Purpose of this work is to present a method to assure the determination of "reliable" reliability data of power diodes, thyristors, transistors ,GTOs starting from field data.

2. LABORATORY RELIABILITY TESTS

Laboratory reliability tests for high power semiconductor devices (FIG. 2) can involve only a small number of devices owing to the high level of power required and the cost of either the equipment or the devices under tests . Usual data involve Total Time of Test (TTT) of the order of ten thousands hours (2).

Using a method like "time compressed test cycle" (3) probably will allow a reduction of the test time necessary to obtain valuable reliability data. At the moment however the application of this technique to press-pack semiconductor is under development within an EC BRITE project.

3. RELIABILITY DATA COLLECTION FROM VEHICLE OPERATION

The field data reported in this work are referred to traction vehicles where the severity of environmental conditions (range of temperature, vibration, humidity, high magnetic field) is emphasized, and where we have the availability of TTT of hundreds millions hours. The complete flow of information feeding the devices reliability data base (4) is shown in FIG 3 .

For all the failure occurring during equipment tests, vehicle tests and vehicle field operations, the faulted assemblies are sent to one repair centre where the failures are diagnosed and the faulted components are replaced; therein the devices failure data are collected. To assure the correct determination of reliability data starting from rough data, a treatment must be performed in order to screen from the total amount of failure the device dependent ones. This treatment requires two strictly connected levels of failure analysis: an equipment level and a device level.

Typical problems encountered in high power module failure analysis are: multiple devices failure; failure induced by control equipment due to random malfunctions, disturbance, abnormal operational conditions; failure causes that can not be determined a posteriori.

Descriptive methods and graphical tools are used for the statistical data analysis as failure rates and down time. When operational data of the single device are concerned both classic and bayesian (5) approaches are employed for statistical inference.

The estimations of failure rate λ (mean value and 90% confidence interval, evaluated in FIT = 10^{-9} h^{-1}) for different high power semiconductors, obtained from traction application, are summarized in FIG.6.

4. HIGH POWER SEMICONDUCTOR FAILURE ANALYSIS

In order to investigate the failure mechanism tools as FMEA are usually applied. FIG.4 and FIG.5 report a manufacturing and a field FMEA for press-pack thyristors.

Methods as FMEA are mainly based on designer and manufacturer experience and should be continuously updated.

Standard physical techniques are used to examine semiconductor failures (6) in order to obtain a correct interpretation of the failure cause.

Post mortems failure analysis equipment laboratory includes:

- Equipment for static and dynamics electrical tests (up to 10000 V and 10000 A);
- SAM (scanning acoustic microscope) for braze voids analysis;
- 1000 x optical microscope;
- Spreading resistance for diffusion profile evaluation;
- Metallization thickness test equipment;
- Helium mass spectrometer for hermeticity test ;
- Flatness, roughness, parallelism test equipment for surface evaluation.

5.CONCLUSION

The availability of hundreds millions hours of total test on high power semiconductors employed on traction vehicles elaborated with statistical method joined with the physical expertise of the devices manufacturer allows the evaluation of reliability data with a good level of confidence.

A constant application of these methods on larger scale should be the right way for reaching a reliability model of high power semiconductors as accurate as the MIL-HDBK-217 for electronic devices.

6. REFERENCES

1. "Military Handbook Prediction of Electronic Equipment (MIL-HDBK 217 E)" Washington, 1986
2. G. Coquery "Thermal cycling: first comparative tests on presspacked and direct bonding technology for GTOs and IGBTs", ESREF 1993
3. V. Loll, L. Rimestad "Time compressed reliability testing" ECR-246, June 1991
4. A. M. Amendola, G. Bifulco, " Reliability data collection system for electric traction vehicles", 7th International conference on reliability and maintainability, 1990 Brest
5. H.F. Marz, R.A. Waller, "Bayesian reliability analysis", Wiley, 1982
6. W.R. Comstock, et. al. " Press - pack post mortems", Solid State State Power Conversion.

FIG.1 : RELIABILITY FEEDBACK LOOP

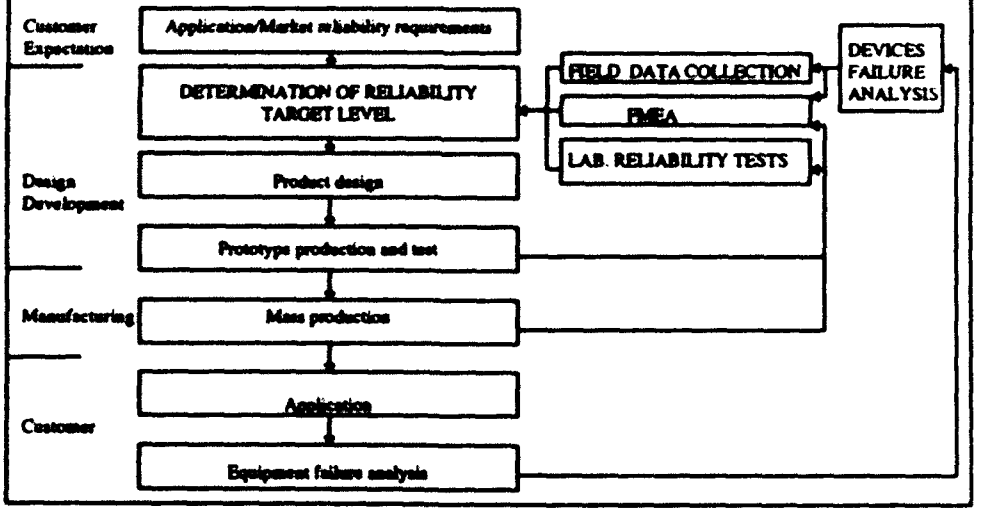


FIG. 2 : TYPICAL SET OF RELIABILITY TESTS FOR HIGH POWER SEMICONDUCTOR DEVICES

TEST	REFERENCE	TEST CONDITION
High temperature blocking voltage test	IEC 147-4 (1976)	Sine wave, 50Hz, Tj=Tj max; t= 168÷1000 h; V=Vmax
Thermal cycling load test	IEC 147-4 (1976)	I (half sine wave, 50 Hz) high enough to heat the devices to Tj max; cooling to not below 40 °C; number of cycles= 1000÷25000
High temperature storage test	IEC 68-2-2 (1974)	T= Tj max; t= 1000÷10000 h
Low temperature storage test	IEC 68-2-2 (1974)	T= Tj min; t=168 h
Sealing test	CEI 50-7 (1985)	Mass spectrometer; Leak rate ≤ 0.5 Pa cm^3 / sec

FIG 3: STRUCTURE OF RELIABILITY DATA BASE

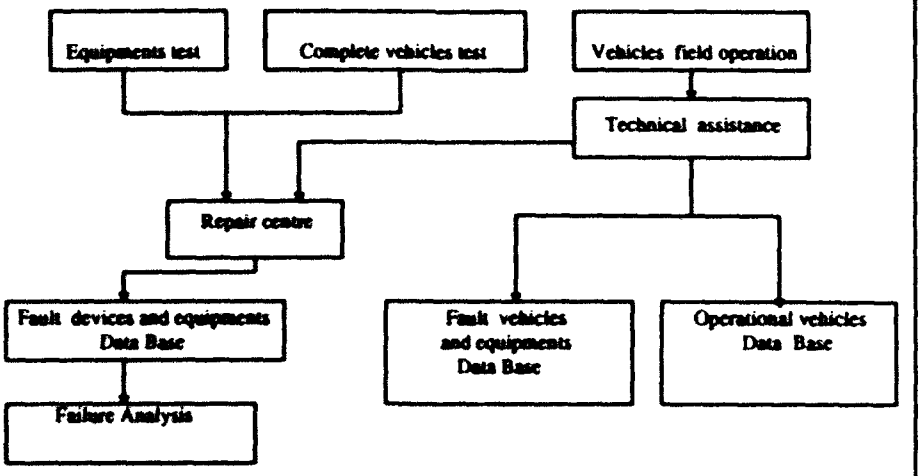


FIG 4 : MANUFACTURING PROCESS FAILURE MODE EFFECT ANALYSIS

PROCESS	FAILURE MODE	FAILURE EFFECT	CORRECTIVE ACTION
Alloy of silicon wafer on molybdenum	Braze voids	Ionic contamination of junction passivation	Monitoring with Scanning Acoustic Microscope of braze layer
Metallization	Thickness out of specification limits; low adherence	Increased on-state dissipation, reduced surge capability	Improvement of engineering instructions; better process inspection
Lifetime control by electron irradiation	Thermal instability of introduced defects (lifetime traps)	Increasing of dynamic losses	Thermal stabilization of irradiated devices
Junction passivation	Contamination; bubbles; low passivation thickness	Blocking voltage instability; increasing of leakage current	Improvement of engineering instructions; better process inspection
Gate contact	Degradation of gate cathode insulation; reduction of spring gate contact	Instability of contact; uncertain firing	Improvement of engineering instructions; better process inspection
Press-pack assembly	Improper positioning of assembly parts; dust contamination	Silicon crack; short circuit	Improvement of engineering instructions; operator training
Welding of ceramic housing	Low housing hermeticity	Degradation of on-state characteristics; higher dissipation	Improvement of engineering instructions; better process inspection
Surface quality of press pack housing	Scratches; low plating thickness; roughness and parallelism higher than specification limits	Increasing of thermal resistance; higher on-state dissipation	Periodic check of pressure contact parts of testing equipments

FIG 5: FIELD APPLICATION FAILURE MODE EFFECT ANALYSIS				
OPERATING CONDITION		FAILURE MODE	FAILURE EFFECTS	
			PHYSICAL	ELECTRICAL
ELECTRICAL STRESS	Overvoltage	Junction breakdown	Melting on silicon edge near or below the passivation	Anode-cathode short
	Overcurrent	Thermal runaway; current filamentation	Large melting zone in silicon	Anode-cathode short
	Excessive power to the gate	High current density	Melting zone in pilot gate area	Anode-cathode short
	High di/dt at turn on	High current density close to the gate area	Small melting zone in cathode area near the gate	Anode-cathode short, lack of direct blocking capability only
	High dV/dt (static or during turn off)	Capacitive current through the device	Small melting zone in cathode area	Anode-cathode short
THERMAL AND MECHANICAL STRESS	Poor thermal cooling	Thermal runaway	Degradation of junction coating	Lowering of blocking capability, anode-cathode short
	High thermal cycling, high mounting force	Thermal fatigue	Degradation of external and internal contact surfaces	Worsening of conduction and blocking characteristics
	Low mounting force	Higher thermal resistance	Degradation of external and internal contact surfaces	Worsening of conduction characteristics

FIG 6 : FIELD FAILURE DATA FOR HIGH POWER SEMICONDUCTOR					
DEVICES	APPLICATION	TTT (x 10^6 h)	FAILURE RATE λ (FIT)	CONFIDENCE INTERVALS 90% (FIT)	
TRANSISTOR MODULE	• Urban Train • Subway	4.0	1875	782	3125
STUD DIODE (PRESS-PACK INSIDE)	• Urban Train • Subway	2.5	210	101	330
PRESS- PACK FAST DIODE	• Chopper locomotive • Auxiliary locomotive loads	92.4	283	202	379
PRESS-PACK FAST THYRISTOR	• Chopper locomotive • Auxiliary locomotive loads	95.4	1565	1165	1948
PRESS-PACK GTO	• Urban Train • Subway	1.6	2770	1875	3856

NOISE DIAGNOSTICS OF PN JUNCTION POWER DEVICES

J. Šikula, B. Koktavý, P. Vašina, Z. Chobola, P. Schauer and M. Kořenská
Department of Physics, Technical University of Brno
Zizkova 17, 60200 Brno, Czech Republic, phone/fax: ++ 425 41 21 11 25

ABSTRACT

In this paper studies of noise properties of power devices, namely, high voltage high power rectifier diodes are reported. It has been found that all samples exhibited nearly ideal behavior in the PN junction forward direction, their noise magnitude being very close to the theoretical value. Attention was therefore focused on the reverse direction where some samples exhibited deviations from the ideal behavior. Most of the samples exhibited extremely low level of the $1/f$ noise which gives evidence of very high quality of their technology. On a very small fraction of the sample number excess noise of the $1/f$ type was measurable. Another group of samples exhibited so called "soft breakdown". On these samples a more pronounced $1/f$ noise, g-r noise, burst and microplasma noise was observed.

1. INTRODUCTION

Noise signal in devices that are produced in a well established and stable technology must be stationary, ergodic and its amplitude distribution must be Gaussian. It has been established that stationary electrical noise can be represented by Markovian processes. For Markovian processes it holds that the random deviation from the mean value of any quantity is directly proportional to its mean value.

Fig. 1. shows a plot of the noise current I_N versus the DC reverse current I measured at f = 380 Hz. The plot can be approximated by a straight line, $I_N = K'.I$, where K' is constant. We propose that $K = K'/\sqrt{\Delta f}$, where Δf is the equivalent noise band width, can be used as a quality indicator. The slopes K for various samples are given in Table 1.

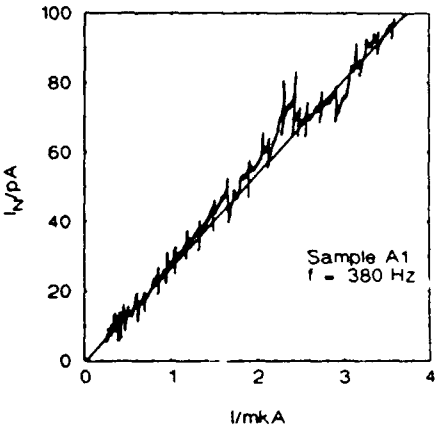


Fig. 1 Noise current I_N versus the reverse current I plot.

Table 1 The slopes K.

sample	A1	A2	A3
K ($Hz^{-1/2}$)	$8{,}0.10^{-6}$	$5{,}5.10^{-6}$	$9{,}2.10^{-6}$
sample	**B2**	**B7**	**B15**
K ($Hz^{-1/2}$)	$5{,}36.10^{-6}$	$1{,}0.10^{-5}$	$2{,}8.10^{-5}$

2. FREQUENCY DEPENDENCE OF THE NOISE SPECTRAL DENSITY

Frequency dependence of the noise spectral density $S_U(f)$ is, in general, a superposition of the $1/f$ noise, g-r noise, burst noise, shot noise and thermal noise. The cutoff frequency between the $1/f$ and thermal noise can be used as a measure of the quality of technology. To distinguish between the g-r and the $1/f$ noise it is suitable to use the $S_U f$ versus frequency plot. In this case, the $1/f$ noise component is manifested by a constant value of $S_U f$ and is easily distinguished from other kinds of noise. The g-r component makes the $S_U f$ curve reach a local maximum, from which the characteristic frequency f_{gr} and the noise time constant τ_{gr} can be easily determined. This is demonstrated in Fig. 2, where it is seen that the characteristic frequency f_{gr} = 19 Hz and the noise time constant $\tau_{gr} = 1/2\pi f_{gr}$ = = 8.4 ms.

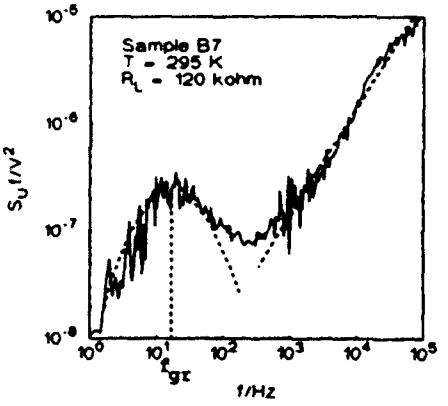


Fig. 2 Plot of the product $S_U f$ vs. frequency.

The other noise generation mechanisms are shown in Fig. 3. At *DC* voltages below 600 V only thermal noise is observed. At 600 volts the curve indicates the presence of a low-frequency $1/f$ noise. When increasing the *DC* voltage over 600 V the $S_U f$ curve reaches a maximum corresponding to a two-state g-r noise with a time constant of about 0,1 s.

Onset of a mechanism generating another g-r noise is apparent at voltages higher than 800 volts.

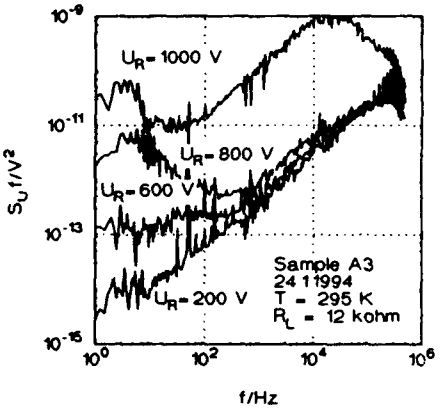


Fig. 3 Plots of the product $S_U f$ vs. frequency for different values of the reverse DC voltage.

The pertinent time constant is of the order of hundreds of microseconds. At still higher voltages avalanche noise is observed.

3. NOISE VOLTAGE SPECTRAL DENSITY VERSUS THE REVERSE DC VOLTAGE

Noise voltage spectral density S_U versus the reverse DC voltage plot was measured across the load resistance. This corresponds to Fig. 4. where a plot of S_U versus the PN junction reverse voltage U is presented. Region *a)* corresponds to the thermal, shot and background noise, region *b)* represents a two-state g-r noise, and, finally, region *c)* may belong to an avalanche noise.

The features of the diodes important for the reliability prediction can be assessed by means of reliability indicators, which were introduced by Jenson and Moltoft in 1986 (Ref.1-2, 4). For the reverse direction we have introduced recently the reliability indicator x_2, which is, by definition $x_2 = U_{BR}/U_S$, where U_{BR} is the breakdown voltage of a nearly ideal sample, U_S is the voltage of a "soft" breakdown. Good diodes feature x_2 = 1, whereas diodes with soft breakdown have x_2 higher than unity.

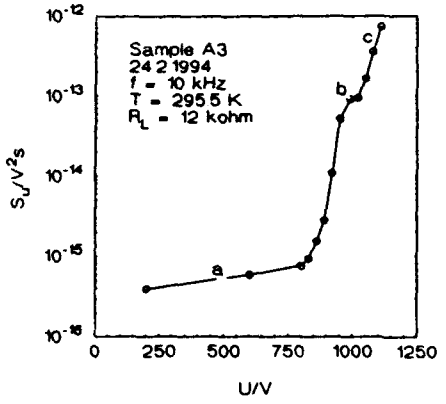


Fig. 4 Noise voltage spectral density S_U versus the reverse voltage U plot indicating several noise sources.

4. BURST NOISE

Burst noise has been observed in many semiconductor devices, bipolar as well as unipolar. It is now established that the burst noise generators are related to defects in the crystal structure of the devices. The nature of the noise source consists in current modulation which is due to charge carrier trapping (Ref. 3).

An usual kind of the burst noise was observed at the room temperature in a fairly wide range of the reverse voltages, typically from 600 V to about 1000 V. As an illustration we present Fig. 5, where the DC reverse voltage across the diode (sample No B7) is 810 V and the AC voltage across the load resistor is displayed in time domain. The amplitude of the current bursts is of the order of 10^{-7} A, the impulse on time depends on the reverse DC voltage and its range is very wide.

Extremely wide range of time constants can be seen. In Fig. 5 the times constant is of the order of a few tens of milliseconds, whereas in Fig. 6 it is as long as several hundreds milliseconds. Occurrence of the burst noise may serve, among others, as a Reliability Indicator.

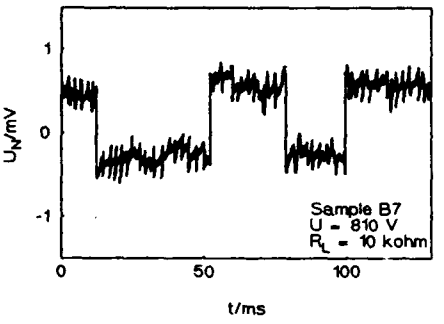


Fig. 5 Burst noise in time domain.

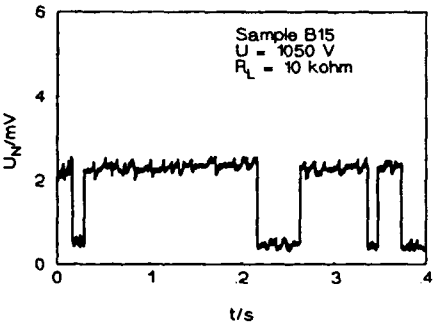


Fig. 6 Burst noise in time domain.

5. CONCLUSIONS

1. No excess noise has been observed within the whole range of measured forward currents.

2. The noise factor K depends on the frequency and on the sample technology. At f = 400 Hz the experimental values are around $K \approx 10^{-6}$ Hz$^{-1/2}$. The factor K is approximately constant over the whole range of reverse currents, being specimen specific.

3. The *g-r* noise is dominant in the low-frequency region. The noise time constant τ depends on the voltage U_R and its experimental value ranges from milliseconds to seconds.

4. Degradation processes are strongly influenced by a high electric field intensity and temperature. This applies particularly to the technology that determines the reverse direction parameters and appears to be of uppermost importance.

6. REFERENCES

1. Šikula J. and Touboul A. "Noise spectroscopy, diagnostics and reliability of electronic devices", Proc. of the 12th Int. Conf. on Noise in Phys. Systems, (ed. Handel P. and Chung A. L.), St.Louis, Aug. 1993, pp. 206-211,

2. Jones B.K. "Electrical Noise as a Measure of Quality and Reliability in Electronic Devices", Advances in Electronics and Electron Physics, Vol.67, 1994, pp 201.

3. Simoen E., Magnusson U. and Claeys C., Appl. Surf. Science, 63 (1993) 285.

4.Wolfgang E., Anger K., Fiegl B., Kuhnert R. and Turkes P. "Thermal management of power semiconductors: modeling, circuit simulation, measurement and reliability", Invited paper, 4th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis, Bordeaux, France, 4 - 7 October, 1993.

ACELERATED TEST OF POWER MOSFET

M.E.Jimenez, E. Nogueira, Q. Perez, J.E. Vila
Reliability & Materials Division
TELEFONICA INVESTIGACION Y DESARROLLO
C/ Emilio Vargas, 6
28043 Madrid
Spain
Phone: 34-1-337.44.89
Fax: 34-1-337.42.12

ABSTRACT

The HAST results on a power MOSFET are presented. The samples were subjected to several power, temperature and humdidity stress conditions and the acceleration factors were evaluated. The same failure mechanism, die attachment failure, was found in the all samples and its activation energy was calculated.

1.INTRODUCTION

This paper presents the results of high accelerated stress test (HAST) on different conditions, carried out on a pressure cooker over a type of power MOSFET.

The objetives of this study were to find the failure mechanism at high temperatures and normal relative humidity conditions, to evaluate the activation energy of the failure mechanism and to calculate the temperature, humidity and electical power acceleration factors in these devices.

The purpose of the presure cooker chamber is to perform accelerated tests instead of tipical test (85 °C/85 %HR) reducing the test time. However there is some controversy about the results that can be obtained with these chambers, because the severe conditions of high temperature, humidity and hight presure, not always reproduce failures which can appear in the real aplication of the devices. Our experience shows that for some componentes, like SRAM and power devices, these chambers are valid, but for other components, like LED, are not valid, because they reproduce failures which are not obtained in the field. One advantage of these chambers test is that it permits to control the humidity over 100° C. This facility allowed us to perform accelerated test in temperature but holding up the humidity close to normal conditions. The purpose is to reproduce the real humidity gradient that these devices present in the aplication.

Due to the fact that these devices show high power dissipation at the semiconductor surface, it was not an objetive to prove the resistance to moisture penetration on plastic packaging. However one test was performed to evaluate the humidity acceleration factor.

2.TEST SAMPLES

Test were carried out on four lots composed each one by five samples of power MOSFET without heat sink. The technology of the transistors was HEXFET and the package type was TO-220. The most commun aplication of these devices is in DC/DC converters and power supplies.

3.FAILURE CRITERIA

The on-resistance $R_{DS(ON)}$ was the selected parameter to detect the failure during the tests. Measurements were made every three minutes and the failure criteria was an on-resistance greater than 1.2 Ohm. That means, aproximately, three times greater than the typical on-resistance at test temperature.

4.TEST CONDITIONS

A pressure cooker was used, that allowed us to control the humidity at temperatures over 100 °C. The following four conditions were selected: 145 °C/65% RH with 1.3 W, 135 °C/85% RH with 1.3 W, 135 °C/65% RH with 1.3 W and 135 °C/65% RH with 1 W.

The transistors were biased in the ohmic region and in the safe operating area. The DC gate and drain power supplies were independent, to avoid the

influence of possible drain failure in one sample, that can affect the bias voltage gate of the rest of the samples. Furthermore this allows us to obtain a good stabilization of the working point.

5.TEST RESULTS

The obtained results were plotted on Weibull paper to evaluate the shape parameter and the mean time to failure (MTTF). The Figure 1 Weilbull plot of the shows the test results and the Table 1 shows the test conditions, shape parameter and MTTF for each test.

The four tests results showed similar shape parameter, that means that only one kind of failure mechanism was present. A failure analysis were performed. The electrical analysis, by a Parameter Semiconductor Analyzer, showed a serial resistance with diode drain-source. The physical analysis, by X-Ray, showed voids on the adhesive between the die and the lead frame, the failure mechanism was related with problems on the die attachment. These results were present in all failed transistors and confirm that only one failure mechanism was present. Below is an explanation with more details about the failure analysis.

In order to evaluate the temperature and humidity acceleration factors, several models were tried (Lawson, Eyering, Weick, Peck, Peck and Zierdt, Strochle). The best one was Peck and Zierdt, Equation 1. As the test 2 and 3 from Table 1 show, the humidity has not to much influence in the reliability of these devices, increassing the relative humidity by a 20%, the MTTF decrease only from 69,4 to 42,5 hours. This result is confirmed by the selected model. Moreover a corrosion failure mechanism was not observed in the failure analysis, this means that no moisture penetration was present.

$$MTTF = A\, e^{E_a/KT}\, e^{BLnH} \quad [1]$$

As is showed from the test 2 and 3 we have obtained a large influence with the power, from this reason, we have decided to include in the model, a factor that consider the power efect, based on the Inverse Power Law, Ecuation 2.

$$\frac{MTBF1}{MTBF2} = \left(\frac{P2}{P1}\right)^n \quad [2]$$

The proposed model is expresed in the Ecuation 3. With this model we obtained an activation energy for the die attachment failure mechanism of 2.34 eV. The value calculated for the B constant is -1.57, the A constant is $5.4*10^{-21}$ and finally the n constant is 4.12.

$$MTTF = Ae^{E_a/KT}\, e^{BLnH} \left(\frac{3.5}{P}\right)^n \quad [3]$$

This model is only applicable for devices without heatsink in case with high dissipation power. The data obtained with the proposed model, were compared with the MIL 217F and the data manufacturer. We observed that the data manufacturer are lightly optimism and the data from MIL are very pessimist.

6.FAILURE ANALYSIS

We carried out an electrical analysis on the failed components to check the mode of failure. The Figure 4 represents the curves of the intrinsic diode drain-source for two devices, one of them good (A) and the other one wrong (B). It can be seen in two curves the integral body-drain diode, but in the wrong case show a serial resistance. This resistance can be calculated by the inverse of the gradient I/V. The slope changed when the body of device was pressed. The electrical behaviour the whole of the failed devices was similar.

In the physical analysis, by X-ray, we observed voids on the adhesive between the die and lead frame. The Figure 2 shows the X-ray photography for a failed transistor and the Figure 3 for a good one. These voids on the adhesive of the component, present in all failed devices, explain the serial resistance finded in the electrical analysis. We can deduce that the failure mechanism in these test, was die attachment.

The faillure mechanism we have obteined with these devices tests is no very usual in their typical application, then the proposed model can not be applicable for general behaviour of these transistors.

7.CONCLUSIONS

We have obtained the following conclusions:

1.- The same failure mechanism, die attachment failure, was found in all tests, the evaluated activation energy was 2.34 eV.

2.- Although a pressure cooker was used, the

failure mechanism was mainly due to temperature and electrical power while the humidity has lower influence.

3.- The proposed model is only applicable for high dissipation power and devives without heat sink.

Table 1.-Test conditions, shape and MTTF results.

	Test-1	Test-2	Test-3	Test-4
T (°C)	145	135	135	135
RH (%)	65	85	65	65
P (W)	1.3	1.3	1.3	1
MTTF (h)	23.4	42.5	64.9	191.6
ß	2	2.1	2.7	2

Figure 1.-Weibull plot.

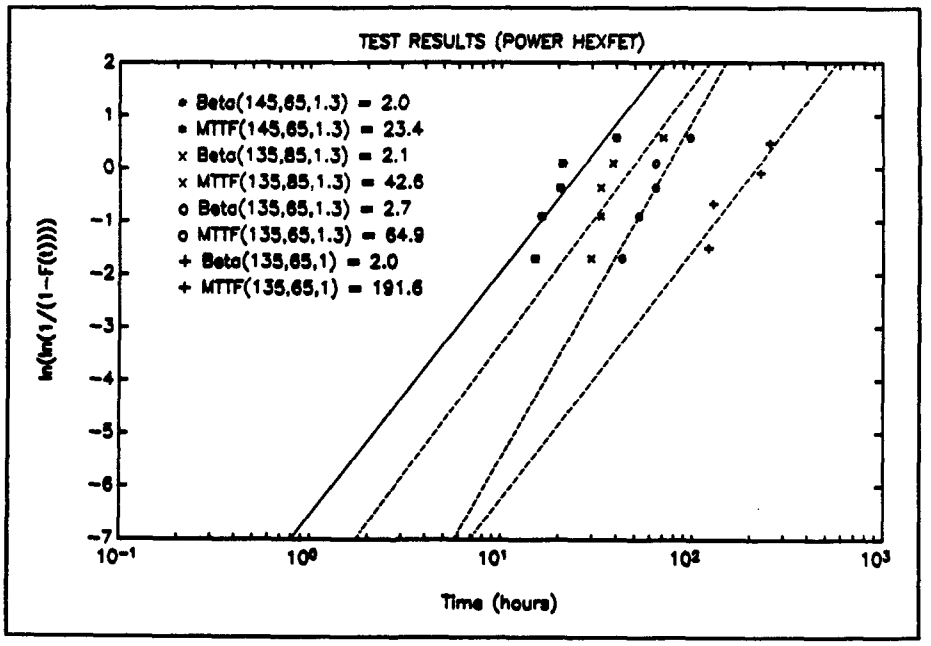


Figure 2.-Photography X-ray in a failed transistor.

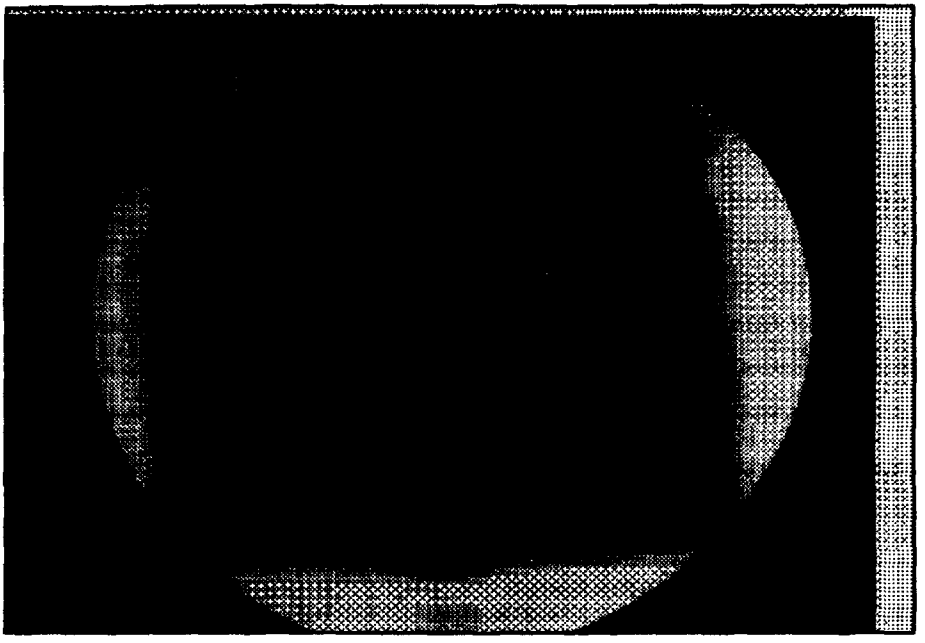


Figure 3.-Photography X-ray in a good transistor.

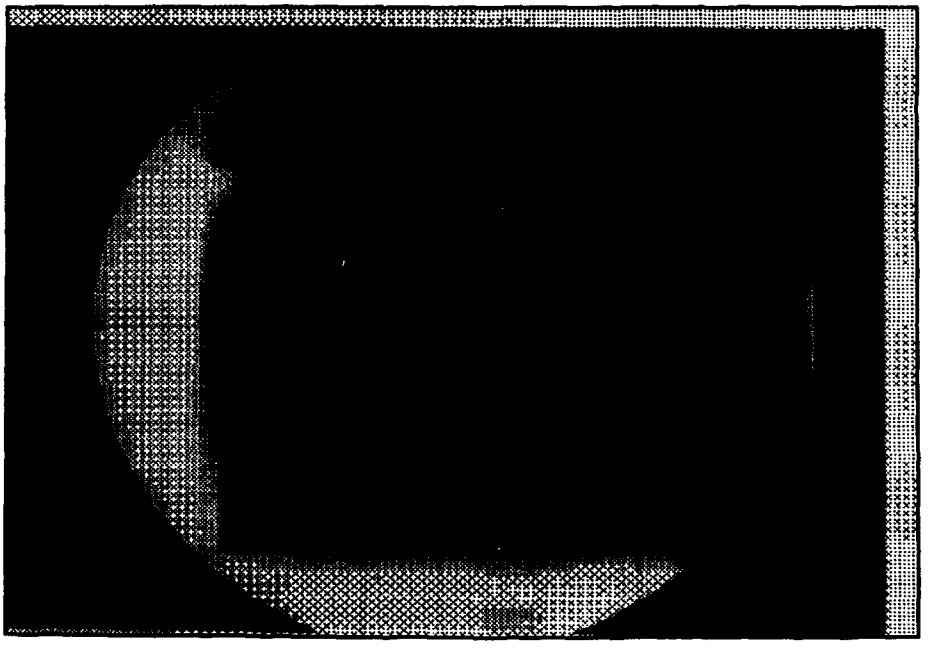
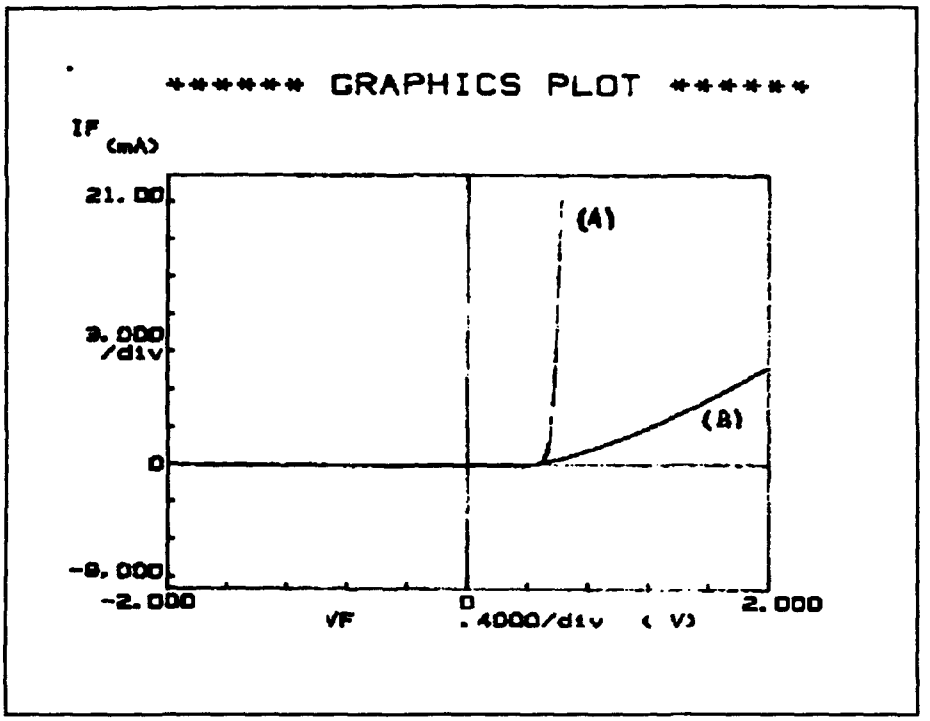


Figure 4.-Curves I-V for a good transistor (A) and failed one (B).



BEHAVIOR OF LASING WAVELENGTH CHANGE DURING DEGRADATION IN BURIED-HETEROSTRUCTURE (BH)-TYPE DFB LASER

Mitsuo FUKUDA, Fumiyoshi KANO, Takeshi KUROSAKI, and Junichi YOSHIDA
NTT Opto-electronics Laboratories
3-1, Morinosato Wakamiya, Atsugi-shi, Kanagawa, 243-01 Japan

ABSTRACT: The behavior of lasing wavelength change during degradation is experimentally clarified in a Buried Heterostructure (BH) InGaAsP/InP DFB laser, and the mechanism of the wavelength change is theoretically analysed.

1. INTRODUCTION

Optical fiber transmission systems employing dense wavelength-division-multiplexing (d-WDM) techniques have attracted special interest very recently. In addition, proto-types of optical fiber transmission systems employing optical fiber amplifiers are beginning to be put in the field. In all those systems, the stability of the lasing wavelength and the spectral linewidth of optical sources are key factors determining reliability. The wavelength stability is often required to be less than a few angstroms all through the service term.

The distributed feedback (DFB) laser is commonly used as an optical source in those systems. Their wavelength stability for long-term aging has been reported, although the correlation between degradation mechanisms and the wavelength change were not clarified [1, 2]. In this paper, we experimentally clarify the behavior of lasing wavelength change in BH type InGaAsP/InP lasers during degradation, and the mechanism of the change is discussed.

The degradation observed in the present study followed a common pattern in BH-type InGaAsP/InP DFB lasers, and scarcely depended on the structure of the active region, bulk structure, Multiple Quantum Well (MQW) structure, and strained-MQW structure. Thus, the correlation between the degradation and the wavelength change is qualitatively the same for all types of InGaAs(P)/InP BH-type DFB lasers. The following sections discuss the behaviors of lasing wavelength change during degradation in BH-type InGaAsP/InP MQW DFB lasers.

2. DEVICES AND AGING CONDITIONS

The devices used are 1.55-μm-band InGaAs/InGaAsP BH-type MQW DFB lasers with a phase shift. An MQW structure with a 1.15-μm-composition guide layer over 0.15 μm thick is grown by MOVPE on an n-type InP substrate [3]. The MQW structure consists of 3-5 InGaAs well layers and InGaAsP barrier layers. These layers are buried by InP layers grown by LPE after the mesa-structure formation. The cavity length is 900 μm, and both facets are coated with an anti-reflecting film. The coupling coefficient of the grating is controlled to 20-35 cm^{-1}. These lasers were mounted on a Si heat sink with junction-up or -down configuration. At 25 ℃, the threshold current was between 20 mA and 40 mA, and the slope efficiency was between 0.1 W/A and 0.15 W/A. More than 100 devices were aged under various conditions.

3. DEGRADATION BEHAVIOR

In BH-type InGaAsP/InP lasers, the main cause of the degradation is BH interface degradation as shown in Fig. 1 [3, 4]. In the initial stage of the degradation, nonradiative recombination current at the BH interface increases while threshold carrier density remains nearly constant. That is, the degradation is in the double heterojunction diode. This degradation mode is quite different from that of AlGaAs/GaAs lasers and peculiar to InGaAsP/InP systems [5]. In the second stage, the degradation of an optical cavity is gradually generated and is added to the degradation of the heterojunction diode. These degradation modes directly or indirectly influence the lasing wavelength.

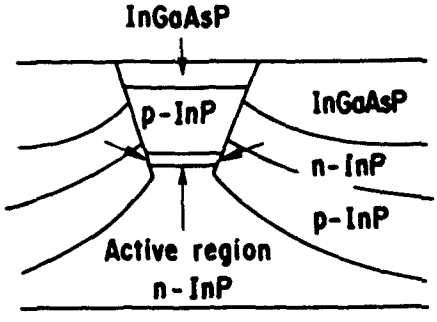


Figure 1 Schematic diagram of BH interface degradation. The arrows show the estimated degradation part.

Typical curves for the change in lasing wavelength are shown in Fig. 2. There is very little change in the initial stage, although the threshold current clearly increases. After severe degradation, the wavelength begins to change. In DFB lasers, the wavelength shift, $\Delta\lambda$, during degradation is determined by the change in effective grating pitch, Λ,

$$\Lambda = \lambda/2n, \qquad (1)$$

where λ is the lasing wavelength within the cavity and n is the refractive index in the active region. This effective grating pitch change is introduced by the refractive index change in the active region. That is, wavelength shortening is caused by the plasma effect due to the injected carrier, $-\Delta\lambda_{plasma}$, and the lengthening due to Joule heating, $\Delta\lambda_{Joule}$, and can be expressed

$$\Delta\lambda = -\Delta\lambda_{plasma} + \Delta\lambda_{Joule}. \qquad (2)$$

Here, $-\Delta\lambda_{plasma}$ is a function of injected carrier density, and may be expressed for 1.55-μm-band InGaAs(P) systems as follows:

$$-\Delta\lambda_{plasma} = -2\,\Delta n\,\Lambda = -2\,\Lambda\,(6 \times 10^{-21})\,\Delta N, \qquad (3)$$

where ΔN is the change in carrier density, which is on the order of 10^{18} cm^{-3} after lasing, Λ is on the order of

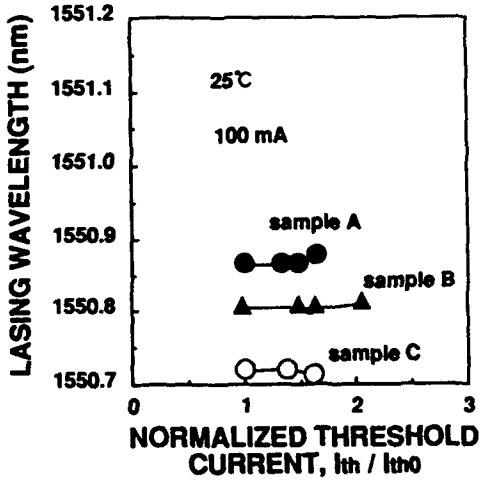


Figure 2 Lasing wavelength change at a constant current of 100 mA during degradation. The aging was carried out under a constant current of 450 mA at 50℃.

10^{-3} cm. Therefore, from Eq (3), the wavelength change due to the plasma effect would be more than 1 nm after a 50% increase in threshold current, if the increase in threshold current was due solely to the increase in threshold carrier density. However, Fig. 1 shows an wavelength shift of less than 0.1 nm. This is because of the constant threshold carrier density in the initial stage

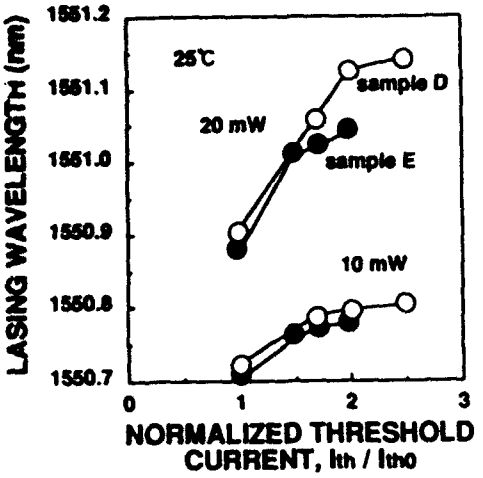


Figure 3 Lasing wavelength change at a constant output power of 10 mW and 20 mW during degradation. The aging was carried out under a constant current of 450 mA at 50℃.

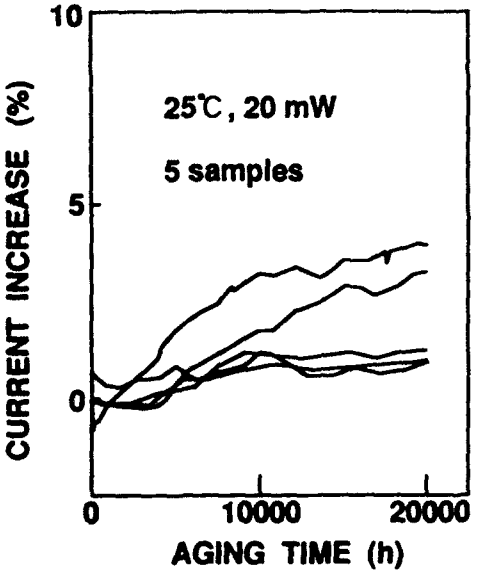


Figure 4 Typical aging of 1.55 μm-BH-type InGaAsP/InP MQW DFB lasers.

of the BH interface degradation as discussed above. For $\Delta\lambda_{cooler}$, the wavelength sfift is about 0.1 nm/℃. Therefore, the wavelength lengthening becomes dominant in BH-type InGaAsP/InP lasers because of the heating in the active region, even with electric-cooler control. This trend is intensified under constant output power operation, as shown in Fig. 3, because the injected current is increased in order to keep output power constant. Therefore, the pattern of wavelength change due to heating differs with the mounting configuration. The devices with junction-up configuration tend to exhibit rapidly increasing wavelength during degradation when compared with those with junction-down configuration [6]. Anyway, these wavelength shifts depend on the device degradation, and devices with no degradation scarcely showed any wavelength change even after 20,000 hours aging (Fig. 4). Consequently, it can be said that we can obtain lasers applicable to the d-WDM systems and systems employing optical fiber amplifiers because of the degradation behaviors indicated above.

4. CONCLUSION

The wavelength change of the DFB laser during degradation is mainly determined by the Joule heating and is quite small because of the behaviors of BH interface degradation. The wavelength change behavior should be considered in estimating reliability of DFB lasers applied to systems requiring long-term stability of lasing wavelength, such as d-WDM systems and systems employing optical fiber amplifiers.

5. REFERENCES

[1] Vodhanel R. S., Krain M., Wagner R. E., and Sessa W. B. "Long-term wavelength drift of the order of -0.01 nm/yr for 15 free-running DFB laser modules", Tech. Dig. OFC'94, IEEE/OSA, (San Jose, California), 1994, pp. 103-104.

[2] Chung Y. C. and Jeong Jichai "Aging-induced wavelength shifts in 1.5-μm DFB lasers", Tech. Dig. OFC'94, IEEE/OSA, (San Jose, California), 1994, pp. 104-105.

[3] Fukuda M., Kano F., Kurosaki T. and Yoshida J. "Degradation behavior of highly coherent 1.55 μm long-cavity multiple quantum well DFB lasers", Proc. ESREF'92, 1992, pp. 241-244.

[4] Chiorboli G., Vanzi M., Ronchini M. C. and Fantini F. "Electrical characterization for failure analysis of DHBC InGaAsP/InP laser diodes", Proc. ESREF'92, 1992, pp. 237-240.

[5] Fukuda M. "Reliability and Degradation of Semiconductor Lasers and LEDs", Artech House, Norwood MA, 1991.

[6] Fukuda M. "Aging characteristics of InGaAsP/InP DFB lasers", OPTOELECTRONICS-Device and Technologies, Mita Press, Tokyo, vol. 3, 1988, pp.177-187.

RELIABILITY ASSESSMENT OF A 1480nm MQW PUMP LASER DIODE FOR ERBIUM DOPED FIBER AMPLIFIERS.

Anthony Hawkridge and Adam Carter
Hewlett-Packard Fibre Optics Components Operation,
Whitehouse Road, Ipswich, Suffolk, IP1 5PB.

1. INTRODUCTION

High power 1480nm pump lasers are required as the pumping sources for erbium doped fiber amplifiers for long haul transmission systems (Ref 1,2). Such systems may require individual component lifetimes in excess of 25 years.

In this paper we describe two 1480nm multi-quantum well (MQW) pump laser structures, and summarise the work undertaken to demonstrate their suitability for such high reliability applications.

2. STRUCTURE AND CHARACTERISATION

Two types of planar buried heterostructures (BH) grown entirely by atmospheric MOVPE (Ref 3) were used in the reliability assessment. These utilised a Graded Refractive-Index Separate Confinement Heterostructure (GRINSCH) or Separate Confinement Heterostructure (SCH).

Figure 1 shows the conduction band diagram of the GRINSCH structure. The structure consists of 9 GaInAsP (λ = 1.55μm) wells with 8 GaInAsP barriers (λ = 1.2μm). The thickness of the quantum wells was determined so that the lasing wavelength would be 1480nm. The graded region has three quaternary layers (λ = 1.2μm, 1.1μm, 1.0μm).

In the SCH structure, shown in figure 2, the confinement region maintains the same composition as that used for the barrier layers. The thickness of the confinement layers was determined so that both the SCH and GRINSCH structures had the same calculated confinement factor.

During subsequent processing, anti-reflective (R = 3-10%) and highly reflective (R = 93%) coatings were deposited on either facet of the 750μm long laser diodes. The devices were mounted p-side down on diamond heatsinks which were themselves bonded to a suitable carrier for the reliability assessment.

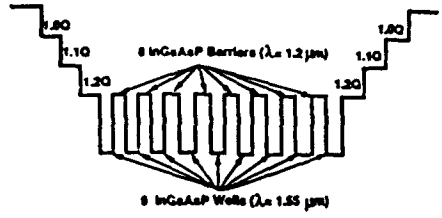


Figure 1:- Conduction band schematic of nine-well GRINSCH structure.

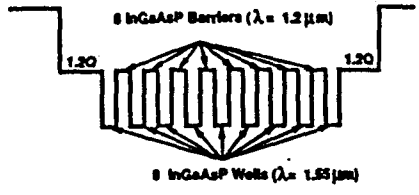


Figure 2:- Conduction band schematic of nine-well SCH structure.

Pump laser diodes gave typical threshold currents of 25mA at 30C; over the temperature range 0C to 60C threshold currents varied between 17 - 40mA with corresponding slope efficiencies of 0.35 - 0.50W/A. Confinement was demonstrated by parallel and perpendicular far fields of 28° and 35°; typical coupling of 50% into single mode fiber realised output powers of 60mW at 350mA and 30C.

3. INITIAL RELIABILITY EVALUATION

3.1 Screening

Devices for the initial reliability evaluation were subjected to a constant current burn-in of 350mA at 125C for 24 hours. A failure criterion of >20% change in threshold was used for screening purposes. This condition is the same as that routinely applied to production devices.

3.2 Evaluation

The reliability of the pump laser diode was initially assessed at a single temperature of 50C. Devices were lifetested using automatic power control (APC) where the optical output of the devices was maintained at a constant level whilst the laser drive current was allowed to vary.

A total of 64 devices completed in excess of 500000 device hours at 50C. A typical optical output power for these devices was 40mW. Drive current degradation rates of <0.7%/1000hours were observed at 50C.

Long term degradation of the buried heterostructure device has previously been modelled using equation (1):-

$$\frac{I - Io}{Io} = At^n \qquad (1)$$

Where Io is the initial drive current, I is the drive current at time t, and A and n are constants (Ref 4). Equation (1) is used in the linear form (2):-

$$\text{Log}\,[\{I - Io\}/Io] = n\text{Log}\,t + \text{Log}\,A \qquad (2)$$

Applying this model to lifetest data using a least squares fitting process allows an estimate of n and A to be made. It is then possible to extrapolate to the failure criteria for current change and obtain a predicted time to failure.

The prediction model (2) was applied to the data from these initial tests. However, due to the low rates of degradation observed, realistic predicted lifetimes were obtained from relatively few devices only. Where it was possible to calculate a time to failure, the lifetime was plotted on a cumulative probability plot. A median lifetime of >1.5e5 hours at 50C was suggested, failure being defined as a 50% increase in drive current for constant light output. As the majority of devices did not show sufficient degradation on lifetest for a time to failure to be calculated, this median lifetime is thought to be a pessimistic value.

4. TEMPERATURE AND CURRENT ACCELERATION

4.1 Test Conditions

The reliability of the pump laser diode under high current stress conditions was considered. A series of lifetests at temperatures of 30C and 70C, and drive currents of 400mA to 550mA commenced in an attempt to measure the effect of temperature and current on lifetime.

The devices were lifetested under a constant drive current condition with interim characterisations made at 30C. A reference optical power of 70mW at 30C was used for monitoring changes in drive current.

The conditions of test, sample sizes used, and test hours completed are summarised in table 1.

Table 1:- Temperature/Current Activation Lifetests

Condition	Sample Size	Test Hours
30C/400mA	10	6208
30C/450mA	7	7118
30C/500mA	9	5955
30C/550mA	8	5855
70C/400mA	6	5878
70C/450mA	7	6892
70C/500mA	8	5749

4.2 Screening

In addition to the constant current screen described in section 3.1, devices for these tests were subjected to a second burn-in screen. Subsequent work to this study had shown the need for a second burn-in process to remove devices which may have shown a higher rate of degradation then desired. The failure criteria suggested was retrospectively applied to the data collected from these lifetest after 100 hours had elapsed. Failure criteria of threshold >50mA at 30C, change in threshold >15%, and optical power <70mW at 30C were used. Approximately 9% of devices failed this second screen.

4.3 Results

Typical results from a 30C lifetest are shown in figure 3. Low rates of degradation were observed. An attempt was made to use equation (2) for predicting lifetimes, but insufficient current degradation had been observed for meaningful predictions to be made even for laser drive currents of 550mA.

Measurable degradation was observed from the 70C lifetests (figure 4) enabling predicted times to failure to be obtained using equation (2). Predicted times to failure were used to produce a cumulative probability distribution for each lifetest condition. The 70C results are summarised in table 2.

Table 2:- 70C Lifetest Results

Condition	Median Lifetime
70C/400mA	3e4 hours
70C/450mA	4e4 hours
70C/500mA	2e4 hours

The results suggested that median lifetime was insensitive to drive current over the current range 400mA to 500mA.

The data from the three 70C lifetests has been pooled and an overall cumulative probability plot for 70C obtained (figure 5). Grouping the data gave an overall median lifetime of 3e4 hours at 70C with a standard deviation (σ) of 0.64. (Ref 5)

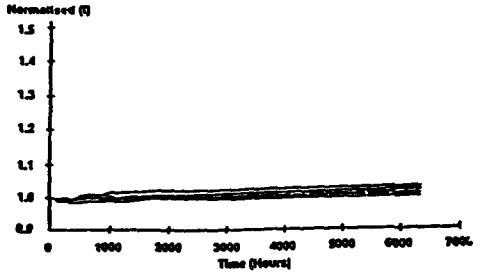


Figure 3:- Typical lifetest results from a pump laser batch at 30C and drive current of 400mA. Normalised drive current for 70mW/30C.

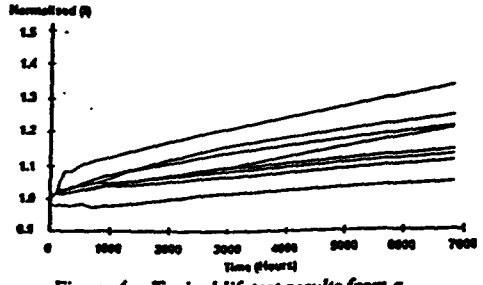


Figure 4:- Typical lifetest results from a pump laser batch at 70C and drive current of 500mA. Normalised drive current for 70mW/30C.

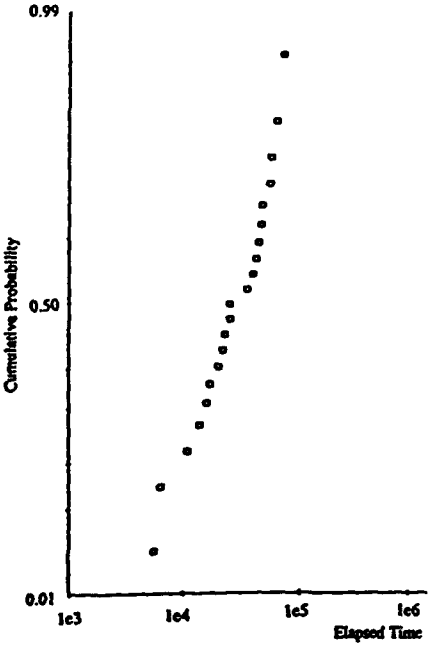


Figure 5:- Cumulative probability plot for grouped 70C data.

4.4 Thermal Activation Energy

The absence of significant degradation at 30C precluded quantifying the thermal activation energy between 30C and 70C. However, the Arrhenius equation (3) was applied to the 70C data and the 50C data from the initial reliability assessment (Ref 5)

$$\text{Acc factor} = \frac{\{Ea(1 - 1)\}}{\{k\ \{T2 - T1\}\}} \qquad (3)$$

Where Ea is the activation energy, k is Boltzmanns constant, and T1 and T2 are absolute temperatures.

The median lifetimes obtained at 70C and 50C suggested an activation energy for current degradation of 0.8eV. Other buried heterostructure designs have previously demonstrated activation energies in the range 0.6 to 1.0eV.

5. WEAR-OUT AND STEADY-STATE FAILURE RATES

The activation energy calculated in section 4.4 suggests a median lifetime of 1.8e6 hours at 25C. Goldthwaite analysis (Ref 5), performed using the σ value of 0.64 derived from the 70C lifetests, indicates that the wear-out failure rate will not exceed 1 FIT in 20 years of service.

An activation energy of 0.35eV for steady state failure, as recommended for active optical devices (Ref 5), has been applied to the testing completed in sections 3 and 4. The equivalent of 4.45e6 device hours at 25C have been completed; this suggests a steady state failure rate of <205 FITs to a 60% confidence level, and <518 FITs to a 90% confidence level. These steady state failure rates are, however, primarily limited by the elapsed hours on test completed and it is anticipated that the demonstrable steady state failure rate will reduce with the collection of further component hours.

6. CONCLUSIONS

The design and device characteristics of a GRINSCH and SCH 1480nm MQW pump laser diode for high reliability applications has been described. Threshold currents of <40mA at 60C were observed. Typical coupling into single mode fiber realised output powers of 60mW at 350mA and 30C.

A single temperature reliability assessment demonstrated a median lifetime of >1.5e5 hours at 50C. Lifetests performed over a range of currents at 70C suggested that median lifetime was insensitive to drive current; grouping the data from these tests gave a median lifetime of 3.0e4 hours at 70C. Low rates of degradation were observed at 30C precluding a prediction of lifetime at this temperature. However, an Arrhenius analysis of 50C and 70C data suggested an activation energy for current degradation of 0.8eV.

This activation energy gave a predicted median lifetime of 1.7e6 hours at 25C. With a sigma value of 0.64 obtained from the 70C test, Goldthwaite analysis indicated that the wear-out failure rate for the 1480nm MQW pump laser diode would not exceed 1 FIT in 20 years of service.

7. ACKNOWLEDGEMENTS

The authors are grateful to the staff of BT Laboratories, Martlesham, Ipswich for assistance with the lifetesting and data collection of the 1480nm pump laser diode reliability assessment.

8. REFERENCES

1. H.Asano, S.Kanano, M.Kawaradani, M.Kitatmura, I.Mito, Optical Fiber Communication Conference, Vol 4 1991 OSA Technical Digest Series (Optical Society of America, Washington DC 1991), paper WG1.

2. S.Akiba, K.Goto, S.Yamaoto, H.Wakabayashi, "Recent Progress of Transoceanic Optical Transmission Systems" in Proceedings of 18th European Conference in Optical Communications, Berlin, Germany, 27 September - 1 October 1992, Vol 2, pp.719-726.

3. D.Cooper, S.Cole, W.Devlin, R.Hobbs, A.Nelson, J.Regnault, A.Skeats, S.Sim, P.Spurdens, "High Performance, High Reliability Buried Heterostructure Lasers By MOVPE", Electronic Letters, Vol 24, pp.519-521, 1988.

4. S.Richie and D.[illegible]wman, "Gradual Degradation of [illegible]ouble Heterostructure Lasers", IEEE J[illegible]um Electron, QE9, No.2, pp.300-305, 1973.

5. Bellcore Technical Advisory TA-TSY-000983, "Reliability Assurance Practices for Optoelectronic Devices for Loop Applications", Issue 1, January 1990.

THE THERMAL RESISTANCE: A RELIABILITY MONITOR FOR LASER-DIODE MODULES

Paolo Malberti and Mauro Ciappa

Swiss Federal Institute of Technology (ETH)
Reliability Laboratory, ETH-Zentrum, CH-8092 Zürich, Switzerland

Abstract – The measurement of the thermal resistance of laser diodes may give useful pieces of information to determine the failure mechanisms not only of the laser diode but also of the module in which the it is housed. This has been clearly shown by failure analysis of commercially-available laser-diode modules for telecom application which failed after a temperature-cycling test. Among the five different failure mechanisms that have been identified, three failure mechanisms are associated with thermal-resistance increase, in consequence of mechanical displacement either of the laser-diode die or of the heat-sink. These two forms of displacement (which have different causes) change the alignment between the die, the lens and the fiber-tip, and consequently change the external efficiency of the laser-diode module, which in some case could also notably increase.
Thus, the thermal resistance proved to be an efficient parameter to monitor the reliability of laser-diode modules, since the it may put into evidence the presence of latent failure mechanisms inside those modules which apparently, relying on the external-efficiency only, are not functionally degrading or, even paradoxically, are functionally improving.

1. INTRODUCTION

The penetration of optical fiber into local-loop communication systems is occurring rapidly. Fiber-in-the-loop installations are undergoing field testing in harsh, unheated or uncooled, environments. These are commonly referred to as uncontrolled environments as opposed to controlled office environment. Laser diode modules in the uncontrolled environment of a pedestal near a subscriber's residence must be designed to withstand rapid temperature cycles, possibly once or twice a day, between storage temperatures as low as –40°C and operating temperatures as high as +85°C, which includes sun loading and heat from the electronics of the powered equipment [1].
The laser diode is a temperature-sensitive device. The most important parameters, e.g. threshold current and emitting wavelength, are strongly temperature dependent. From the reliability point of view, the laser diode needs a stable, homogeneous, low resistive, and low self-heating die-attach to guarantee a uniform distribution of the injected current through the active layer and an efficient drain of the dissipated power to the heat-sink. Accelerated life tests have demonstrated that only laser diodes with a perfect die attach may reach high reliability performance.
Thermal-cycling is one of the most appropriate tests to evaluate the mechanical integrity of a module design. Repeated switching from high to low temperatures stress joints, bonds, interfaces and lens mounting due to the different thermal expansion coefficients of the internal parts. A recent paper [2] reports the results of a thermal-cycling stress test of laser diode modules for telecom application, and a simple life-time model to compute the number of cycles to failure. In order to make predictions on the operational reliability of laser-diode modules, it is essential to have an understanding of the involved failure mechanisms. The measurement of the thermal resistance may give useful pieces of information for this purpose.

2. EXPERIMENTAL

2.1 Samples

Eighty commercially-available laser diode modules of two different manufacturers (A and B) have been tested. The packages are cylindrical-type with single-mode silica-fiber pigtail. In order to take pistoning effects into account [3], half of the modules have been provided with an optical connector. For both module type, the internal hermetically-sealed metal-can contains a ridge-type 1300-nm-wavelength InP laser diode, a focusing and collimating lens, and a rear-faced silicon monitor diode.
The A-type laser diodes are soldered active-side-up onto a silicon heat-sink by an eutectic Au/Sn die-attach. The heat-sink is soldered onto a gold-plated copper stem using a soft Pb/Sn solder. A collimating and focusing zirconium ball lens is positioned close to the front mirror of the laser diode. The lens is soldered into a square pyramid-shaped window in the silicon mounting.
The B-type laser diodes are mounted active-side-down onto a diamond heat-sink by an eutectic Au/Sn die-attach. The heat-sink is soldered onto a gold-plated copper stem using a soft Pb/Sn solder. The Ti/Pt/Au metallization layer of the active-side is isolated from the crystal area surrounding the contact strip by a silicon nitride dielectric layer. The collimating lens is positioned using a metal clip mounting.

2.2 Thermal-Cycling Stress Test

The thermal-cycling stress test has been divided into

four test programs, as indicated in Tab. 1.
Each test program includes a set of ten devices of both manufacturers, and is intended to investigate the homogeneity and the time-dependence of the failure mechanisms activated by different temperature swing, range, and gradient. No electrical power has been supplied to the samples during thermal cycling. The thermal cycling test has been performed contemporaneously at different laboratories using two-chamber climatic systems.
All the devices have been automatically tested before, at scheduled intermediate check points, and at the end of thermal cycling. Before and after thermal cycling, the main electrical and optical parameters of the laser diode and of the monitor diode have been measured at the bench in order to obtain supplementary information for future failure analysis.
Since the investigated laser diode modules are intended for installation in complex telecommunication systems, the tolerances for the drift of the electro-optical parameters must be tighter than usual. Thus the devices are assumed to fail as soon as any electro-optical parameter decreases by 10% (≈ – 0,5 dB) of its initial value.

2.3 R_{th} Measurement Before Thermal Cycling

The junction-to-heat-sink thermal resistance (R_{th}) has been measured for all the devices before and after thermal cycling. The thermal resistance measurements have been performed using a dedicated switch-and-measure circuit, based on the terminal voltage method, [4] and [5].
Seventy-seven out of the eighty devices have been measured. The only exception are three devices (#A26, #B7, and #B23) because of their excessive leakage current. The mean value (μ) and the standard deviation (σ) of the R_{th} values, measured in [°C/W], before thermal cycling are : $\mu = 35.5$, $\sigma = 6.0$ for Man.A , and $\mu = 12.2$, $\sigma = 7.0$ for Man.B . These mean values are typical [6] of InP laser-diodes soldered active-side-up and active-side-down, respectively.

2.4 Results After Thermal Cycling

Among the monitored parameters the external efficiency (EE) resulted to be particularly affected by thermal cycling.
The external efficiency (shortened form of: slope efficiency of the fiber output power) is defined as: $EE = P_f / (I_F - I_{TH})$, where I_{TH} is the threshold current, I_F is the forward driving current, and P_f is the corresponding radiative power emitted from the fiber at a constant reference value (200 μA) of the current induced into the rear monitor diode.
The diagrams in Fig.1a and Fig.1b are two representative plots showing in details the external efficiency drift ΔEE, computed (in dB) as: $\Delta EE = 10 \text{ Log } (EE / EE_0)$, during the test program 1B (– 40 / +85°C) for both manufacturers. In the case of Man.A the external efficiency decreases monotonically with the number of cycles, while for the devices of Man.B there is no systematic trend. Some devices are even, paradoxically, improving. Assuming the aforementioned failure criteria ($\Delta EE \leq -0,5$ dB), the percentage of failed devices, at the end of the whole thermal-cycling program, is about 90% for Man.A and about 77% for Man.B.
The mean value and the standard deviation of the R_{th} values, measured in [°C/W], after thermal cycling are: $\mu = 37.8$, $\sigma = 7.6$ for Man.A and $\mu = 21.2$, $\sigma = 10.9$ for Man.B . Thus, after thermal cycling, the R_{th} mean values have increased about 6% for Man.A and about 74% for Man.B. Fig.2 and Fig.3 show the variation of the thermal resistance of all the laser diodes of Man.A and Man.B, respectively, before and after thermal cycling.

3. FAILURE ANALYSIS

Sixteen devices (seven of Man.A, and nine of Man.B) with the most representative failure modes have been submitted to failure analysis. The failure mode of almost all the analyzed modules is decrease of the external-efficiency. The only exception is a device (#B23) which failed because of a short circuit at the terminals of the laser diode.
The first failure-analysis step was to verify the fiber integrity. For this purpose, the internal guiding-wave structure of the pig-tailed modules has been analyzed by optical reflectrometry [7]. The reflection signature of all the analyzed modules has shown that the interfaces of the fiber-end, metal-can window, lens and laser-diode mirrors are localized at standard position, within the resolution limit of this technique. The reflected peak intensities at the different interfaces, when compared with those of a good reference device, are normal too. Therefore it may be inferred that thermal cycling caused no damage to the fiber and the failures are localized inside the internal metal can, where the emitter and the lens are hermetically encapsulated. The observed failure mechanisms are specific for each manufacturer.

3.1 Manufacturer A

The main failure mechanism of Man.A is the cracking of the lens-holder due to brittle fracture of the silicon (Fig.4). The percentage of the main failure mechanism is about 95%. For the remaining 5% of failures (devices # A22 and #A28) , the drift of the external efficiency is accompanied by an R_{th} increase due to cold bonding of the laser diode die onto the heat-sink (Fig. 5). The initial R_{th} values of these two devices was already high. The device # A22 passed from 40.8 °C/W up to 63.8 °C/W (+ 56%), while the device #A28 passed from 45.1 °C/W up to 54.3 °C/W (+ 18%).

3.2 Manufacturer B

The main failure mechanism of Man.B is the peeling of the nitride dielectric layer deposited under the metallization of the active side of the laser diode. The percentage of the main failure mechanism is about 70%. Since Man.B uses an active-side-down mounting, the peeling of the nitride dielectric layer results in a

random movement of the die, with consequent deviation of the light beam. Furthermore the degradation at the interface produces a large increase of the thermal resistance. Fig 6 shows the peeled-off dielectric layer of the device #B13. The thermal resistance of this laser diode has increased from 14.7 °C/W up to 27.9 °C/W (+ 90%). The die promptly lifts off from the heat-sink after applying a small shearing force by a pointed wooden tooth-pick.

The random increase of the external efficiency of the modules from Man.B during thermal-cycling (Fig.1b) can be explained by the originally not perfect die-alignment, which may then be corrected by random movement of the die.

There are a second and a third failure mechanisms, with 25% and 5% percentage of failure, respectively. The second failure mechanism is the disbonding of the diamond heat-sink (Fig. 7). The initial R_{th} value of this device (# B9) was already extremely high, and has increased from 36.4 °C/W up to 53.2 °C/W (+ 46%).

The third failure mechanism, observed in device #B23, is the interaction between the crystal and the die-solder (Fig. 8), which is a typical jeopardy of the active-side-down mounting. This laser diode had already a very high leakage current before thermal cycling.

4. CONCLUSIONS

The measurement of the thermal resistance of laser diodes may give useful pieces of information to determine the failure mechanisms not only of the laser diode but also of the module in which it is housed. This has been clearly shown by the failure analysis of commercially-available laser diode modules for telecom application from two manufacturers, which failed after an extensive thermal-cycling test.

All the samples have been electro-optically characterized at the start, at intermediate number of cycles, and at the end of thermal cycling. Before and after the test, the thermal resistance of each laser diode has been measured too.

Among all the parameters that have been monitored, only the thermal resistance was found to have notably changed after thermal cycling for a large number of devices.

Five different failure mechanisms have been identified: two for Man.A and three for Man.B . The identified failure mechanisms are:

1. cracking of the silicon lens-holder
2. cold bonding of the laser-diode die onto the heat-sink
3. peeling of the nitride layer of the laser diode
4. poor bonding of the diamond heat-sink onto the copper-stem
5. interaction between the solder and the laser-diode die

Three failure mechanisms (i.e. failure mechanisms n.2, 3, and 4, which affect 10% of the devices of Man.A and 70% of the devices of Man.B) are directly associated with an increase of the thermal resistance.

The R_{th}-increase caused no evident parametric drift of the laser-diode die (e.g. threshold-voltage or series-resistance increase) but directly affected the external efficiency of the module, in consequence of mechanical displacement either of the laser diode die or of the heat-sink. These two forms of displacement (which have different causes) change the alignment between the laser diode, the lens and the fiber-tip, and consequently change the external efficiency of the laser diode module, which in some case could also notably increase.

Thus, the thermal resistance proved to be an efficient parameter to monitor the reliability of laser-diode modules, since the it may put into evidence the presence of a latent failure mechanism inside those modules which apparently, relying on the external-efficiency only, are not functionally degrading or, even, paradoxically, are functionally improving.

5. ACKNOWLEDGMENTS

This work has been supported by the European Community in the framework of the RACE Project. The authors are very indebted with Reinhard Pusch, Gerhard Herrmann, Karl Hanssen, and Christian Boisrobert for conducting tasks important to this work, and with Carlos Salla for the optical-reflectrometry analysis of failed devices.

6. REFERENCES

[1] B.A. Unger and J.L. Spencer, "*Fibre in the loop reliability: the impact for fibre optic components*", Optical Engineering, June 1991, Vol. 30 No. 6, pp. 815-820

[2] M. Ciappa and P. Malberti, "*Reliability of Laser Diode Modules in Temperature-Uncontrolled Environment*", Proceedings of the 32nd International Reliability Physics Symposium, IRPS 1994, pp. 466-469

[3] P.Su and B.A.Unger, "*Temperature cycling tests of laser modules*", Semiconductor Device Reliability, Kluwer Academic Publishers, pp. 363-378

[4] P. Malberti and C.A. Salla, "*A practical way to measure the thermal resistance of laser diodes*", Proceedings of the 3rd European Symposium on Reliability of Electron Devices, Failure Physics and Analysis, ESREF 1992, pp. 483-490

[5] P. Malberti and M. Ciappa, "*The thermal resistance measurement: an efficient failure-analysis tool for laser diodes*", Proceedings of the 20th International Symposium on Testing and Failure Analysis, ISTFA 1994

[6] W.Nakwaski, "*Thermal properties of P-side-down ridge-waveguide lasers*", Optica Applicata, 1990, Vol. XX, No. 2, pp. 143-163

[7] C.Y. Boisrobert et al., "*Low coherence optical reflectrometry of laser diode waveguides*", Proceedings of SPIE Conference on Optical Technology for Signal Processing, Vol. 1474, April 1991

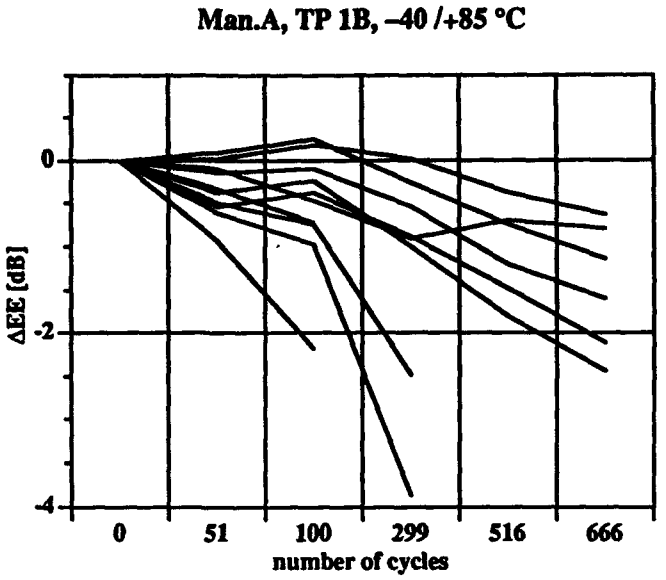


Fig.1a External efficiency drift during test program 1B (– 40/+85°C) of Man.A . The external efficiency decreases almost monotonically with the number of cycles.

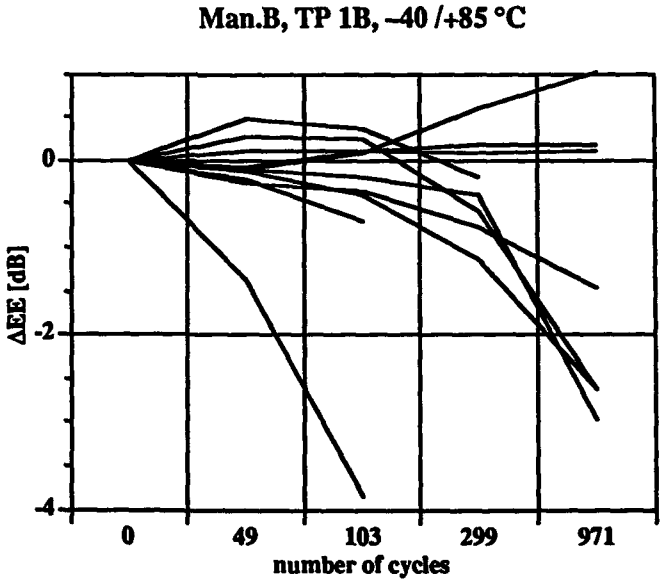


Fig.1b External efficiency drift during test program 1B (– 40/+85°C) of Man.B. The external efficiency drifts with no systematic trend. Some devices are even, paradoxically, functionally improving.

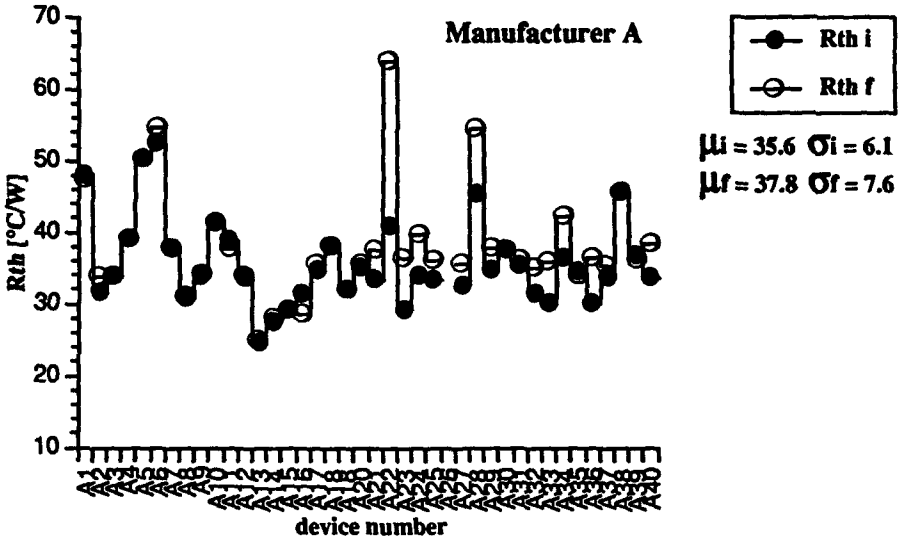


Fig.2 Thermal resistance of the laser diodes of Man.A before (i; black dots) and after (f; white dots) thermal-cycling (μ: mean value ; σ: standard deviation).

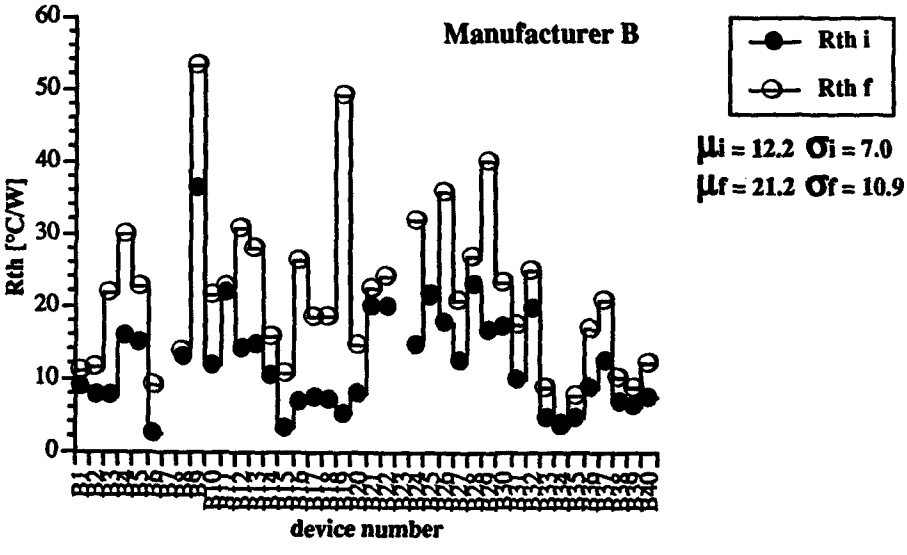


Fig.3 Thermal resistance of the laser diodes of Man.B before (i; black dots) and after (f; white dots) thermal-cycling (μ: mean value ; σ: standard deviation). After thermal cycling the mean value has increase about 74% .

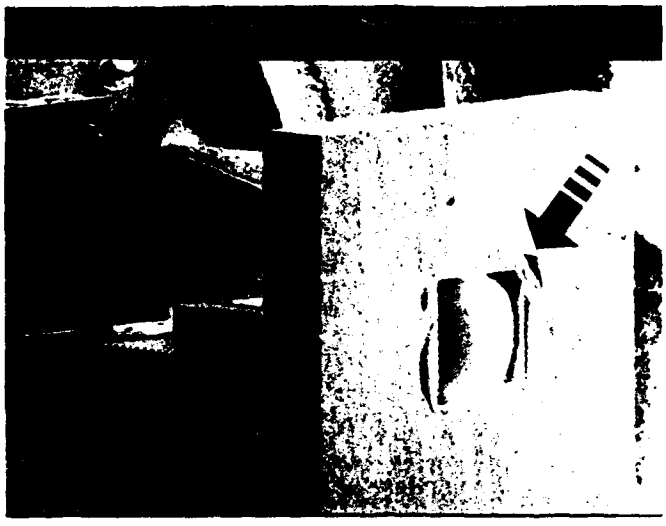


Fig.4 SEM image of the cracked lens holder. This is the main failure mechanism observed in the laser-diode modules of Man.A .

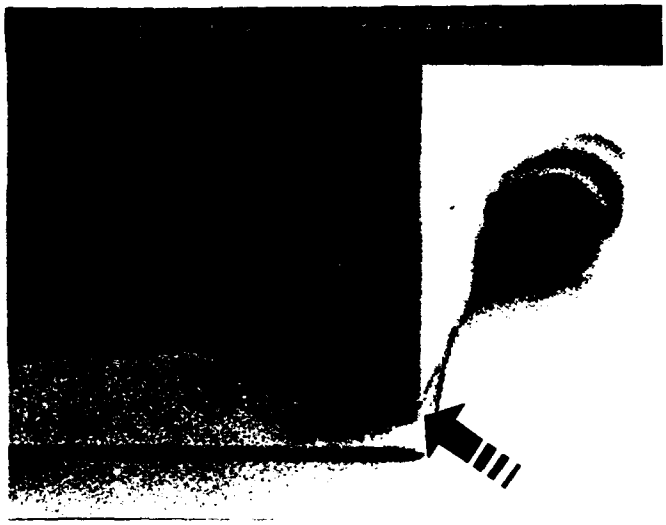


Fig.5 SEM image of the gap between the die and the bonding pad on the heat-sink, which caused a large R_{th} increase. This is the second failure mechanism observed in the laser-diode modules of Man.A. The thermal resistance of this device (# A22) increased from 40.8 °C/W up to 63.8 °C/W.

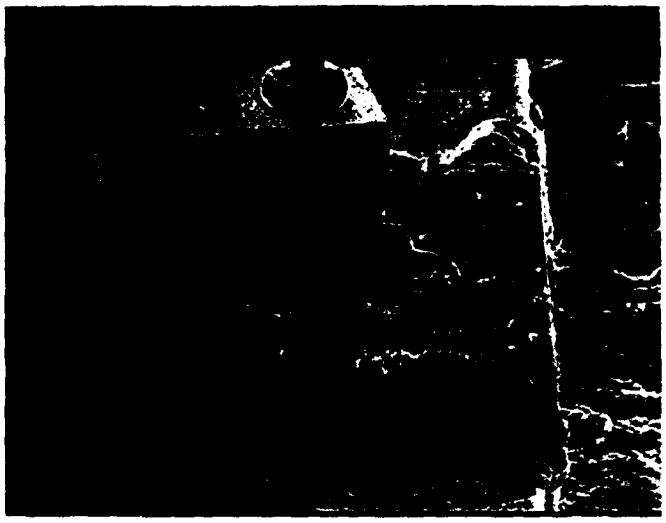


Fig.6 SEM image of the peeled-off dielectric layer of the device # B13. The thermal resistance of this laser diode has increased from 14.7 °C/W up to 27.9 °C/W. The die promptly lifts-off from the heat-sink after applying a small shearing force by a pointed wooden tooth-pick. This is the main failure mechanism of Man.B .

Fig.7 SEM image of the lifted heat-sink because of poor bonding. This is the second failure mechanism of Man. B. The initial R_{th} value of this device (#B9) was already extremely high, and has increased from 36.4 °C/W up to 53.2 °C/W (+ 46%).

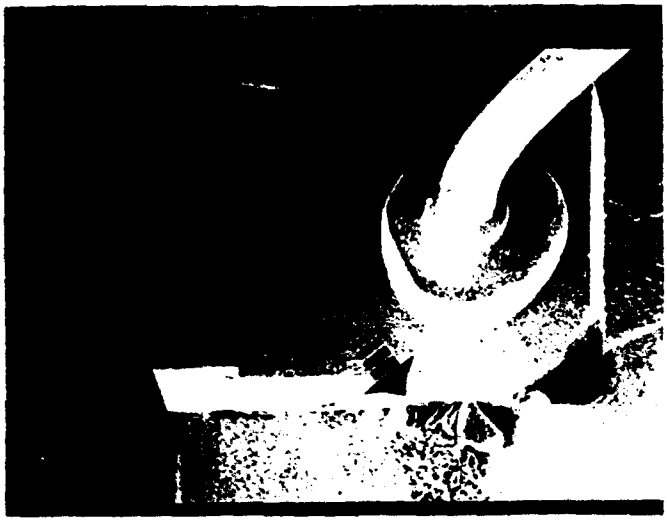


Fig.8 SEM image of the short-circuited laser diode because of interaction between solder and die. This is the third failure mechanism of Man.B. This laser diode (# B23) had already a very high leakage current before thermal cycling.

test program	1A	1B	2	3
temperature range	-40 /+70 °C	-40 /+85 °C	-10 /+60 °C	0 /+50 °C
DUTs Man A Man B	 10 * 10 *	 10 * 10 *	 10 * 10 *	 10 * 10 *
dwell time	3 hours	3 hours	3 hours	1,5 hours
# of cycles	660	660	2200	5000
test after [100 cycles]	0.5; 1; 3	0.5; 1; 3	1; 5; 10	5; 10; 20

(* *5 devices with fiber connector*)

Tab.1 Temperature-cycling stress test

BLUE SHIFT AND MIRROR DEGRADATION IN InGaAs/GaAs STRAINED QUANTUM WELL LASERS

L. Serra, G. A. Azzini, P. Montangero

CSELT SpA - Via G. Reiss Romoli 274
10148 Torino - Italy
Phone +39 11 2285471 - Fax +39 11 2285695

ABSTRACT

A high decrease in optical power and a strong blue shift have been observed in InGaAs/GaAs strained quantum well gain guided lasers failed during aging tests. The degradation of the front facet and the consequent formation of a dark spot defect close to the mirror have been evidenced with infrared and scanning optical microscopy.

1. INTRODUCTION

The high power InGaAs/GaAs strained quantum well (QW) lasers are suitable in many fields of power application, for example as pump sources for solid state lasers or fiber amplifiers. A high stability of the pumping wavelength and a high output power are the main requirements for long-term reliable devices [1]. In many cases these are critical points because the electro-optical characteristics can be easely degraded by different failure mechanisms occuring at facet or bulk level into the device st .cture.

Despite the enhanced catastrophical optical damage (COD) level (>20 MW/cm^2) reported by some authors [2] in comparison to the AlGaAs/GaAs lasers, these devices can still suffer for strong degradation mechanisms developing on the mirrors. As an example, during aging tests in constant output power mode, a gradual degradation of 0.98 μm lasers due to facet oxidation has been observed by Okayasu et al [3].

In the present work it is reported a study on high power InGaAs/GaAs strained QW gain guided lasers suffering for rapid degradation during step stress tests performed in the constant current mode. The pronounced changes in the electro-optical characteristics have been analyzed and the degradation mechanisms highlighted.

2. EXPERIMENTAL

The devices under study are prototypes of high power InGaAs/GaAs strained quantum well lasers with a gain guided confining structure and a 50 μm stripe. The output power is 300 mW at an operating current ranging from 550 to 600 mA.

The electro-optical characterization has been performed at different temperatures obtaining forward and reverse current-voltage (I-V) curves, light-current (L-I) characteristics and derivative and optical spectra under different operation currents.

After electro-optical measurements, the lasers have been operated in a costant current mode during a step stress test of about 2000 h. The test temperature has been increased from 25 °C to 50°C with steps of 5°C and the operating current changed from 650 to 850 mA with steps of 50 mA, in order to keep the output power always close to the maximum rating. The failure criteria imposed was the reduction of 50% or more of the emitted optical power.

The failed devices have been analyzed with investigative techniques based on infrared microscopy, scanning electron microscopy (SEM) and scanning optical microscopy (SOM), which offer different approaches to the problem of identification and localization of crystallographic defects into the device structure [4]. In particular, after adequate preparation, the lasers have been analyzed on the front facet and from the n- or p-side in the electroluminescence (EL), electron beam induced current (EBIC), optical beam induced current (OBIC) and photoluminescence (PL) modes.

3. RESULTS

3.1 Aging tests and characterization

The devices suffered from rapid degradation occuring at different steps of the aging test.

The light-current characteristics for strongly degraded devices, i.e. for lasers having a reduction in optical power higher than 50%, showed a strong increase in threshold current and a relevant loss of quantum efficiency, as reported in Fig. 1. A kink can be seen in the L-I curve obtained after failure.

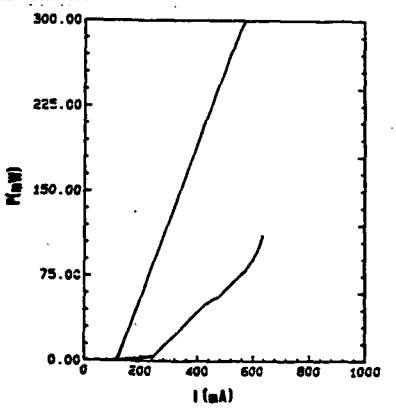


Fig.1- L-I curve before and after step stress for a strongly degraded laser.

Also the optical spectrum changed and new emission peaks at lower wavelengths appeared. Their intensities were directly connected to the degradation level, as visible in Fig. 2 for a weakly and a strongly degraded laser. Both the spectra have new peaks at shorter wavelengths, corresponding to an increase in the energy levels transition of about 60 meV. Anyway, in the first case the most intensive peaks are still at the original wavelengths, while in the second one the most intensive peaks are blue shifted.

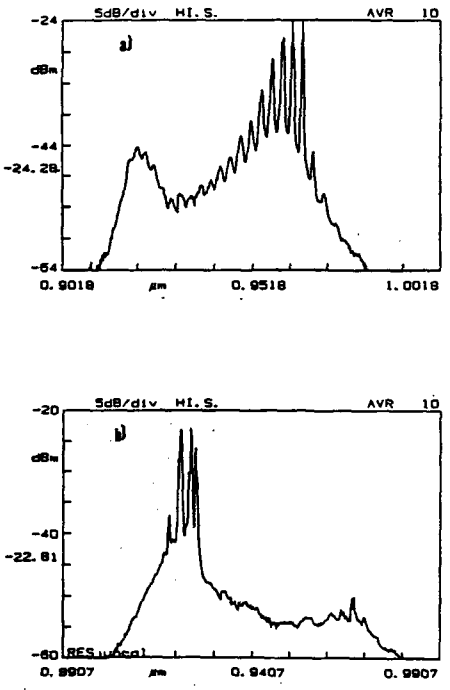


Fig.2- Optical spectrum for a) a weakly and b) a strongly degraded laser.

Optical spectra obtained on different points of the emitting area have demonstrated that this phenomena is not spatially dependent. On the contrary, it is clearly linked to the level of injected current. In fact, the kink in the L-I characteristics is directly connected to the appearance of the wavelength shift, as visible in Fig. 3.

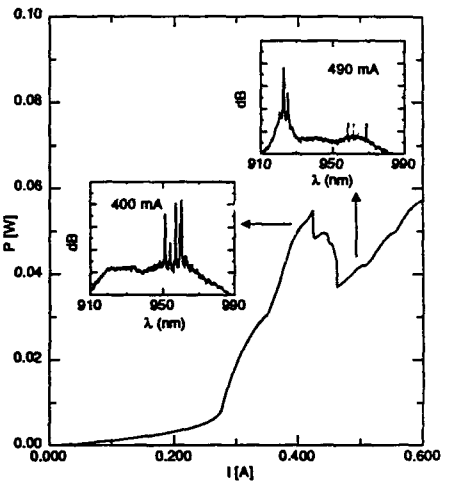


Fig.3- L-I and optical spectra before (400 mA) and after (490 mA) kink.

Furthermore, changes in power emission appeared when the lasers front facet was chemically treated. The L-I curve of a failed device after mirror treatment shows a power recovery of one order of magnitude.

3.2 Failure analysis

The EL analysis on the facet of strongly degraded devices showed the total

a)

b)

Fig.4- EL from the facet of a laser a) before and b) after strong degradation (I=500 mA).

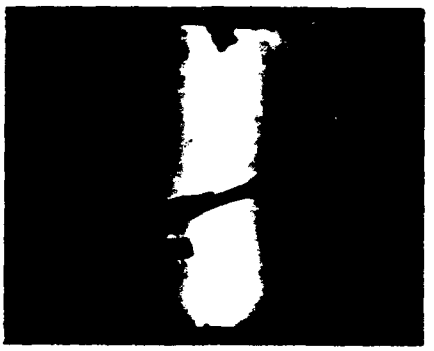


Fig.5- EL (3 mA) from the n-side of a failed laser. A DSD is visible near the front facet.

absence of light emission from the central part of the active stripe. An example is given in Fig. 4 showing the near field image of one of these lasers before and after failure.

After the opening of an optical window, the n-side analysis has been performed. It highlighted the presence of a large dark spot defect (DSD) just close to the front facet of the lasers. As an example, the electroluminescence (EL) image from the n-side reported in Fig. 5 shows one of these defects, which act as strong non-radiative recombination centers for the electron-hole couples present in the QW layer and in its confining structure.

This result has been confirmed by the OBIC analysis performed with a selective excitation of the active layer. Fig. 6 is the OBIC image of another DSD obtained using a 980 nm excitation wavelength.

Finally, some devices have also been detached from their submount, soldered p-side up on another carrier and electrically bonded in order to perform a p-side analysis with a scanning electron microscope. The obtained images give a further evidence of the presence and position of the dark spot defect. Fig. 7 is an EBIC image at 35 kV, where a large defective area is visible near the front facet.

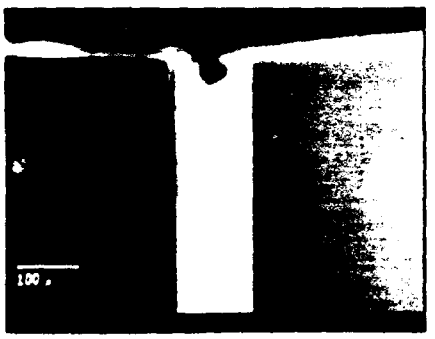


Fig.6- OBIC (980 nm) from part of the active stripe of a failed laser. A DSD is present near the front facet..

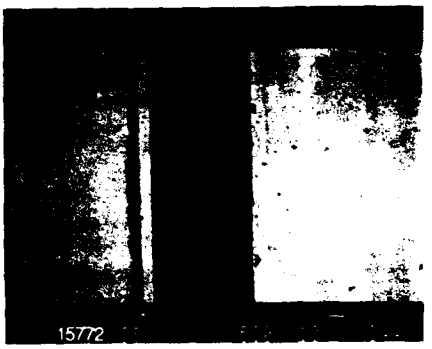


Fig.7- EBIC (35kV) from part of the active region of a failed laser. A large dark area near the front facet is visible.

4. DISCUSSION

As already discussed, the electro-optical characterization gives the indication of a strong reduction in optical power for most of the failed lasers, which is accompanied by a relevant increase in threshold current and a blue shift in wavelength emission. There has also been the evidence that the new peaks are not spatially but current dependent, occuring at a precise current level and increasing while enhancing it.

The chemical treatment of the front facet and the failure analysis reveals the presence of a mirror degradation, which spreads into the inner part of the cavity producing the dark spot defects observed. The dimensions of these defects seem directly connected to the degradation level, suggesting the idea that they can strongly suppress the lasing performances of the devices.

The mirror losses can cause an increase in threshold current and a lowering in the emission wavelength, as reported also by other authors [5]. In this particular case, the change of about 60 meV in the energy levels transition could correspond to the shift from the fundamental n=1e-hh transition to an higher energy transition. This transition is favoured by the changes in the optical cavity occured after degradation, which cause a gain saturation of the lower energy transition for a definite current value.

5. CONCLUSION

High power InGaAs/GaAs strained QW lasers have been suddenly degraded in a step stress test. The failure mode was a dramatic increase in threshold current and a blue shift in the optical spectrum. The changes in the electro-optical characteristics were caused by mirror degradation, which favoured an higher energy transition.

6. REFERENCES

[1] A. R. Goodwin, A. J. Collar, I. G. A. Davies, G. D. Henshall, P. J. Anslow, A. Rashid, SPIE' s 'Laser Diode Technology and Applications V' Conference Proceedings, Los Angeles (CA), 190 (1993)

[2] W. J. Gignac, R. R. Craig, D. F. Welch, J. G. Endriz, J. S. Major Jr., SPIE's 'Laser Diode Technology and Applications V' Conference Proceedings, Los Angeles (CA) (1993).

[3] M. Okayasu, M. Fukuda, T. Takeshita, S. Uehara and K. Kurumada, J. Appl. Phys. 69 (12), 8346 (1991)

[4] P. Montangero, G. A. Azzini, R. De Franceschi, M. Mancini and L. Serra, Material Science and Engineering B24, 175 (1994)

[5] T. Takeshita, M. Okayasu, O. Kogure and S. Uehara, Jap. J. Appl. Phys., 29 (7), 1138 (1990)

RELIABILITY RESULTS ON 980 NM HIGH POWER LASERS

F. Magistrali, A. Righetti, G. Salmini
PIRELLI CAVI S.p.A.-Divisione Italia - Milano (Italy)

W.J. Gignac
SDL Inc. , San Josè, CA (USA)

ABSTRACT

Reliability demonstration for GaAs-based pump lasers is one of the key issues for large field deployment of Erbium Doped Fiber Amplifiers (EDFAs) working at 980 nm.
To that purpose, a long term reliability experiment has been designed and performed, allowing for measurement of very low degradation rates at operating conditions.
So far, these results demonstrate the suitability of 980 nm lasers as reliable pumping sources for telecom application.

1) INTRODUCTION

In the last years the increasing request for advanced telecomunication services (ISDN, CATV, long-haul links) has boosted an impressive technological breakthrough in the field of optical transmission. New systems operating at very high bit rates (1.2 or 2.4 Gb/s) are now available as commercial units, while optical transmitters at even higher bit-rates (20-40 Gb/s) have already been demonstrated in the laboratory.
Either in the long-haul trunk connections or in the local distribution environment the start of the technology revolution has been made possible by the appearance of active-fiber devices, which have given a definitive answer to the problem of all-optical compensation of fiber losses and bit-rate transparency. Erbium-doped fiber amplifiers (EDFAs), demonstrated as practical components of telecomunication links in the second half of the eighties, have gained universal acceptance among the telecommunication system designers, outperforming any alternative approach to optical regeneration in the 1.5 micron band.
The availability of an efficient gain medium has been the basis for the rapid development of EDFA technology. In its basic form (Fig.1) the optical amplifier is made of three fundamental components: a pump source, an active fiber, and a specialized device able to multiplex on the same fiber the signal and the pump.

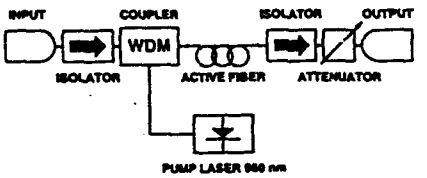


Fig. 1 : Typical structure of an Er-doped fiber amplifier.

The operation at 980 nm, the best wavelength in terms of performance, has been made possible in the recent years after a large development work on the pump-lasers; these devices feature high coupled optical power (50-80 mW) with low operating current (<200 mA) and good spectral stability.
The structure that allows to reach these characteristics is based an InGaAs Quantum Well (QW) active layer, with asymmetric facet coating, enabling high power emission from the front facet; vertical confinement is providedby AlGaAs GRIN layers.

Given the extremely high optical power density (>>10 e 6 W/cm^2) it is quite obvious that reliability is a crucial issue, mainly associated with Catastrophic Optical Damage (COD) at the facets (see for example [1]) several works have addressed this problem, which can be solved by means of appropriate facet treatment before coating [2] or by use of InGaP cladding layers [3] due to better surface recombination characteristics with respect to AlGaAs.
The influence of Dark Line Defects (DLDs), that are quite common to GaAs-based devices, seems to be greatly reduced by the presence of Indium in the active area [4] and they can anyway be avoided by optimized growth and process control [5]
In this work, we will present the results of reliability tests performed under moderate stress conditions, with the goal of assessing long term degradation behavior at operating conditions; tentative accelerating factors for power and temperature dependence will also be derived.
Short term experiments on instantaneous COD will be briefly discussed.

2) DEVICE CHARACTERISTICS

We have tested commercially available pump lasers produced during 1992, sealed in metal TO package with glass window and back facet monitor diode; a typical CW characteristic is shown in Fig.2, where. the normal operating conditions are indicated (150 mW front facet power at 180 mA driving current); the capability for single mode operation up to very high power is also shown in the figure
When operating under CW conditions in Junction-down configuration, these devices can work up to 1A (500 mW maximum power) without short term degradation; much higher power can be reached under low duty cycle operation, avoiding thermal effects (Fig.3).

3) TEST CONDITIONS

120 Lasers are being tested, under 4 different power/temperature conditions, as reported in Tab.I; tests are running in APC mode (Automatic Power Control) with individual feedback circuits that keep optical power constant by means of the bach facet monitor current.

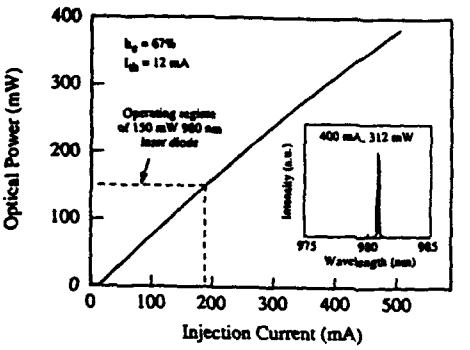


Fig. 2 : Optical power versus current and output spectrum for a typical 980 nm laser diode.

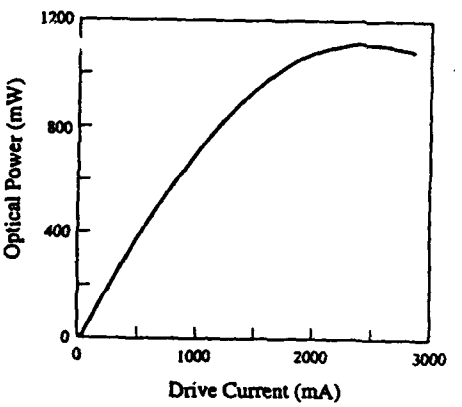


Fig. 3: Peak optical power versus current for a 980 nm laser subjected 200 to *ns* current pulses at 10000 Hz.

Laser Q.ty	Conditions (mW/°C)	Test time (Khrs)
40	50/25	10,8
20	50/50	10,2
20	100/25	10,5
40	120/70	9,8

Tab. I : Test conditions

Continuous monitor of the operating current is performed and, periodically, a complete electro-optical characterization is taken, including Power/Current/Volatge curve and spectral characteristics versus temperature.
Test conditions were intentionally chosen not to be particularly severe, as we wanted to investigate gradua, degradation that can be hidden at extreme conditions by specific failure mechanisms.
A quite conventional failure criterion of 25% increase of driving current was used, that can be arbitrarly changed, due to the good linearity observed in the degradation of the lasers under test. No device has actually reached the failure criterion, so that reliability estimation is based on linear extrapolation,
As the degradation rates can be very low, calculations are made by averaging the current for each individual laser over 50 hours in order to eliminate spurious effects from measurement resolution and repeatibility.
Besides long term tests, characterization up to very high current was performed, to assess device robustness against instantaneous COD.

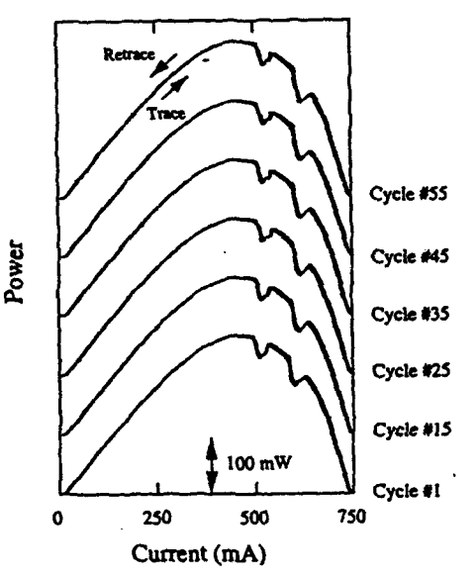


Fig. 4: CW optical power versus current for a 980 nm laser diode cycled repeatedly to 750 mA.

4) TEST RESULTS

4.1 Instantaneous COD

This test was performed on devices mounted junction up, by ramping up the current to 750 mA.
As shown in Fig.4 the devices can undergo numerous overdrive cycles without a change in operating performance. The laser reaches a maximum output power level of 300 mW at a current level of -450 mA. A further increase in current causes the output power to drop to zero at -750 mA drive level due to overheating of the chip. However, by reducing the current, the laser will be able to retrace its initial power versus current curve. Selected lasers have been cycled over 1,000 times, under these conditions.

4.2 Long term tests

So far the devices have been operating for approximately 10,000 hours with slight parametic degradations, as reported in Fig.5; the degradation behavior, when measurable, fits quite well a linear law so that linear estrapolation was used up to 25% degradation.

As previously reported, we averaged the current for each laser over 50 hours in order to smooth perturbations. Moreover, we did not consider very low dedradation rates which are below the measurement resolution, and therefore can impact the reliability estimation by introducing scatter into the data (in practice we did not consider any extrapolated failure time exceeding 10 e 6 hours).
At the end, the data were fitted to a lognormal law, yielding the results of Tab. II.

Conditions (mW/°C)	MTF (Khrs)	Standard variation
50/25	8×10^3	2
50/50	3×10^3	1,9
100/25	5.4×10^3	1,2
120/70	2.2×10^3	1,5

Tab.II:
Extrapolated lognormal parameters

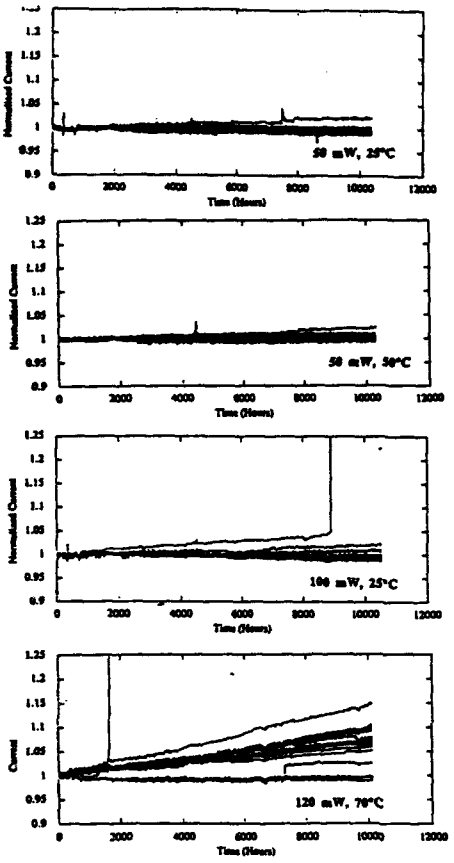


Fig. 5: Percentage driving current variation at the four test conditions.

These data can be eveluated in order to get accelerating factors for optical power and temperature; it is better to use active area, rather than heatsink, temperatures because the variation in drive currents for each to do that, it is better to use actual active area temperature because driving current for each of the 4 sets of test conditions will produce differences between the laser active areas and heatsink temperatures.
For junction down devices, we can the use a value of 35÷40°C/W junction to case thermal resistance, along with the results of Table II to extract acceleration factors consistent with the following model.

$$MTF = P^{-\alpha}\exp(Ea/KT)$$

where $\alpha = 1/2$ and $Ea = 0.35 \div 0.45$ eV
The small Arrhenius type activation energy and power law exponent give a clear indication of weak dependence of laser degradation rate on temperature and power (at least in the considered range).
The periodic complete electro-optical characterization gave no other indication, of laser degradation; in particular spectral measurements remained stable over the test.

Finally random catastrophic failures occurred; with no correlation with parametric degradation; some of these failure have been clearly inferred by external events, such as computer failure or handling; excluding these cases, the failures came from the same wafer, posing the question of better screening procedures.

5) DISCUSSION AND CONCLUSIONS

Long term degradation behaviour of 980 nm laser diodes has been evaluated by operating lifetests at moderate stress level; so far the results indicate that a degradation mechanism is present, with low linear degradation rate.
The dependence of this phenomenon; on both optical power and temperature is low, in accordance with previously reported results; nevertheless, extrapolation to normal operating conditions (P= 120÷150 mW, T_{laser} = 25°C with Peltier cooling) yields a median life well exceeding 10 e.6 hours even with a tight failure criteria of 25% increase of the driving current.
Therefore, it is possible conclude that long term degradation is not a limit for operational reliability of 980 nm pump lasers.
Regarding COD, a brief characterization was carried out to assess robustress against current ramp: devices mounted in junction up configuration showed thermally limited behaviour, without catastrophic degradation up to 300 mA and recovery of initial characteristics when ramped several times up to 750 mt.

Still, a better comprehension has to be gained about long term COD: based on the cases we have observed during our lifetests, a work is in progress trying to find out if this phenomenon can significantly be reduced by proprer screening.

REFERENCES

[1] M. Fukuda
"Reliability and degradation of semiconductor lasers and LEDs"
Artech House, 1991.

[2] A. Moser, E.E. Latta
"Arrhenius parameters for the rate process leading to catastrophic damage of AlGaAs-GaAs lasers facets".
J. of Appl. Phys 71 (10), pp 4848-4853, May 1992.

[3] H. Asonen et al.
"Aluminum-Free 980nm GaInAs/GaInAsP GaInP Pump lasers".
IEEE, J. of Quantum Elect., Vol. 30, n.2, fEB, 1994.

[4] R.G. Waters et al.
"Inhibited dark-line defect formation in strained InGaAs/AlGaAs QW lasers, IEEE .Photon Techn Lett. Vol. 2 pp. 531-533, 1990.

[5 N. Chand et al.
"Growth and fabrication of high performance 980 nm strained InGaAs Quantum Well lasers for Erbium-doped fiber amplifiers".
IEEE Journal of Quantum Elect, Vol. 30, n.2 pp. 424-440, Feb. 1994.

INFLUENCE OF PACKAGE ATMOSPHERE ON 980nm HIGH POWER CHIP RELIABILITY

Peter Whitney, David Darby, Laurie Scheffel, Dale Flanders, *Lasertron, Inc.*
Reinhard Pusch, *Alcatel SEL*

Recent advances in the development of high power 980nm QW laser diodes have allowed the production of reliable devices of this type operating at over 100mW of ex-facet optical power. Substantial data now exist which demonstrate that the sudden failure of devices due to catastrophic optical mirror damage (COMD) can be eliminated (Refs 1,2). The available data are largely based on lifetests of laser chips performed in ambient atmosphere, and very little data has been presented on the reliability of these devices in sealed hermetic modules. On the other hand, it is well known that high power semiconductor lasers can be degraded in the presence of hydrocarbons in hermetic atmospheres, which has led researchers to propose incorporation of dry oxygen in hermetic modules (Refs 3,4).

In this paper, we report experimental results on the effect of the ambient environments on the reliability of high power 980nm laser diodes. It is demonstrated that trace organic contaminants within a sealed laser module can lead to sudden failure of the laser diode in the absence of oxygen. It is confirmed that oxygen in the module atmosphere can prevent this effect and result in reliable operation.

"Standard" laser modules used in this study are sealed in dry nitrogen and typically contain organic contamination levels below 10ppm, as determined by residual gas analysis. Our lifetest studies on such modules reveal a failure rate in the first few thousand hours which is significantly higher than that predicted from "chip-on-carrier" lifetest studies conducted in air. Such module failures are typically preceded by a period in which the rear facet power increases with time, while the front facet power decreases, as shown in Figure 1. The change in the front-to-back ratio of optical power is an indication that the facet reflectivity is changing over time, and it is hypothesized that this change is caused by the accumulation of nonvolatile deposits on the facet during operation.

We have performed a series of experiments intended to support the above hypothesis. The first experiment involves the intentional contamination of the front facet of the laser diode with carbon, accomplished by exposing the laser front facet area to a 30kV electron beam in a vacuum chamber using an electron microscope. Such exposure is known to deposit trace carbon contamination due to break down of residual organics in the vacuum chamber. The device is then placed in a vacuum chamber with a controllable atmosphere where it can be operated while monitoring the rear facet power with an in-situ large area detector.

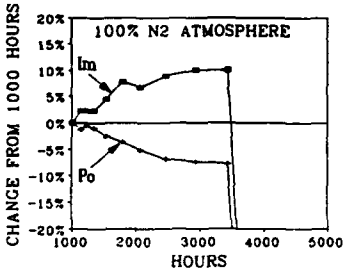


Figure 1 - Example of pump laser module lifetest failure using N_2 sealing atmosphere. The curves show the change in back facet monitor current (Im) and fiber coupled power (Po) over time.

After performing the electron beam exposure, we observe an increase in the rear facet power of the laser, similar to that discussed above in "standard" laser modules. Subsequent operation of the device in the absence of oxygen results in little or no additional change in rear facet power. However, introduction of oxygen into the ambient while the device is operating results in an immediate and rapid decrease in the rear facet power, indicating the decomposition and removal of the facet contamination. Continued operation of the device eventually results in a return to the original condition.

Figure 2 shows the relative change in the back facet power as a function of time under various oxygen levels and drive currents. The fact that the

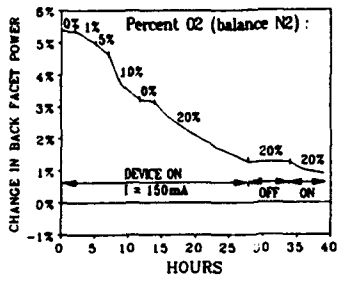


Figure 2 - Change in back facet power as a function of time for a contaminated 980nm laser under various ambient atmospheres.

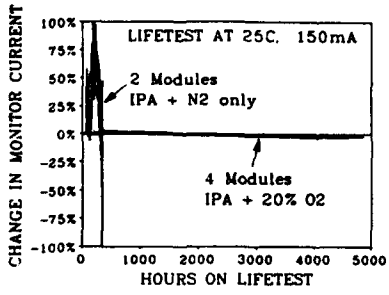


Figure 4 - Lifetest comparison for intentionally contaminated modules with and with 20% O_2 in lid sealing atmosphere.

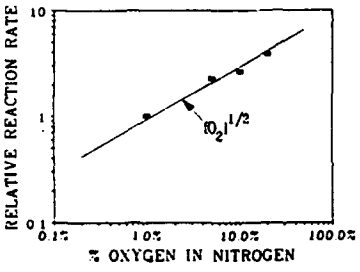


Figure 3 - Dependence of decontamination rate on oxygen concentration.

recovery process occurs only in the presence of oxygen and only if the device is actually operating is indicative of a photo- and/or thermally-induced reaction process between oxygen and the carbonaceous deposit.

Figure 3 shows the relative reaction rate for this process as a function of oxygen concentration. The decontamination rate shows approximately a square root dependance on oxygen concentration and proceeds quite rapidly even oxygen levels as low as 1%.

We have also demonstrated this effect by intentionally contaminating pump laser modules with a drop of isopropyl alcohol just before lid sealing. Contaminated modules of this type were sealed in dry nitrogen with and without 20% oxygen and were subsequently placed under operation while monitoring the back facet power.

Modules sealed without oxygen showed extremely unstable operation and eventually failed within the first few hundred hours of operation. On the other hand, modules sealed with oxygen exhibited very stable operation and showed no sign of degradation even after 5000 hours. The results, shown in Figure 4, clearly demonstrate that incorporating oxygen to the laser modules can suppress the facet contamination process even with abnormally high residual contamination levels.

In conclusion, we have directly observed the effect of organic contamination on the facets of 980nm laser diodes and have shown that such contamination can produce failures in hermetically sealed modules. We have shown that oxygen can effectively reverse the facet contamination process and that reliable operation of hermetic modules is possible by incorporating oxygen in the module atmosphere.

The authors would like to acknowledge the assistance of Gianni Pinelli and Scott Solimine in carrying out the experiments.

REFERENCES

1. Latta E., Moser A., Oosenbrug A., Gasser M. and Forster T. "Reliability of High-Power AlGaAs SQW GRINSCH Lasers,", Proc. 12th IEEE Int'l Semiconductor Laser Conference, Davos, Switzerland, Q-4, 1990.

2. Gignac W., Major J., Jr., Plano W., Welch D., Scifres D., "Highly Reliable, High Power InGaAs/AlGaAs Single-Mode Laser Diodes Suitable for Pumping Fiber Amplifiers," Proc. 18th European Conference on Optical Communications, pp. 69-72, 1992.

3. Japanese Patent No. 1723710, "Semiconductor Light Emitting Device", filed March 5, 1984, awarded September 21, 1985.

4. Jakobson P., Sharps J. and Hall D., "Requirements to Avert Packaging Induced Failure (PIF) of High Power 980nm Laser Diodes", Proc. LEOS, San Jose, CA, Nov. 1993.

HOT-ELECTRON-INDUCED EFFECTS, LIGHT EMISSION, BREAKDOWN AND RELIABILITY PROBLEMS IN GaAs MESFET'S, AlGaAs/GaAs HEMT'S AND AlGaAs/InGaAs PSEUDOMORPHIC HEMT'S

E. Zanoni, A. Neviani, G. Meneghesso, E. De Bortoli, L. Vendrame and A. Rizzato
Dipartimento di Elettronica ed Informatica — Università di Padova
via Gradenigo, 6/A — I-35131 Padova, Italy

ABSTRACT

This paper describes experimental results which demonstrate the existence of reliability problems due to hot-carriers in GaAs MESFET's, AlGaAs/GaAs HEMT's and AlGaAs/InGaAs pseudomorphic HEMT's. Both permanent and recoverable changes of I_d, g_m and V_p have been observed, due to different mechanisms such as: (a) trap creation and electron trapping in the drain access region, possibly at the passivation/surface interface; (b) trap creation and electron trapping under the gate, possibly at the donor/channel interface; (c) thermally-activated or recombination-induced detrapping of electrons from deep levels.

1. INTRODUCTION

As device dimensions are scaled down in the deep submicrometer region, hot-electron phenomena markedly influence their electrical characteristics; in particular, the design of optimized power microwave High Electron Mobility Transistors (HEMT's) requires a detailed evaluation of impact-ionization effects, in order to obtain high values of breakdown voltage, and to improve the output power. Device breakdown (often accompanied by burn-out) is only the most dramatic effect of impact ionization; increase in noise power, kinks in the I–V characteristics, and increase in gate current are other undesired effects which must be controlled in order to obtain a reliable operation of devices.

In Si MOS transistors, hot-electron effects induce reliability problems due to electron and hole trapping in the SiO_2 gate oxide and to the generation of interface traps. These degradation mechanisms are not present in GaAs-based Field Effect Transistors, which obviously do not adopt a MOS structure; unfortunately, several other failure mechanisms can be originated by hot carriers due to degradation of the GaAs surface, to the generation of deep levels under the gate and in the drain access region, to trapping/detrapping phenomena on pre-existing deep levels. The aim of this paper is to describe experimental results which demonstrate the existence of reliability problems due to hot-carriers in GaAs MESFET's, AlGaAs/GaAs HEMT's and AlGaAs/InGaAs pseudomorphic HEMT's (PM-HEMT's). According to device structure, technology and passivation the observed failure modes can consist in:

a) a permanent decrease of drain saturation current I_{ds} and transconductance g_m due to the increase of the drain parasitic resistance R_d, following trap creation and electron trapping in the drain access region;

b) a permanent decrease of I_{ds} due to a decrease of the absolute value of the pinch-off voltage V_p induced by trap creation and electron trapping under the gate;

c) a recoverable increase of I_{ds} and $|V_p|$ due to detrapping or recombination of electrons trapped under the gate.

After having reviewed electrical and optical methods for the characterization of impact-ionization phenomena in MESFET's and HEMT's, this paper will describe the results obtained during hot-electron accelerated tests of both commercially available devices and laboratory test structures.

2. SAMPLES DESCRIPTION AND CHARACTERIZATION OF HOT-ELECTRON EFFECTS

Samples used in this work were: ion-implanted GaAs MESFET's, manufactured by Alenia for this study, commercially-available S8901 and S8902 AlGaAs/GaAs Toshiba HEMT's, and MGF 4317D AlGaAs/InGaAs pseudomorphic HEMT's manufactured by Mitsubishi. Alenia MESFET's have been fabricated on semi-insulating ⟨100⟩ GaAs substrates, implanted with a $^{28}Si^+$ dose of 5×10^{12} cm^{-2} at 100 keV (channel implant) and with 1×10^{13} cm^{-2} at 40 keV (n^+ shallow implant for the ohmic contact regions). Source and drain ohmic contacts

were achieved by alloying a AuGeNi multilayer at 450 °C for 1 min., while the $0.5 \times 300\ \mu m^2$ recessed gate electrode was based on thermally evaporated aluminium. Next, SiN or SiO passivating layers were deposited on different parts of the same processed wafer. SiN was deposited by Plasma Enhanced Chemical Vapour Deposition (PECVD) at 13.56 MHz from a gas mixture of SiH_4, NH_3 and N_2; the substrate temperature was 250 °C. SiO was thermally deposited from an SiO crucible, with the GaAs substrate at room temperature. The gate was aligned to the (110) direction, in which most MESFET's are fabricated. The channel implant under the gate has a peak concentration of $2 \times 10^{12}\ cm^{-2}$. The MESFET layout was defined with gate length $L_g = 0.5\ \mu m$, gate-source spacing $L_{gs} = 0.25\ \mu m$, and gate-drain spacing $L_{gd} = 0.25\ \mu m$.

Toshiba HEMT S8901 and S8902 devices were characterized by a recessed gate with gate length $L_g = 0.3\ \mu m$ and gate width $W = 200\ \mu m$, and by gate-to-source and gate-to-drain contact spacing $L_{gs} = 0.5\ \mu m$ and $L_{gd} = 1.5\ \mu m$, respectively. The heterojunction is formed by an AlGaAs layer, $\cong$ 30 nm thick, on an undoped GaAs layer, $\cong$ 350 nm thick. An AlGaAs/GaAs (approximately 30 nm thick) superlattice buffer is used for separating the active device from the semi-insulating substrate. Devices have recessed Al/Ti gates and an n^+ GaAs cap layer, which is approximately 60 nm thick and extends from the gate edges to the ohmic contacts.

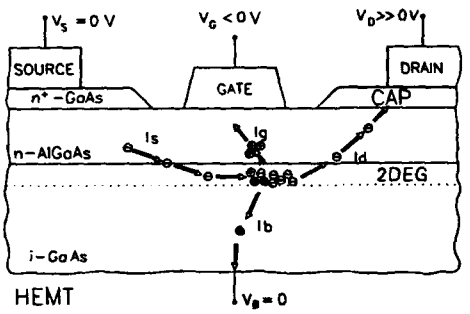


Figure 1: *Sketch of impact ionization phenomenon in HEMT's*

Mitsubishi MGF4317D pseudomorphic HEMT's adopt a 420 Å $Al_{0.2}Ga_{0.8}As$ donor layer and a 150 Å $In_{0.18}Ga_{0.82}As$ channel, $L_g = 0.25\ \mu m$ and $W = 200\ \mu m$. Fig. 1 shows the schematic cross-section of an AlGaAs/GaAs HEMT and identifies impact-ionization-induced currents I_g and I_b. When the device is operated in open-channel conditions at high drain-to-source voltage V_{ds}, electrons can reach an energy sufficient to generate hole-electron pairs due to impact-ionization; the electric field in the channel separates the generated pairs; electrons are collected at the drain, while holes are collected partly by the source, partly by the gate and partly by the substrate, as schematically shown in the figure. As a consequence, a noticeable increase in the gate current I_g, exiting from the device, is observed, see Fig. 2. I_g, due to the impact-ionization generated holes, is usually much larger than the gate-drain diode reverse current I_{gdo}, measured with the source kept floating.

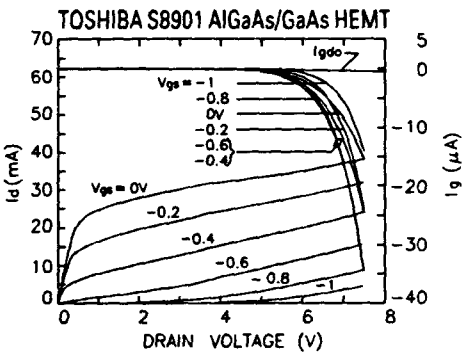


Figure 2: *I_d and I_g characteristics vs V_{ds} at fixed V_{gs}, for a typical Toshiba S8901 AlGaAs/GaAs HEMT. I_{gdo} is the gate-drain diode reverse current measured with the source kept floating.*

Figure 2 shows the drain and gate characteristics of a Toshiba HEMT at different fixed values of the gate-to-source voltage V_{gs}, with the range of drain-to-source voltage V_{ds} extending up to 6 V. Notice that the maximum drain voltage specified by the supplier for these components is $V_{ds} = 4$ V. The gate-drain diode reverse current I_{gdo} measured with the source contact floating, is also reported. The onset of the "pre-breakdown" regime is marked by the noticeable (negative) increase of the gate current I_g, when V_{ds} is moved beyond 5 V. The excess gate current comes from the collection of holes generated through impact ionization by electrons drifting in the conductive channel of the device operated as a transistor. In fact, with source floating, I_{gdo} remains negligible with respect to I_g in the whole range of drain-to-gate voltages explored. Moreover $|I_g|$ increases going from $V_{gs} = -1$ to -0.5 V, i.e. decreasing the gate-to-drain reverse voltage, in opposition to what expected if the contribution to I_g came from gate-drain diode reverse current.

The increase in I_g actually represents a limiting factor for the operation at high V_{ds}. As shown in Fig. 3, a non monotonic behaviour of I_g as a function of V_{gs} is usually observed in MESFET's and HEMT's when biased at fixed V_{ds} in impact ionization regime. Starting from pinch-off, the increase in I_g is due to

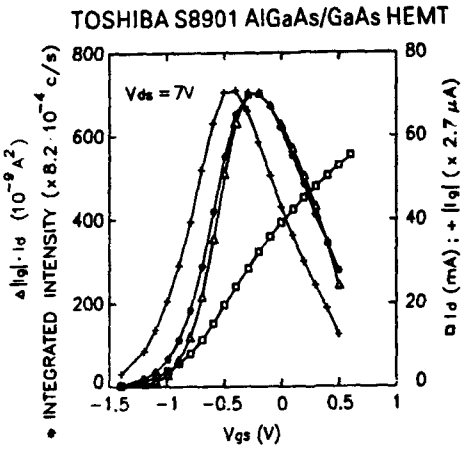


Figure 3: *I_g, $I_$ light intensity (integrated over the $1.7 \div 3.1$ eV energy range) and $|I_f|$, I_d as a function of V_{gs} with $V_{ds} = 7$ V measured in a Toshiba HEMT. Normalizing constants have been used for graphical reasons.*

the increase in I_d, i.e. in the amount of the highly energetic electrons. The subsequent decrease in I_g is due to the following reasons:

(a) the decrease of the longitudinal electric field in the gate drain region due to the increased V_{gs}, with consequent decrease in the impact-ionization rate;

(b) the possible real-space-transfer of electrons to the gate, which gives rise to a positive gate current contribution, of opposite sign with respect to the hole current.

The measurement of the gate (and substrate) current in both MESFET's and HEMT's can be used as a tool to characterize impact-ionization effects in the channel of those devices. Hot-carrier effects can be also correlated with light emission; electroluminescence spectroscopy is therefore another way of characterizing hot-electron phenomena in short-channel devices. In the following, we analyze, as a representative example, the features of the spectra emitted by AlGaAs/GaAs Toshiba HEMT's.

Optical measurements were performed using the experimental setup described in [1].

Spectra have been corrected for the optical response of the components of the apparatus, for the transmittance of the monochromator and for the quantum efficiency of the phototubes.

Figure 4 shows the photon energy distribution from 1.1 to 3.1 eV obtained in a typical Toshiba device on increasing V_{ds}. The spectra can be divided into three energy ranges. The first region spans approximately from 1.1 to 1.5 eV, and is characterized by a kink or a peak at about 1.4 eV, in correspondence of the GaAs energy gap. In the second region, from 1.5 to 2.6 eV, the experimental data fit an exponential distribution (dashed lines). The equivalent temperatures of these distributions can be evaluated from the slope of the energy spectra and lie in the $800 \div 2600$ K temperature range. Both the intensity of the emitted light and the equivalent electron temperature increase on increasing drain voltages. It should be stressed, however, that the emitted spectra most possibly result from different emission contributions, so that the observed photon distribution can not be directly related to the original electron energy distribution.

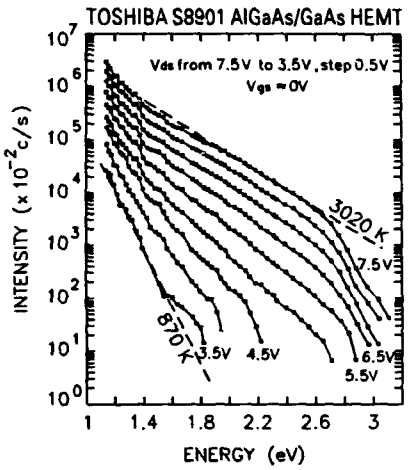


Figure 4: *Emitted light intensity as a function of energy at $T = 300$ K at various V_{ds} and $V_{gs} = 0$ V. Experimental data refer to Toshiba HEMT. Electron temperatures extrapolated from the slope of the spectra (dashed line) are shown.*

The shape of the spectra at high energy values ($h\nu > 2.6$ eV) is distorted by light absorption in the n^+-GaAs cap layer, due to the increase in the GaAs absorption coefficient taking place beyond 2.6 eV. The spectra does not show the presence of any peak in proximity of $1.7 \div 1.8$ eV, i.e. corresponding to the value of the energy gap of $Al_xGa_{1-x}As$ with x in the $0.2 \div 0.3$ range. The absence of this peak suggests that a detectable direct band-to-band recombination does not take place in the AlGaAs layer. Since a great amount of holes are generated by impact ionization and are collected at the gate electrode, thus crossing the AlGaAs layer, this would suggest that, at room temperature, AlGaAs may have a smaller radiative recombination probability than that of GaAs, or that very few electrons travel in the AlGaAs layer.

In order to discriminate the emission mechanisms of photons with energy greater than E_g, we integrated the light intensity I_{ph} in the $1.7 \div 3.2$ eV energy range and analyzed it as a function of V_{gs}. Figure 3 shows $|I_g|$, I_d, $I_g \times I_d$ and I_{ph} as a function of V_{gs} for a Toshiba HEMT biased at $V_{ds} = 7$ V. I_{ph} is proportional to the $I_g \times I_d$ product. Since I_g is proportional to the hole current, while I_d samples the electron current, this correlation, which holds for more than five orders of magnitude, Fig. 5, suggests recombination as the dominating emission mechanism of high-energy photons.

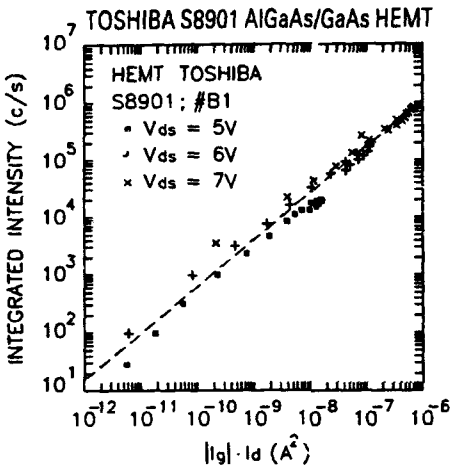


Figure 5: *Linear dependence of light intensity on $|I_g| \cdot I_d$ in a Toshiba HEMT. The device has been biased as a function of V_{gs} for the three V_{ds} values reported in the figure.*

A theoretical investigation of impact-ionization and associated light emission in GaAs MESFET's has been published by G. Zandler et al. [2]. Self-consistent Monte Carlo simulations have been used for obtaining hot electron and hole distribution functions; the absorption and emission spectra due to direct interband transitions have been calculated in the framework of relativistic non-local empirical pseudopotential theory. Simulations have been performed on the same Alenia MESFET previously described; the analysis shows that at $V_{ds} = 7$ V the below-gap radiation is controlled by conduction band to conduction band transitions, while, above gap, conduction to valence band recombination dominates. This last result is in agreement with the correlation of the integrated visible light with the product of drain and gate current above described.

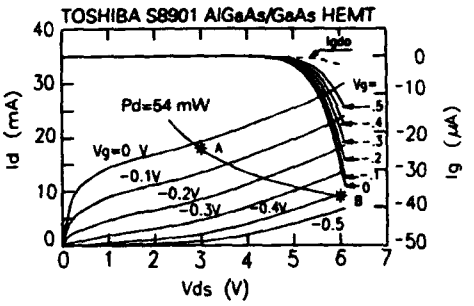


Figure 6: *I_d and I_g vs V_{ds} in a typical AlGaAs/GaAs HEMT. The curve at equal dissipated power, P_d=54 mW, has been superimposed. A. and B. show life-test bias points /A=no hot electron regime, B=hot electron regime).*

3. AGING OF MESFET'S AND HEMT'S DUE TO HOT-ELECTRON EFFECTS

3.1. Permanent Degradation of GaAs MESFET's and AlGaAs/GaAs HEMT's Due to Trap Creation in the Drain Access Region

In this section we describe surface/interface degradation effects due to hot-carriers in AlGaAs/GaAs Toshiba HEMT's and Alenia MESFET's.

Unpassivated Toshiba HEMT's were subjected to two sets of tests at room temperature.

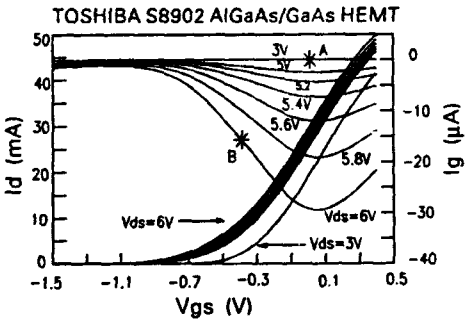


Figure 7: *I_d and I_g vs V_{gs} at various V_{ds} in a typical HEMT device. A and B show life test bias points at equal dissipated power.*

The first set was run keeping devices biased at constant dissipated power, $P_d = 54$ mW, as indicated by points A and B in Figs. 6 and 7. In this way devices have the same channel temperature during the tests, thus ruling out difference in aging due to different device self heating. The operating point A ($V_{ds} = 3$ V, $V_{gs} = -0.02$ V, $I_d = 1$ [illegible] $= -65$

nA) is characterized by negligible hot-electron effects,
impact-ionization and I_g, while in point B (V_{ds} = 6 V, V_{gs} = −0.41 V, I_d = 9 mA, I_g = −16.1 μA) significant impact-ionization occurs and I_g is high.

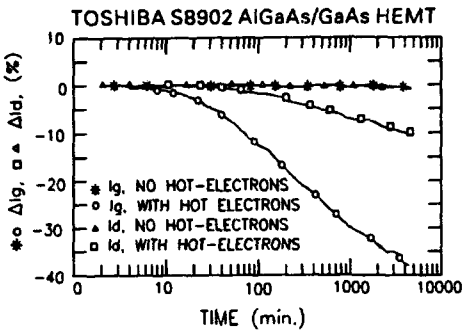


Figure 8: *Variation of I_d and I_g at the bias point during stress test at point A (no hot electrons) and B (hot electrons).*

As shown in Fig. 8, devices biased at point A showed no change in their characteristics even after a 132 hours life test, while only 79 hrs. (4740 mins.) of hot electron test in point B induced a 10% decrease in drain current, as measured at V_{ds} = 3 V and V_{gs} = 0 V. The decrease in I_d is accompanied by a decrease in the impact-ionization-induced gate current I_g.

Table I: Accelerated test at constant V_{ds} = 6.2 *V, AlGaAs/GaAs HEMT's*

V_{gs}	$I_d(mA)$	$I_g(\mu A)$	I_g/I_d $\times 10^{-3}$	Dissip. Power (mW	ΔI_{ds} at 4000 min.
-0.07	34.85	-56.30	1.61	216	-1.1
-0.28	24.27	-58.70	2.42	150	-2
-0.50	12.39	-37.85	3.05	77	-3.6
-0.76	5.80	-19.90	3.43	36	-3.9
-1.00	1.28	- 5.81	4.54	7.9	-5
-1.20	0.24	-1.47	6.14	1.5	-13

The second life test set includes tests in impact ionization regime at constant drain voltage, V_{ds} = 6.2 V, at various V_{gs} in order to obtain different I_d, I_g, I_g/I_d ratios and dissipated power, Tab. I and Fig. 9. Figure 10 reports the percentual decrease of I_d measured at V_{ds} = 3 V and V_{gs} = 0 V at various steps during this second set of stress tests. The ratio I_g/I_d kept during each test is also reported: degradation is more rapid at increasing I_g/I_d. The percentual decrease of I_d at a fixed time, 4000 mins., has been taken from Fig. 9 and superimposed to the I_g/I_d ratio in Fig. 10. A good proportionality between the I_d degradation and the corresponding I_g/I_d ratio of the tests is clearly detectable.

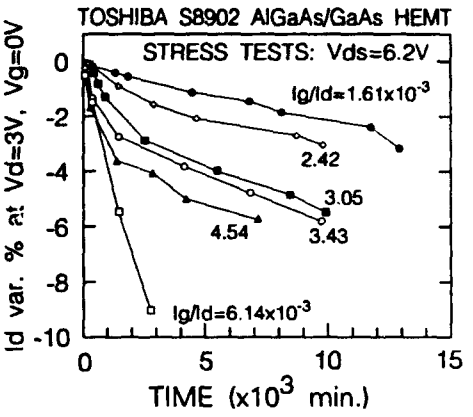


Figure 9: *Percentual decrease of I_d vs t at V_{ds} = 3 V and V_g = 0 V during tests at V_{ds} = 6.2 V with various I_g/I_d.*

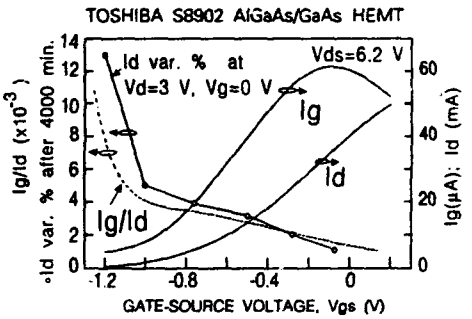


Figure 10: *I_g, I_d and the ratio I_g/I_d vs V_{gs} at V_{ds} = 6.2 V. The percentual decrease of I_d measured at V_{ds} = 3 V and V_g = 0 V after 4000 mins. of stress test at V_{ds} = 6.2 V has been superimposed.*

Figure 11 shows the drain output characteristics of a device before and after hot-electron stress (136 hrs.) at $I_g/I_d = 3.43 \cdot 10^{-3}$. As occurs for Si MOSFET's, the damage induced by hot-electrons is localized in the gate-drain region, and can be interpreted in terms of an increase in the drain parasitic resistance, Fig. 12. For example during the accelerated test performed at $I_g/I_d = 3.43 \cdot 10^{-3}$, we observed a percentual increase of R_d by about 35% after 8000 mins.. On the contrary, the parasitic source resistance remains constant. If the ouput characteristics of an aged device are measured exchanging source and drain electrodes, the damaged region is now in series with the source, causing an increase in R_s, which has a larger influence on I_{ds} than R_d; the I_{ds}

decrease observed in these 'reverse' conditions is actually larger than that observed in the 'normal' ones, as shown in Fig. 13.

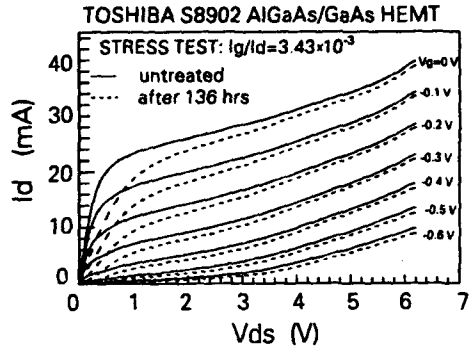


Figure 11: *Typical output characteristics before and after 136 hours of stress test at* $I_g/I_d = 3.43 \cdot 10^{-3}$.

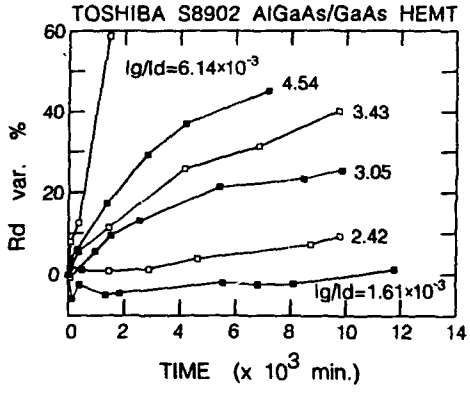


Figure 12: *Percentual increase of parasitic drain resistance measured during stress test at* $V_{ds} = 6.2$ *V The ratio* I_g/I_d *is also reported.*

In order to identify the failure mechanism induced by hot-electrons, we performed the analysis of the frequency dispersion of transconductance. Figure 14 shows normalized $g_m(f)$ measured at $V_{ds} = 100$ mV on increasing the stress time at $I_g/I_d = 3.43 \cdot 10^{-3}$. Hot-electron test induced a decrease of almost 25% in $g_m(f)$ after 136 hrs. The $g_m(f)$ decrease at a fixed time during various stress tests is larger on increasing the I_g/I_d ratio. The decrease in transconductance on increasing frequency is normally attributed to the presence of surface or interface states, in gate-drain and gate-source access regions [3]. The charge on this states can be modulated by V_{go}, provided that

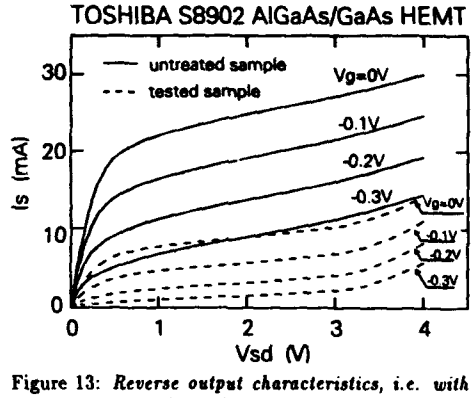


Figure 13: *Reverse output characteristics, i.e. with source contact exchanged with drain, measured in the same device of Fig. 11 before and after 136 hours of stress test.*

the period of the modulating signal is long compared to traps emission and capture times. On increasing the frequency, the charge can not be modulated anymore, and the transconductance decreases. According to this picture, hot-electron would induce the creation of surface/interface states in the gate-drain access region. This hypothesis is confirmed by the observed increase in the gate-drain breakdown voltage, observed in Fig. 15 after 162 hrs. of aging at $I_g/I_d = 3.43 \cdot 10^{-3}$; in fact, the negative charge on surface states reduces the gate-drain electric field, thus increasing BV_{gdo} [3].

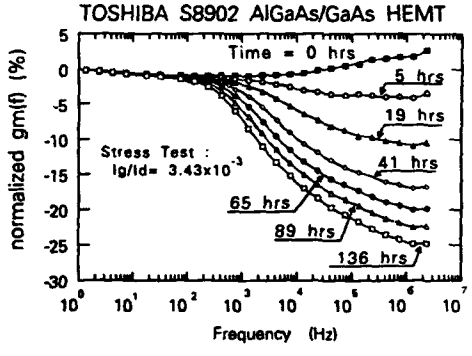


Figure 14: *Normalized transconductance frequency dispersion* $g_m(f)$ *in as received sample and at various time during stress tests at* $I_g/I_d = 3.43 \cdot 10^{-3}$.

By measuring the temperature dependence of the frequency at which the decrease of $g_m(f)$ occurs, an activation energy $E_a = 0.26 \pm 0.02$ eV is obtained, Fig. 16.

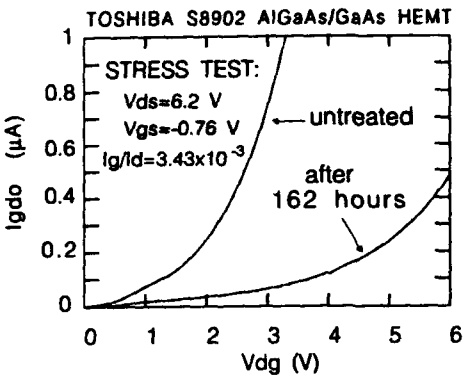


Figure 15: *Gate-drain diode reverse current measured with source floating, I_{gdo}, before and after 162 hrs. of stress test. A clear improvement of the reverse gate-drain junction characteristics can be detected.*

The behaviour of GaAs Alenia MESFET's is similar to that above reported for Toshiba S8902 HEMT's. The degradation was found to be extremely fast in unpassivated devices, while SiN passivated ones presented remarkably smaller degradations. Figure 17 compares I_d degradation as a function of time in *SiN passivated* and *SiO passivated* Alenia MESFET's. The degradation of MESFET's is also accompanied by an increase of the transconductance frequency dispersion. The transconductance decreases at high frequency, thus indicating traps in the access regions as it occurs for Toshiba HEMT's.

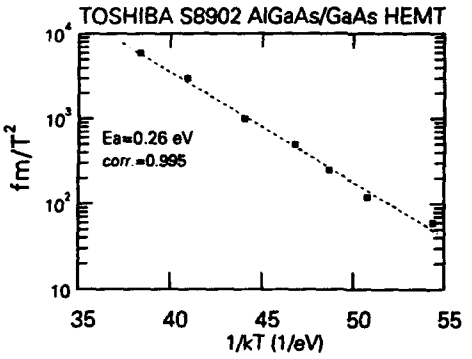


Figure 16: *Arrhenius plot of the frequency corresponding to the half-maximum decrease of transconductance measured as a function of temperature in a treated sample after 136 hrs. of test at V_{ds}.*

Figure 18 shows the behaviour of the time needed to achieve a 5% I_{ds} decrease as a function of I_g/I_d in Toshiba S8901 unpassivated HEMT's and SiO-passivated Alenia MESFET's.

The improvement of reverse gate-drain breakdown and the decrease in $g_m(f)$ suggest that the damage induced by hot electrons is related to the development of deep levels in the access region between gate and drain, as already reported [3]. All the observed degradations are permanent, i.e. they can not be annealed even after a high-temperature storage (up to 200 °C, 10 hrs.).

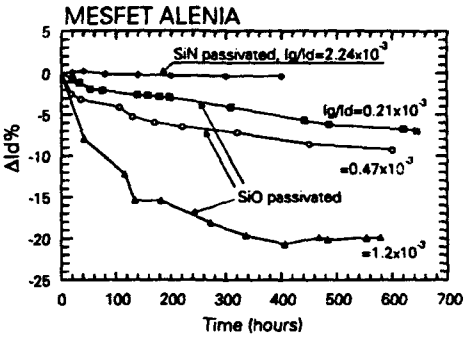


Figure 17: *Percentual decrease of I_d in SiN and SiO passivated Alenia MESFET's. The ratio I_g/I_d of stress test is reported.*

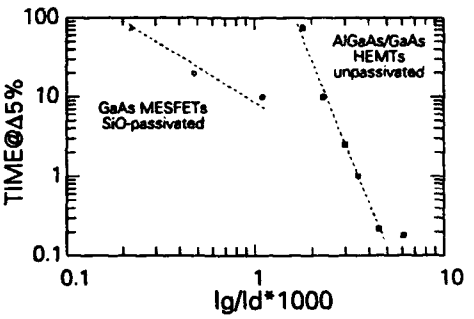


Figure 18: *Comparison of degradation rate of AlGaAs/GaAs HEMT's vs GaAs MESFET's.*

The above reported results suggest a method for accelerated testing of hot-electron effects in GaAs-based FET's, and show that hot-electron degradation can be a real concern for the reliability of high power MESFET and HEMT devices. The observed dependence on the I_g/I_d ratio can be exploited to identify acceleration factors and appropriate design rules.

To our knowledge the only available data on hot electron induced degradation of GaAs based devices

biased as transistors refer to MESFET's [4]. Devices were biased at high V_{ds} in impact ionization regime and showed a rapid decrease in I_d. Degradation was found only in SiO_2 and not in SiN passivated devices and was attributed to the injection of hot carriers into the interface between GaAs surface and passivation film, thus causing charge trapping and enlarging surface depletion layer. The degradation rate was found to be proportional to I_g.

In [5], AlGaAs/GaAs HEMT's were stressed by applying a high reverse gate voltage with respect to drain/source, thus allowing a remarkable breakdown current (1 mA/mm) to flow through the gate. Breakdown "walkout" [5] and permanent device degradation were found only in unpassivated devices and not in SiN passivated one. The proposed explanation was the oxidation of the AlGaAs surface near the gate edges.

In Toshiba and Alenia devices the degradation rate was found to be proportional to I_g/I_d, i.e. to the maximum longitudinal electric field and/or the electron energy. In particular at V_{ds} = const. a higher degradation is found for more negative V_{gs} values. This is different from what reported in [4] , where degradation was found to be proportional to I_g and to decrease at V_{ds} = const. for more negative V_{gs} values.

The improvement of reverse gate-drain breakdown and the decrease of $g_m(f)$ suggest that the damage induced by hot electrons is due to the formation of deep levels in the access region between gate and drain [5]. In our devices the activation energy of these deep levels was found to be $E_a = 0.26 \pm 0.02$ eV. We think that these traps can be located at the interface between passivation film and n^+ GaAs cap layer or between AlGaAs and GaAs channel.

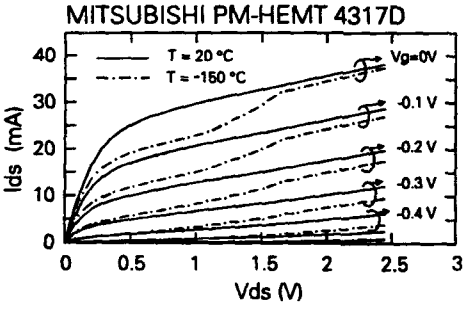


Figure 19: *Drain current characteristics at room and low temperature (-150 o C);a kink in I_ds characteristics is observed at low temperature.*

3.2. Instabilities in Pseudomorphic AlGaAs/InGaAs PM-HEMT's Due to Thermally-Activated or Impact-Ionization-Induced Electron Detrapping

In pseudomorphic AlGaAs/InGaAs HEMT's, hot-electron and impact-ionization effects can be enhanced due to the low electron effective mass and small energy-gap of InGaAs [6]. Previous studies have demonstrated that these devices may be also affected by deep levels in the vicinity of the AlGaAs/InGaAs interface. P. Audren et al. [7] have observed electron emission and capture from deep centers in pseudomorphic HEMT's similar to devices adopted in this study. The activation energy of the electron capture process is $E_a = 0.55 \div 0.59$ eV, while the emission process has $E_a = 0.74 \div 0.79$ eV, possibly corresponding to EL2 in GaAs; after aging at $V_{ds} = 2$ V, $I_{ds} = 10$ mA at $T_{ch} = 180$ °C, however, a new deep center is observed, having an emission activation energy $E_a = 0.56$ eV, wich was attributed to the n-AlGaAs strained layer.

R. Plana et al. [8] have studied noise in AlGaAs/InGaAs pseudomorphic HEMT's from 10 Hz to 18 GHz. They attributed the origin of noise with a corner frequency in the 3 kHz range observed in Mitsubishi MGF 4313 PM-HEMT's to trapping and detrapping on DX centers of the AlGaAs layer, occurring mostly at the AlGaAs/InGaAs interface.

Y. Hori et al. [9] have observed a collapse of the output I–V characteristics at low temperatures in both $Al_{0.3}Ga_{0.7}As/GaAs$ HEMT's and $Al_{0.25}Ga_{0.75}As/In_{0.15}Ga_{0.85}As$ pseudomorphic HEMT's. The amount of I_d degradation decreases at high V_{ds}. In fact, in their devices, applying a stress drain voltage of more than 1.6 V resulted in a recovery from the I–V collapse. They explained the I_d collapse as due to electron trapping on DX centers; the recovery by drain stress was associated with the ionization of DX centers by capture of holes generated by impact-ionization.

Pseudomorphic HEMT's MGF4317D manufactured by Mitsubishi are also affected by trapping effects. Fig. 19 shows the comparison between device output characteristics at 20 °C and at −150 °C. A kink in I_d is observed at low temperature, testimonying the presence of trapping effects. Electron trapping stress tests have been performed at −180 °C by biasing the devices at various V_{ds} values. Fig. 20 shows the output characteristic for $V_{gs} = 0$ V for the as-received sample at room temperature and at −180 °C, and after an electron trapping test at $V_{ds} = +0.5$ V, $V_{gs} = +0.6$ V for 1000 seconds. A noticeable decrease in the absolute value of the pinch-off voltage takes place, as shown in the $g_m(V_{gs})$ characteristics, Fig. 21. The collapse factor, defined as the ratio of g_m measured after electron trapping tests at $V_{ds} = 90$mV, $V_{gs} = 0$V to initial g_m is showed in Fig. 22 as a function of the drain

voltage V_{ds} applied during the tests. Our results confirm those obtained by Y. Hori et al. [9]; in fact, the amount of degradation first increases, then decreases on increasing V_{ds}. The reduced degradation observed at high V_{ds} is possibly due to holes generated by impact-ionization, which recombine with trapped electrons, thus reducing the net trapped negative charge and partially recovering $|V_p|$.

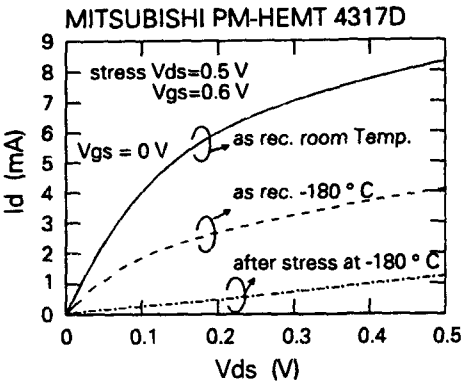


Figure 20: *I_d characteristics vs V_{ds} for V_{gs} = 0 V for the as-received sample at room temperature and at $-180°C$ and after a stress test at V_{ds} = +0.5 V, V_{gs} = 0.6 V for 1000 seconds.*

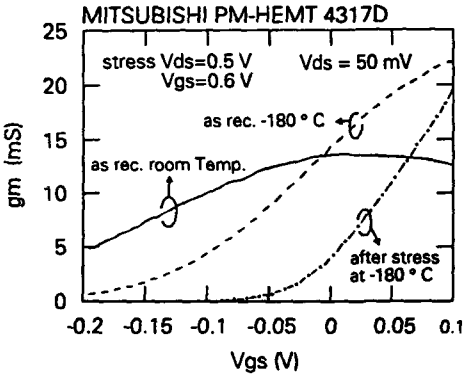


Figure 21: *$g_m(V_{gs})$ characteristics for V_{ds} = +90 mV for the as-received sample at room temperature and at $-180°C$ and after a stress test at V_{ds} = +0.5 V, V_{gs} = 0.6 V for 1000 seconds.*

The drain output characteristics of a Mitsubishi PM-HEMT taken at room temperature are shown in Fig. 23. Beyond V_{ds} = 4 V remarkable impact-ionization effects take place, with increase in (negative) I_g.

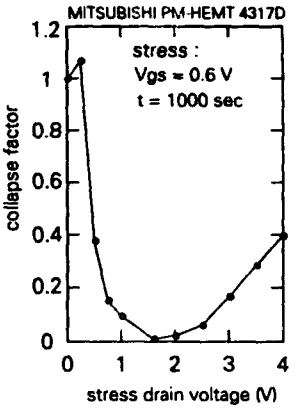


Figure 22: *Dependance of collapse factor of transconductance on stress drain voltage after 1000 seconds of stress at fixed V_{gs}.*

The presence of traps in these devices induces instabilities also at room temperature. To understand the related degradation mechanisms, we performed two sets of tests: storage tests at high temperature without bias and life tests at room temperature with the device biased in hot electron and impact-ionization regime.

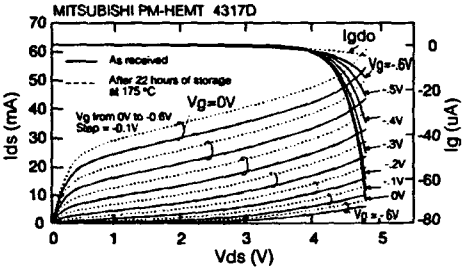


Figure 23: *Continuos lines : I_d and I_g is an as received device; I_{gdo} is the gate-drain reverse current, measured with source floating. Dotted lines : I_d after 22 hrs. of storage at $T = 175°C$.*

After 22 hours of storage at T = 175 °C without applied bias, a remarkable increase in I_{ds} is observed, as shown in Fig. 23. An even larger increase in I_{ds} takes place after 15 hours of hot-electron stress at V_{ds} = 4.7 V, V_{gs} = −0.1 V, Fig. 24. In both cases, the increase in I_d is due to an increase in the absolute value of the pinch-off voltage $|V_p|$, as shown in Fig. 25 for the hot-electron test; this increase is accompanied by a decrease in $|I_g|$ and I_{gdo} (not shown). The I_{ds} increase is therefore accelerated both by the pres-

ence of hot electrons (or hot holes) and by high temperatures. Fig. 26 shows the amount of Ids increase for different tests performed at $V_{ds} = 4.7$ V for 15 hours, as a function of $|I_g|$ which is proportional to the amount of the impact-ionization-generated holes. I_g was changed by varying V_{gs}.

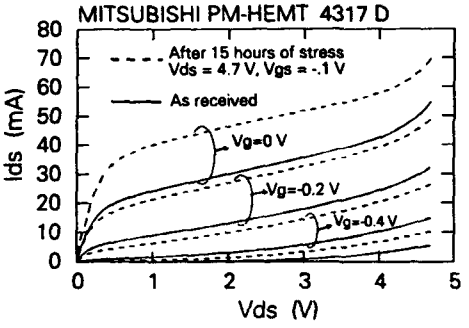


Figure 24: *Continuos lines : I_d in an as received device; Dashed lines : I_d after 15 hrs. of stress at at $V_{ds} = +4.7$ V, $V_{gs} = -0.1$ V.*

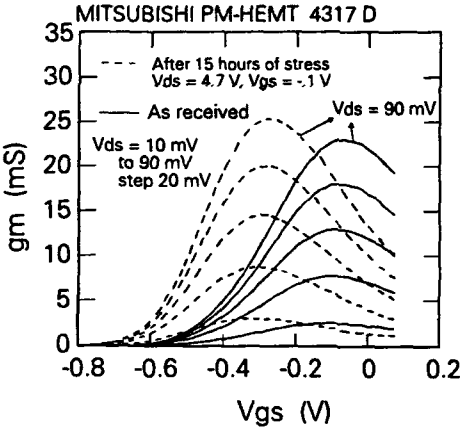


Figure 25: *Continuos lines : $g_m(V_{gs})$ in an as received device; Dashed lines : $g_m(V_{gs})$ after 15 hrs. of stress at at $V_{ds} = +4.7$ V, $V_{gs} = -0.1$ V.*

ΔI_{ds} increases with I_g, then saturates, and eventually increases again when device self-heating due to increased power dissipation contributes to device degradation. The role of impact-ionized holes in accelerating the I_{ds} increase is demonstrated by tests performed at the same dissipated power, but with different I_g; enhanced device degradation was found at high I_g values, Fig. 27 and Table II. An increase in I_{ds} is also observed when holes are generated by biasing the devices in gate-drain breakdown conditions with the source open, see Table II.

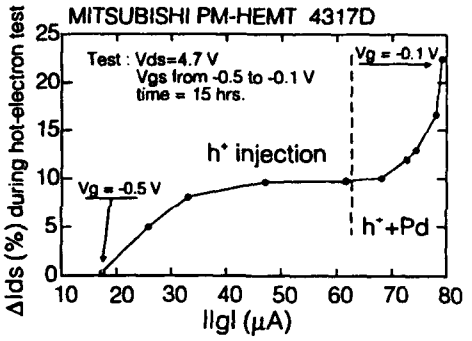


Figure 26: *ΔI_{ds} as a function of impact-ionization hole current $|I_{ds}|$, for tests at $V_{ds} = 4.7$ V, t=15 hrs.*

Table II: Increase in saturation current I_{ds} observed during accelerated test at $P_d = 98mW$ and in gate-drain breakdown conditions with source floating.

Test point	A	B	C	G-D Break.
Dissip. Pow.	*98 mW*	*98 mW*	*98 mW*	*0.38mW*
I_g during test	$-43.1\mu A$	-77 nA	756 nA	$-48\mu A$
ΔI_{ds} during test	9.4%	3.18%	0.2%	-
ΔI_{ds0}; $V_{ds} = 1.4$ V $V_{gs} = 0$ V	+10.9 %	+4.5 %	+3.2 %	+6.9 %
ΔI_{gdo}; $V_{gd} = -5.3$ V $I_s = 0$ mA	-53.6 %	-15 %	-11 %	-62 %

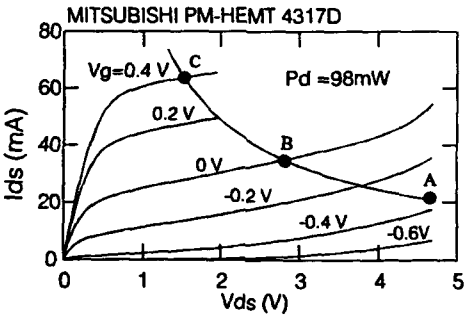


Figure 27: *I_d characteristics superimposed to the load line that identifies $P_d = 98$ mW; A, B and C are various test point.*

For both storage and hot-electron tests the degradation is not permanent, see Fig. 28; in fact, the I_d increase is recovered after a room temperature or low temperature storage. It should be noticed that at room temperature the recovery time is usually much longer than that required for causing the I_d increase either during hot-electron or storage tests.

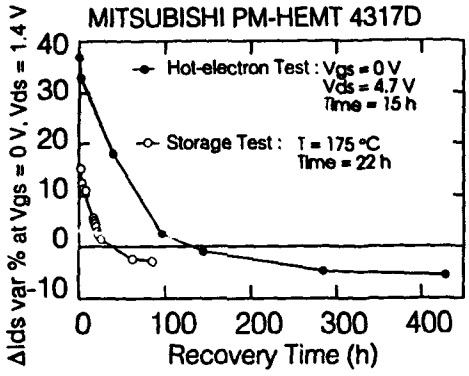


Figure 28: *Typical I_{ds}-recovery characteristics for storage and hot-electron tests at 300 K.*

A trapping/detrapping mechanism could be the possible cause of all parasitic effects observed in these pseudomorphic devices. Noise [8] and pulsed [7] measurements have demonstrated the presence of deep levels in these PM-HEMT's, also confirmed by the kinks and collapse of the I–V characteristics at low temperatures previously described.

The non-permanent nature of the I_d increase after room temperature storage or hot-electron tests suggests that also these phenomena could be related to trapping effects. We therefore explain the observed instabilities as follows. The negative charge stored due to electron trapping even at room temperature can be removed either by thermal activation of trapped electrons or by recombination with holes generated by impact-ionization. This causes an increase in $|V_p|$, and therefore an increase in I_d. When the devices are stored at room (or lower) temperature, electrons slowly fill the ionized traps and I_{ds} and V_p recover to their original value. The existence of a detrapping mechanisms due to recombination with holes generated by impact-ionization also explains why the amount of the decrease of I_{ds} during electron trapping at low temperatures decreases on increasing V_{ds} beyond 2 V, see Fig. 22. Recombination of holes with electrons trapped at DX centers has been observed by L. Dobaczewski et al. [10] in $Al_{0.35}Ga_{0.65}As/GaAs$ structures.

In conclusion, we have shown that the presence of traps can induce long-term instability effects in AlGaAs/InGaAs PMHEMT's; the described failure mechanism is accelerated by high temperature and impact-ionization effects, and therefore requires particular consideration during the design and application of power devices, which typically operate at high V_{ds}, dissipated power and T_{ch}.

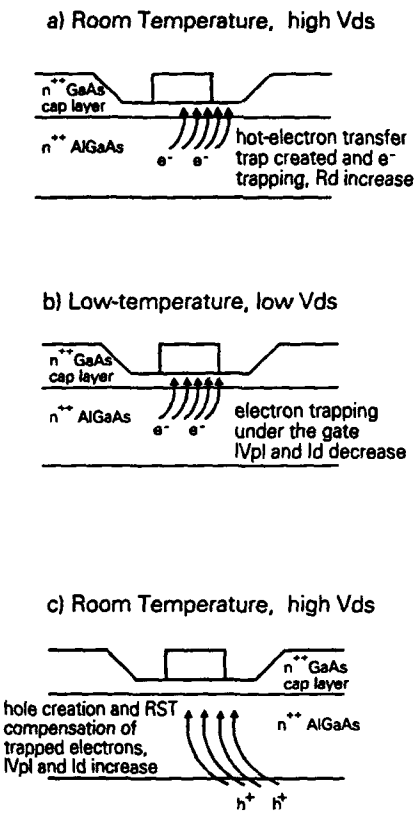


Figure 29: *Schematic cross section of a generic HEMT device, summarizing observed failure mechanism a) at room temperature in Toshiba HEMT's and Alenia MESFET's; b) and c) in Mitsubishi PM-HEMT's at low temperature and room temperature respectively.*

4. CONCLUSIONS

Hot-electron-induced degradation of commercially available GaAs-based FET's has been studied by means of high V_{ds} accelerated testing and by corre-

lation with impact-ionization effects, as monitored by gate current measurements.

Observed failure mechanisms are schematically shown in Fig. 29. In both MESFET's and AlGaAs/GaAs HEMT's we have observed increase in R_d due to trap creation and electron trapping in the drain access region (see fig. 29a). This degradation mechanism is permanent and markedly depends on the passivation adopted, SiN-passivated device being more reliable than unpassivated or SiO-passivated. Pseudomorphic AlGaAs/InGaAs HEMT's suffer of I_{ds} instabilities due to hot electron trapping and detrapping on deep levels under the gate causing threshold voltage shifts (fig. 29b and fig. 29c). This degradation is completely recoverable and obviously depends on temperature and on the presence of impact-ionization-generated holes. Permanent theresold shifts due to creation of traps under the gate following hot-electron testing have also been observed in PM-HEMT's [11].

All these failure mechanisms may limit the reliability of power devices operating at high V_{ds}. While $g_m(f)$, DLTS [12] pulse and noise measurements have proved to be excellent diagnostic techniques for the analysis of these effects, more experiments on well characterized devices and materials are needed in order to identify the physical origin of the described degradations and to correlate them with fabrication processes.

ACKNOWLEDGEMENTS

The authors would like to thank C. Canali and C. Tedesco (now with Hewlett-Packard) for useful discussions and R. Chieu, C. Morico and M. Cenedese for assistance in electrical measurements.

We are indebted to A. Cetronio and C. Lanzieri (Alenia) for providing MESFET samples and for useful suggestions.

Work partially supported by the ESPRIT project MANPOWER.

REFERENCES

[1] E. Zanoni, M. Manfredi, S. Bigliardi, A. Paccagnella, P. Pisoni, C. Tedesco and C. Canali, "Impact ionization and light emission in AlGaAs/GaAs HEMT's", *IEEE Trans. on Electron devices*, vol. ED-39, n. 8, pp. 1849-1857, 1992.

[2] G. Zandler, A Di Carlo, P. Vogl and P. Lugli, "Monte Carlo simulation of minority carrier transport and light emission phenomena in GaAs devices", in J. F. Ryan and A. C. Maciel, editors, *Semiconductor Science and Technology*, volume 9, pp. 666-670. Institute of Physics, May 1994.

[3] P.H. Ladbrooke and S.R. Blight, "Low-Field Low-Frequency Dispersion of Transconductance in GaAs MESFET's with Inplication for Other Rate-Dependent Anomalies", *IEEE Trans. on Electron Devices*, vol. ED-35, n. 3, pp. 257-267, March 1988.

[4] A. Watanabe, K. Fuijmoto, M. Oda, T. Tadayoshi and A. Tamura, "Rapid Degradation of WSi Self-Aligned Gate GaAs MESFET by Hot Carrier Effect", *Proc. of IRPS*, pp. 127-130, 1992.

[5] P.C. Chao, M. Shur, M.Y. Kao and B.R. Lee, "Breakdown Walkout in AlGaAs/GaAs HEMT's", *IEEE Trans. on Electron Devices*, vol. ED-39, n. 3, pp. 738-740, March 1992.

[6] C. Tedesco, E. Zanoni, C. Canali, S. Bigliardi, M. Manfredi, D.C. Streit and W.T. Anderson, "Impact Ionization and Light Emission in High-Power Pseudomorphics AlGaAs/InGaAs HEMT's", *IEEE Trans. on Electron Devices*, vol. ED-40, n. 7, pp. 1211-1214, 1993.

[7] P. Audren, J.F. Dumas, S. Motet, C. Vuchener, J. Paugram and M.P. Favennec, "Stability of Interfaces in Pseudomorphic InGaAs HEMTs", *Quality and Reliability Engineering International*, vol. 8, pp. 279-282, 1992.

[8] R. Plana, L. Escotte, O. Llopis, H. Amine, T. Parra, M. Gayral and J. Graffeuil, "Noise in AlGaAs/InGaAs/GaAs Pseudomorphic HEMT's from 10 Hz to 18 GHz", *IEEE Trans. on Electron Devices*, , n. 5, pp. 852-858, May 1993.

[9] Y. Hori, M. Kuzuhara, N. Samoto and T. Itoh, "Bias-dependent Collapse and its Recovery Phenomenon in AlGaAs/GaAs 2 DEGFET's at Low Temperature", *IEEE Trans. on Electron Devices*, vol. 39, n. 12, pp. 2720-2725, 1992.

[10] L. Dobacezewski, P. Kaczor, Z.R. Zytkiewicz, M. Missous, F. Saleemi, P. Dawson and A.R. Peaker, "Hole Capture at the DX(Si) and DX(Te) defects in $Al_xGa_{1-x}As$", *J. Appl. Phys.*, vol. 72, n. 7, pp. 3198-3200, 1992.

[11] E. Zanoni *et al.*, "Hot-electron-induced effects in pseudomorphic HEMT's", unpublished.

[12] E. Muñoz, E. Calleja, F. Garcia, I. Izpura and J.L. Sancez Rojas, "Influence of DX centers on HEMT performance", *Alta Frequenza - Rivista di elettronica*, vol. 5, n. 2, pp. 3/51 10/58, Marzo-Aprile 1993.

QUALITY EVALUATION OF AlGaAs/GaAs AND AlGaAs/InGaAs/GaAs HEMTs BY LF EXCESS NOISE ANALYSIS

N. Saysset, C. Maneux, N. Labat, A. Touboul, Y. Danto
Laboratoire IXL, URA 846-CNRS - University Bordeaux I - 351 Cours de la Libération,
F 33405 TALENCE - Tel : 33 56 84 65 49 - Fax : 33 56 37 15 45
J.M. Dumas
ENSIL - 83 rue d'Isle - F 87000 LIMOGES - Tel : 33 55 43 36 70 - Fax : 33 55 43 36 80

Abstract : The quality of AlGaAs/GaAs/buffer layer on GaAs HEMTs and of AlGaAs/InGaAs/GaAs HEMTs is compared on the basis of technological parameters influence : Al molefraction in the n- AlGaAs layer, type of the buffer layer (p- or n- doped GaAs or AlGaAs), In molefraction in pseudomorphic structures.
From the LF drain noise behaviour versus gate and drain biases and temperature (90K to 300K), pseudomorphic devices are found to present lower drain current noise and less G-R contributions when compared to conventional HEMTs. This discrepancy might result from a lower deep levels concentration.

1. INTRODUCTION

High Electron Mobility Transistors exhibit better noise figure performances compared with that of GaAs MESFETs. The interest of the analysis of LF noise is found in many applications of the HEMTs as oscillators and mixers where the non linearities of the device cause intermixing of the noise spectral components and generates spurious phase noise.

The parasitic effects in conventional HEMTs are related to deep levels which affect the functional operation of both discrete devices and integrated circuits. These anomalies mainly appear as frequency and temperature dependent current or voltage transients (ref. 1). The location of the contributed traps depends on the operating mode of the devices. They are related to the manufacturing processes and to the electrical and physical properties of each layer and of their interfaces. The influence of these levels can be characterised by LF noise analysis through the 1/f noise level and the possible G-R noise contributions appearing in the spectra.

The dominant trap level in conventional AlGaAs/GaAs HEMTs has been identified as the DX center which lies from 0.5 to 0.4 eV below the conduction band (refs. 2,3). It has been demonstrated that a low Al molefraction in n-AlGaAs must be achieved to minimize the DX traps related effects. A low molefraction of Al also reduces the heterojunction conduction band discontinuity and then increases the probability of electron injection across the potential barrier. This mechanism is a possible cause of the LF noise increase due to the enhancement of electron capture or emission mechanisms.

The quantum well pseudomorphic AlGaAs/InGaAs structure maintains a significant potential barrier while reducing the Al molefraction in the n-AlGaAs layer. In the AlGaAs/GaAs system, an Al molefraction greater than 0.22 is required to provide a significant conduction band offset (ΔEc) while in the AlGaAs/InGaAs system, an Al molefraction of only 0.15 is needed to produce an efficient confinement of the free carriers. However, due to the lattice mismatch ($\varepsilon = 1.3\%$) of this system, the InGaAs layer critical thickness must not be exceeded to preserve the transport properties advantages of the pseudomorphic strained heterostructures (ref. 4).

Thus, the strained - layer AlGaAs/InGaAs HEMT derives its superior performances from both the higher mobility value in the InGaAs than in the GaAs material and from the reduction of the DX related effects while maintaining a high electron gas density. According to these concepts, pseudomorphic HEMTs should provide less G-R noise components than conventional devices.

2. DESCRIPTION OF THE DEVICES

The devices under test are based on the AlGaAs/GaAs-buffer layer and AlGaAs/InGaAs/GaAs modulation doped heterostructures respectively for the conventional HEMTs (S-HEMTs) and for the pseudomorphic HEMTs (PM-HEMTs). Three types of conventional devices (A, B, C) are distinguished according to the nature of the buffer layer (table I).
An AlGaAs buffer is used to create an energy barrier in the conduction band and to reduce electron injection into the buffer, as it has been proposed by Morkoc (ref. 5) The epitaxial growth process is a Vapor Phase Epitaxy using metallorganic sources (MOCVD). The Al molecular fraction of the n-doped AlGaAs layer deposited onto a non-doped AlGaAs spacer is 0.28. The InGaAs layer of the PM-HEMTs (D type) has been grown by Molecular Beam Epitaxy (MBE). The molefraction of In was minimized to 0.15 to reduce lattice mismatch at the heterointerface and the Al mole fraction is 0.22.

Device	n+GaAs cap layer thickness (Å)	AlGaAs layer thickness (Å)	InGaAs layer thickness (Å)	GaAs buffer dopant type	AlGaAs buffer thickness (Å)
A	300	400	absent	p-	1000
B	300	500	absent	p-	absent
C	500	480	absent	n-	absent
D	300	450	150	p-	absent

Table I : Technological characteristics of devices under test

The double barrier quantum well structure (device D) presents two conduction band discontinuities : $\Delta Ec1 = 0.3$ eV and $\Delta Ec2 = 0.12$ eV while the potential barrier energy of the single heterostructure is 0.22 eV in the conventional FET as represented in figure 1.

The ohmic contacts are deposited onto a n+doped GaAs cap layer ($1x10^{18}$ cm$^{-3} \leq ND \leq 2x10^{18}$ cm^{-3}). Identical masks were used to define the gate and ohmic contacts of all devices ; the six gate fingers are 0.5 µm long and 33 µm wide.

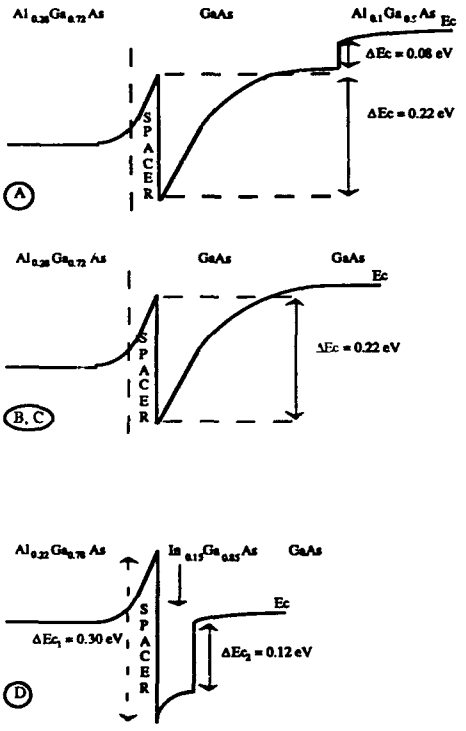


Figure 1 : Conduction band diagram :
A : AlGaAs/GaAs/AlGaAs HEMT
B, C : AlGaAs/GaAs/GaAs HEMT
D : AlGaAs/InGaAs/GaAs HEMT

3. EXPERIMENTAL RESULTS

LF noise investigations were conducted to evaluate and compare the influence of the physical characteristics due to the heterostructure layers on the global quality of devices. The measurements were performed in the ohmic regime and in the frequency range : 1 Hz - 100 kHz.

To obtain some insight into the sources of the LF excess noise, its evolution was analysed as a function of the gate and of the drain biases in the 90 K - 300 K range. The evolution of the current noise spectra as a function of the temperature has allowed to identify deep levels responsible for parasitic effects in the HEMT.

3.1 1/f noise analysis vs. drain voltage

As verified in figure 2, the 1/f noise spectra of all devices follow the Hooge relation [1]:

$$S_{I_D} = \frac{\alpha_H I_{DS}^2}{N} \frac{1}{f} \qquad [1]$$

where N is the total number of free electrons in the channel. Moreover, the current noise spectral intensity normalised to the squared drain current remains independent of the drain bias in the ohmic regime (20 mV ≤ Vds ≤ 70 mV). At f=10Hz, the values of the α_H/N factor (where α_H is the Hooge parameter) are respectively $1.1x10^{-10}$, $9.3x10^{-12}$ and $4.7x10^{-12}$ for A, B and C devices. So, the 1/f noise level is more important for A devices than for B and C devices. The residual dopants type of the GaAs buffer layer (p^- type for B devices and n^- type for C devices) has no influence on the channel noise level of conventional devices.

Owing to the presence of the AlGaAs buffer layer in A devices which induces a band conduction offset of 0.08 eV (fig 1.A), the free electrons are shielded from the substrate traps. This suggests that the 1/f noise of conventional devices in the ohmic regime is not issued from free carriers density fluctuations related to the injection of electrons into the substrate. The high 1/f current noise level could originate from structural defects in the GaAs buffer layer resulting from the deep GaAs/AlGaAs interface.

A value of 1.2×10^{-12} is obtained for α_H/N on PM-HEMT (device D), i.e about one or two orders of magnitude lower than for A, B and C devices. This confirms the better 1/f noise performances of PM-HEMTs.

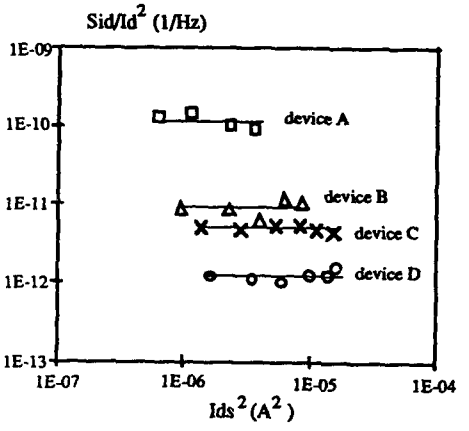


figure 2 : Normalised current noise spectral intensity vs. the squared drain current in the ohmic regime ($V_{GS} = 0V$)

3.2 1/f excess noise variation vs. gate bias

In the following, a comparison is performed between the best conventional HEMT (C device) and the PM-HEMT (D device) performances. In figure 3, the normalised drain noise spectral intensity is reported as a function of the inverse of the relative gate voltage $1/V_{GS}-V_T$. V_T is the threshold voltage of the HEMT precisely calculated from noise measurements in the ohmic regime.

As previously reported for conventional HEMTs (ref. 6), the evolution of the PM-HEMT normalised drain current noise versus $1/V_{GS}-V_T$ gives information on the origin of the 1/f excess noise and presents three different regions.

The total drain noise in the ohmic regime can be written as :

$$\frac{S_{I_D}}{I_{DS}^2} = \frac{S_{R_T}}{R_T^2} \simeq \frac{S_{R_{CH}} + S_{R_S}}{(R_{CH} + R_S)^2} \qquad [2]$$

$$\frac{S_{I_{DS}}}{I_{DS}^2} = \frac{1}{R_T^2}\left[\frac{\alpha_{CH}\, q \mu_{CH}}{L_G^2} R_{CH}^3 + \frac{\alpha_S\, q \mu_S}{(L - L_G)^2} R_S^3\right]\frac{1}{f} \qquad [3]$$

where $S_{R_{ch}}$ is the intrinsic channel noise, S_{R_s} represents the noise contribution of the access zones (from gate to drain and source), L_G the gate length and L the source-drain distance.

For values of $1/V_{GS}-V_T$ higher than 2, the series resistance equivalent noise is neglected when compared to the intrinsic channel noise. The slope of the plot is 1 or 3, depending upon the respective values of R_{CH} and R_S .

In the slope 1 region, the total resistance R_T (between drain and source) is reduced to the intrinsic channel resistance R_{CH} .

Then, the equation [3] becomes :

$$\frac{S_{I_D}}{I_{DS}^2} \simeq \frac{\alpha_{CH}\, q\, d}{L_G W \varepsilon} \frac{1}{(V_{GS} - V_T)} \frac{1}{f} \qquad [4]$$

replacing R_{CH} by :

$$R_{CH} = \frac{L_G\, d}{\mu_{CH} W \varepsilon} \frac{1}{(V_{GS} - V_T)} \qquad [5]$$

This assumption is verified by experimental results (figure 3) and α_{CH} appears to be independent of the gate bias in this region. An identical value of 3.8×10^{-4} is obtained for C and D devices ; this results from the higher number of free electrons in the intrinsic channel (2DEG) for PM-HEMTs than for S-HEMTs.

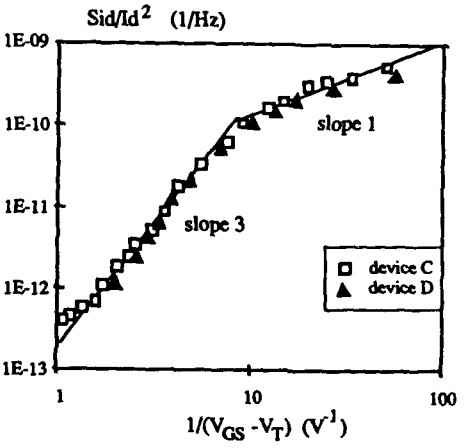


figure 3 : Evolution of the normalised drain current noise vs. the inverse of the relative gate voltage (f=20Hz - V_{DS} =50mV)

In the slope 3 region, the channel resistance R_{CH} is smaller in comparison with R_S. Then, the equation [3] gives :

$$\frac{S_{ID}}{I_{DS}^2} = \frac{\alpha_{CH}\, q\, \mu_{CH}}{L_G^2} \frac{R_{CH}^3}{R_S^2} \frac{1}{f} \qquad [6]$$

The theoretical expression [5] of R_{CH} is no more valid. A square dependence of R_T on $1/V_{GS}-V_T$ is observed (figure 4). The value of the slope results in this zone from the variations of $\frac{R_{CH}^3}{R_S^2}$ versus $1/V_{GS}-V_T$.

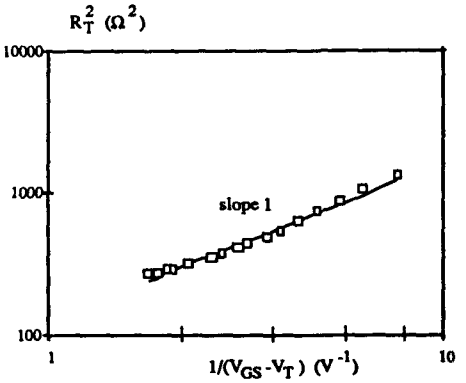


figure 4 : Total resistance R_T vs. the inverse of the relative gate voltage

For values of $1/V_{GS}-V_T$ lower than 2, the access resistance noise S_{RS} is dominating and depends on $1/V_{GS}-V_T$ as R_S. Then, the equation [3] is reduced to :

$$\frac{S_{ID}}{I_{DS}^2} = \frac{\alpha_S\, q\, \mu_S}{(L - L_G)^2} R_S \frac{1}{f} \qquad [7]$$

Following these assumptions, it is not trivial to separate the respective contributions of the series resistance and of the intrinsic channel to the current noise in the slope 3 region of the S_{ID}/I_D^2 plot reported in figure 3.

3.3 G-R noise vs. gate bias

As seen in figure 5, the PM-HEMT current noise behavior shows a G-R component which seems less preeminent than in S-HEMTs devices. This could originate from the lower Al molefraction of the n-AlGaAs in PM-HEMT. As already demonstrated by several authors (refs. 7,8), the preeminent G-R noise component which cut-off frequency is around 10 kHz (at ambient temperature) is supposed to be related to the DX traps contribution in the conventional HEMTs if the Al molefraction is higher than 0.22 (ref. 9).

Drain current noise measurements on PM-HEMT for gate voltage between -0.5V and 0V at fixed drain bias (50 mV) have been performed. For values of the gate bias V_{GS} close to the threshold voltage of the device, the excess channel noise presents a 1/f behavior at room temperature. When increasing the gate bias, G-R noise contributions appear in the noise spectra (figure 6). This evolution is similar to that observed on conventional HEMTs (ref. 10).

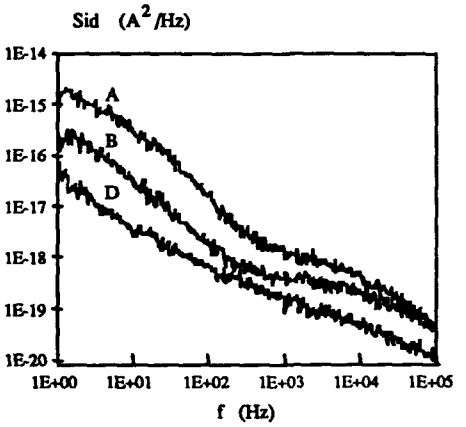


figure 5 : Comparison of the drain current noise spectra (measured at Ids=1.9mA) for A, B and D devices

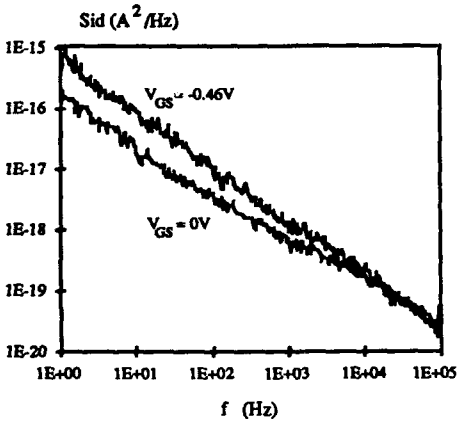


figure 6 : Evolution of the PM-HEMTs drain current G-R noise contributions as a function of the gate bias

3.4 G-R noise vs. temperature

At cryogenic temperatures, the HEMTs were operated in the ohmic regime at a constant drain bias.

For the temperature range 200K - 300K, the corner frequency of G-R contributions shifts towards the lower frequencies for B (figure 7), C and D devices.

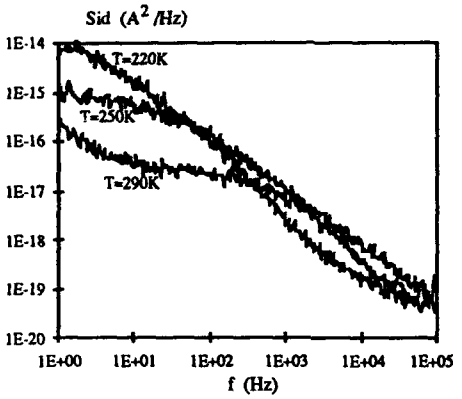


figure 7 : Evolution of the drain current noise spectral intensity as a function of temperature (for B device)

The shift of the corner frequency fc versus temperature is drawn on the Arrhenius plot of $\log(T^2/fc)$ (figure 8). The activation energy Ea and the capture cross-section σ are obtained from a least-squares linear regression. In table II, Ea and σ are reported for B, C and D devices. The values of these parameters clearly identify levels which can be classed as the generic DX center.

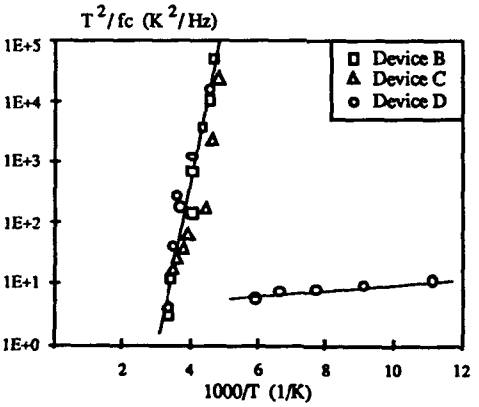


figure 8 : Arrhenius plot to identify :
- the DX center in B, C and D devices
- the specific trap in PM-HEMTs

For the PM-HEMT, another G-R noise contribution appears in the temperature range of 170K to 90K. The activation energy Ea and the capture cross-section σ of this specific level are reported in table II.

Device type	Ea ± 0.05 (eV)	σ (cm²)
B	DX traps : 0.5	1.2×10^{-15}
C	DX traps : 0.4	5.3×10^{-17}
D	DX traps : 0.55	3.2×10^{-14}
D	specific level : 0.06	1.3×10^{-21}

Table II : Characteristics of traps identified in B, C et D devices

In reference 11, a trap with an activation energy of 0.08 eV has been reported in PM-HEMTs in the temperature range of 250K-320K and its origin attributed to the InGaAs-GaAs interface.

4. CONCLUSIONS

The LF drain current noise analysis of conventional and pseudomorphic HEMTs (issued from identical metallurgical processes) has been performed to evaluate the global quality of these devices through the α_H/N factor. The lower PM-HEMT α_H/N parameter value verifies that a better carrier confinement is achieved from the significant conduction band discontinuity at the AlGaAs/InGaAs interface and from the deep offset at the InGaAs/GaAs buffer. A higher sheet charge density is maintained in PM-HEMTs than in S-HEMTs but very similar intrinsic channel contributions to the LF noise current are observed through an identical value of the α_{CH} parameter : 3.8×10^{-4}.

DX centers have been characterised in conventional and pseudomorphic HEMTs according to the G-R noise dependence on temperature. It has been verified that a low Al molefraction in the n-AlGaAs layer implies a lower G-R noise level and therefore a weaker deep traps concentration. The detection and the characterisation of a shallow trap in PM-HEMTs through the G-R noise evolution versus temperature has been related to the AlGaAs/InGaAs interface structure.

5. REFERENCES

1. P. Audren, J.M. Dumas, M.P. Favennec and S. Mottet "Etude des pièges dans les transistors à haute mobilité électronique sur GaAs à l'aide de la méthode dite de 'relaxation isotherme'. Corrélation avec les anomalies de fonctionnement", J. Phys. III, 3, February 1993, pp 185-206.

2. S. Kugler "Evaluation of the trap concentration in highly doped semiconductors from low frequency noise spectra" J. Appl. Phys, 66 (1), July 1989, pp 219-222

3. J.R. Kirtley, T.N. Theis, P.M. Mooney and S.L. Wright "Noise spectroscopy of deep level (DX) centers in GaAs-$Al_xGa_{1-x}As$ heterostructures" J. Appl. Phys, 63 (5), March 1988, pp 1541-1548

4. F. Ali and A. Gupta "HEMTs and HBTs devices, Fabrication and Circuits", Ed. Artech House (USA), 1991, p 106

5. H. Morkoc, W.F. Kopp, T.J. Drummond, S.L. Su, R.E Thorne and R. Fisher "Submicron gate GaAs/$Al_{0.3}Ga_{0.7}As$ MESFETs with extremely sharp interfaces (40 Å)", IEEE Trans. Electr. Dev., Vol. ED-29, N°6, June 1982, pp 1013-1018

6. J.M. Peransin, P. Vignaud, D. Rigaud and L.K.J. Vandamme "1/f noise in MODFETs at low drain bias", IEEE Trans. Electr. Dev., Vol. ED-37, N°10, October 1990, pp 2250-2253

7. F. Hofman, R.J.J. Zijlstra, J.M. Bettencourt de Freitas and J.C.M. Henning "Generation-recombinaison noise in $Al_xGa_{1-x}As$", Sol. St. Elect., Vol. 34, N°1, 1991, pp 23-32

8. N. Labat, N. Saysset, D. Ouro Bodi, A. Touboul, Y. Danto, C. Tedesco, A. Paccagnella and C. Lanzieri "Investigations on the origin of AlGaAs-GaAs HEMTs LF channel noise", Proc. ICNF, St Louis (USA), August 1993, pp 264-267

9. idem reference 4, p 83

10. N. Labat, D. Ouro Bodi, A. Touboul, Y. Danto and J.M. Dumas "Low frequency noise analysis to detect the influence of deep levels in AlGaAs/GaAs HEMTs", QREI, Vol. 8, 1992, pp 301-305

11. R. Plana, L. Escotte, O. Llopis, H. Amine, T. Parra, M. Gayral and J. Graffeuil "Noise in AlGaAs/InGaAs/GaAs pseudomorphic HEMTs from 10 Hz to 18 GHz", IEEE Trans. Electr. Dev., Vol. ED-40, N°5, May 1993, pp 852-858

GaAs MMIC SWITCH RELIABILITY

W. T. Anderson, J. A. Roussos, and K. A. Christianson
Naval Research Laboratory, Code 6855, Washington, DC, USA

ABSTRACT

Accelerated high temperature RF life testing was conducted on MMIC single pole single throw (SPST) and single pole double throw (SPDT) switches. The primary failure mechanism was found to be degradation of the passivation separating the first and second level metalizations.

1. INTRODUCTION

Previous reliability studies of GaAs monolithic microwave integrated circuits (MMICs) have concentrated on low noise and power amplifiers (Refs. 1-4). In the present paper we report on a study of switch reliability which are, to our knowledge, the first quantitative results reported to date. Even though only a limited number of devices were available for this study, it was interesting to determine if there was enough statistical validity to obtain useful information. As we have found in the present MIMIC Program, only a small number of expensive circuits are frequently all that are available for reliability studies. For example, only 10 GaAs MMIC modules of each type will be delivered for performance, reliability, and other types of testing.

2. RF LIFE TESTING PROCEDURES

MMIC single pole single throw (SPST) switches were RF life tested under accelerated high temperature conditions at baseplate temperatures of 170°C and 200°C, following the JEDEC JEP118 Guidelines (Ref. 5) as closely as possible. During life testing, the SPST switches were driven by an RF signal at a frequency of 8.4 GHz and a level of +4.0 dBm. To simulate actual use conditions, the switches were modulated by a 0V to -5V square wave with a 10 kHz repetition rate and a 50% duty cycle. Single pole double throw (SPDT) switches were also life tested under similar conditions but at baseplate temperatures of 170°C, 200°C, and 230°C with an RF signal level of -10.0 dBm. The square wave bias was used to constantly switch the devices between the ON and OFF states, thus simulating the operating conditions they would experience in practice.

Prior to RF life testing, small signal s-parameters were measured vs. frequency on a network analyzer. All the switches were measured in both the ON and OFF states. The MMICs were then placed on the life test apparatus and measured under the same dc bias conditions, but with the RF input signal set to the levels specified above. Initial values of output power, insertion loss or gain were then recorded both at room temperature and at a baseplate temperature of 125°C and used as pre-stress figures of merit. The MMICs were then heated to stress temperature (170°C, 200°C, or 230°C) and the 10 kHz square wave bias was applied. The square wave bias was maintained whenever the switches were at stress temperature. Approximately every 500 hours, output power and insertion loss or gain were remeasured under constant bias at room temperature and at 125°C and compared to the initial values. For purposes of evaluation, failure was defined as a ±0.7 dB change in insertion loss, gain, or pass band ripple, a ±1.8 dB change in isolation, or a ±2.4 dB change in return loss.

The RF life test stations were designed so that the microwave input power, dc bias, and baseplate temperature of each device on test could be individually monitored and controlled. Input power to each device was adjusted by means of a variable attenuator and monitored, via a 20 dB directional coupler, to ensure that the level remained constant throughout the test. Crystal detectors were used to monitor the output power of each device. Temperature control was achieved with a cartridge heater, a two-wire thermocouple, and an individual temperature controller for each device. In the latter stages of the study, computer control capability was implemented with a 286 PC and an analog to digital interface consisting of 16:1 multiplexer boards, solid state relays, a 16 channel A/D board and a 96 channel digital I/O board. A QuickBasic program was written to control measurement and data storage of baseplate temperature, dc bias level, and RF output power, and to shut off all power to a device in the event of catastrophic failure.

As there were only eight SPST MMICs available for life testing it was decided to use only two temperatures in order to have a minimum number for statistical analysis. Accelerated high temperature RF life testing was performed on the SPST switches at baseplate temperatures of 170°C and 200°C. In Figure 1, the time of failure for each SPST switch at a given temperature is plotted versus cumulative percentage of failures at that temperature, assuming a lognormal distribution. The data at each temperature were then fitted on computer by exponential least squares and a standard deviation, σ, was calculated for each curve. In Figure 2 the median time of failure (median life, t50) at each stress temperature is plotted versus channel temperature in accordance with the Arrhenius model of temperature acceleration. Since negligible dc current flows in either the SPST or the SPDT switches, baseplate temperature is a good approximation of channel temperature. The data were then fitted on computer with a logarithmic curve fit to enable extrapolation over temperature. For example, the median life of the SPST switches extrapolated to a channel temperature of 125°C is 1.4×10^4 hours with an activation energy of 0.41 eV.

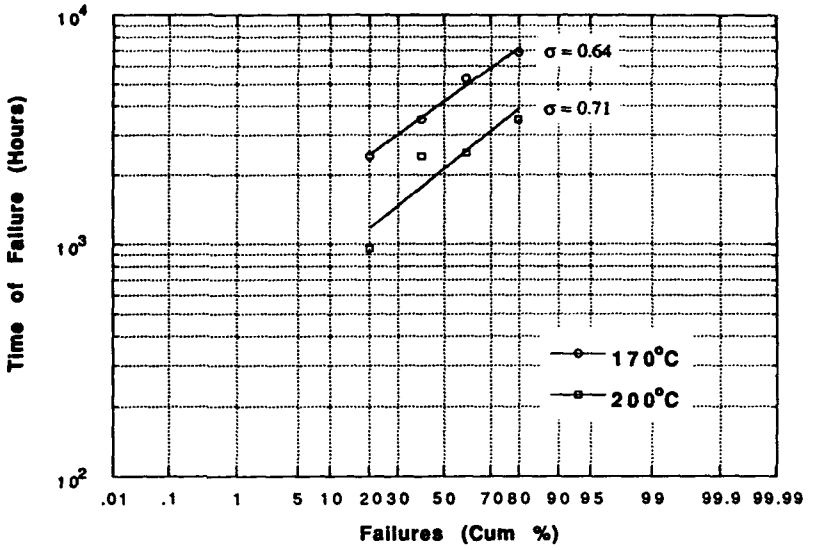


Figure 1. Lognormal plot of time of failure vs. cumulative percentage of failures for the SPST switches, RF life tested at baseplate temperatures of 170°C and 200°C.

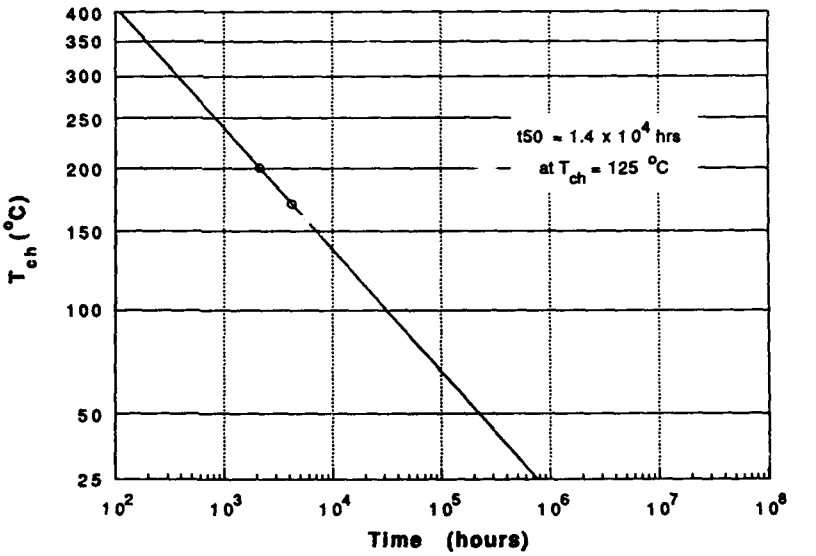


Figure 2. Arrhenius plot of median life, t50, vs. channel temperature for the SPST switches, RF life tested at T_{ch}=170°C and 200°C, corresponding to an activation energy of 0.41 eV.

In the case of the SPDT MMICs, there were nine switches available and a three temperature life test was conducted. This allowed a comparison with the two temperature life tests with the SPST switches. Because the designs are very similar and the fabrications methods almost identical, similar life test results were expected. Accelerated high temperature RF life testing was performed on the SPDT switches at baseplate temperatures of 170°C, 200°C, and 230°C. As with the SPST switches, lognormal and Arrhenius plots were made for the SPDT switches and are shown in Figures 3 and 4 respectively. From the Arrhenius plot, we estimate a median life of 3.4×10^4 hours at a channel temperature of 125°C. The corresponding activation energy is 0.54 eV.

3. FAILURE ANALYSIS

Failure analysis was carried out on the SPST and SPDT MMICs that failed during life testing. Discussed below are the results of analyses using optical microscopy, scanning electron microscopy (SEM), and energy dispersive x-ray analysis (EDAX).

MMIC SPST Switches

All eight of the SPST switches have had failure analysis performed on them. A summary of the observed characteristics is given in Table 1. For the catastrophic failures (#1, 6, 7, and 10) this has included both optical and scanning electron microscopy. The parametric failures (#2, 4, 5, and 9) had only optical microscopy performed on them in case further electrical measurements need to be made.

The most interesting feature observed is that the damage observed, with the exception of MMIC #7, has not been at the transistors, but rather along the control (C and C') lines running to the transistors. For both of these MMICs damage occurred along the line running from the control input pad to the lowest center most transistor structure. This has been labeled Q17 in the SPST switch layout shown in figure 5. In the case of MMIC #1 there was also damage extending upwards from the control pad towards the airbridge structure. The existence of similar damage for both catastrophically failed (e.g. #1) and parametrically failed MMICs (e.g. #2) indicates the same failure mechanism is responsible for both types of failure.

Analysis of the transistors of the circuits showed no visible damage with the exception of MMIC #6, which showed bits of metalization everywhere including the transistors. The damage to MMICs #9 and #10 also dealt with the control lines, but at a different location. For both of these MMICs, the initial cross-over point between control and control' was damaged.

From the evidence gathered thus far, the likely failure mode is a shorting of the control and control' lines, due to a failure of isolation due to dielectric degradation between the two lines at cross-over points. It does not appear that the transistors have failed. Analysis of FETs, by measuring transistors disconnected from the circuit, is planned.

MMIC SPDT Switches

The nine GaAs MMIC SPDT switches have had failure analysis performed on them. A summary of the observed characteristics is given in Table 2. As can be seen in this table, six of the SPDT switches have damage to the control and control' lines. Figure 6 shows the position of these lines in the SPDT switch layout. For three of the switches (#'s 1, 4, and 5), the damage consisted of a damaged control and/or control' line. For the other three switches (#'s 3, 9, and 10), the damage was more severe, with melted lines at the cross-over points. For SPDT switch #3, the damage extended from the C input to a cross-over point with C'. For SPDT switches #'s 9 and 10, the damage was more extensive, with damage extending further along the control lines. Like the SPST switches, this type of damage implies failure at the control line cross-over points due to dielectric degradation for these six switches. The physical and/or chemical causes of the degradation has not yet been determined.

4. CONCLUSIONS

Accelerated high temperature RF life testing was conducted on GaAs MMIC single pole single throw (SPST) and single pole double throw (SPDT) switches. The results differ from most previous MMIC life test studies, in which the failure site was almost always at the hot spot of one of the active devices in the circuit. Although the population was small for both types of switches, because of the limited number of devices available, the median life (t50) data do appear to be statistically significant. A minimum of nine devices should be studied in a three temperature life test. The results indicate that the lifetimes are only of the order of 2 to 4 years at a channel temperature of 125°C. Improved technology, particularly in regard to passivation, will be required to extend the median life by one or two orders of magnitude required for most field applications.

5. REFERENCES

1. Anderson, W. T. and Christou, A. "Status of Compound Semiconductor Device Reliability," in Semiconductor Device Reliability, A. Christou and B.A. Unger, eds., (NATO Scientific Affairs Division and Kluwer Academic Publisher, 1990), pp. 423-437.

2. Reliability of Gallium Arsenide MMICs, A. Christou, editor, (Wiley, Chichester, England, 1992).

3. Anderson, W. T., Roussos, J. A., and Christianson, K. A. "GaAs MMIC Reliability Studies," Quality and Reliability Engineering International, Vol. 8, pp. 295-300 (1992).

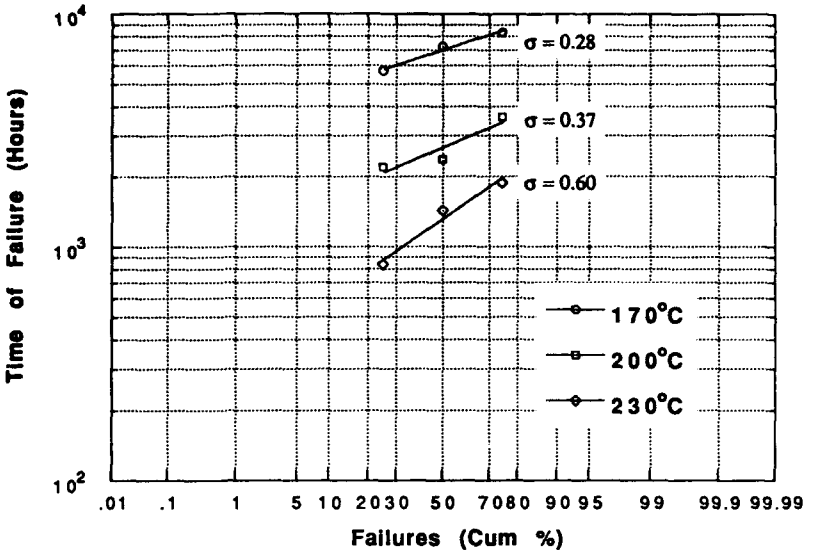


Figure 3. Lognormal plot of time of failure vs. cumulative percentage of failures for the SPDT switches, RF life tested at baseplate temperatures of 170°C, 200°C and 230°C.

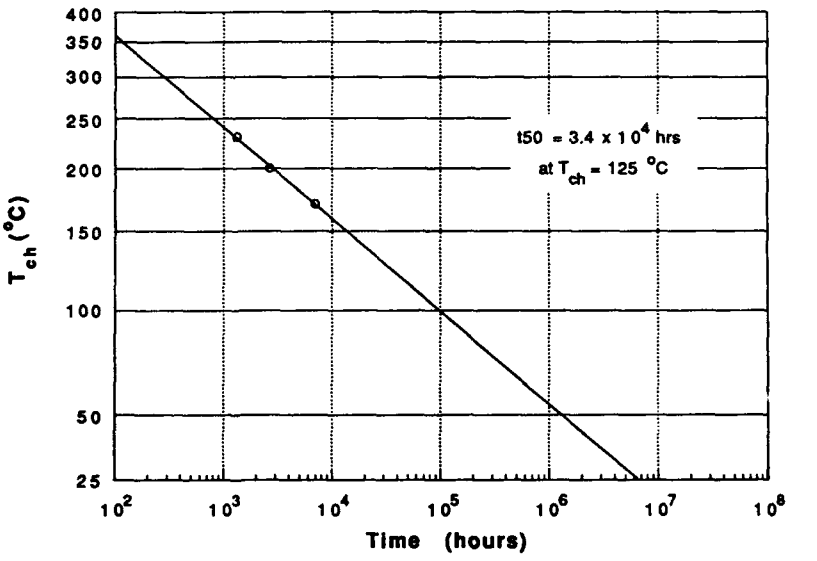


Figure 4. Arrhenius plot of t50 vs. channel temperature for the SPDT switches, RF life tested at T_{ch}=170°C, 200°C and 230°C, corresponding to an activation energy of 0.54 eV.

4. Christianson, K. A., Roussos, J. A., and Anderson, W. T. "Failure Analysis of MMIC Power Amplifiers," Proceedings of 4th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF 93), (University of Bordeaux, 1993), pp. 453-458.

5. JEDEC Publication JEP-118, "Guidelines for GaAs MMIC and FET Life Testing," (Electronic Industries Association, Washington, D.C., January 1993).

ACKNOWLEDGMENTS: This work was supported by the Office of Naval Research. The authors wish to thank J. A. Mittereder of SFA, Inc. for computer control of the life test system and assistance in the failure analysis.

Table 1 Observations of failed MMIC SPST switches.

MMIC:	Observations:
1	damage along control line to Q17 damage along control line above C input bits of metalization at airbridge near Q22
2	damage along control line to Q17
4	top left 10% of chip gone damaged area near control line input
5	no visible damage
6	damage along control line to Q17 bits of metalization everywhere
7	damage to Q20, Q21 gate, source, drain metalization melted
9	damage at cross-over intersection between C and C'
10	damage at cross-over intersection between C and C'

Table 2. Observations of failed MMIC SPDT switches.

MMIC:	Observations:
1	damage at intersection of C and C'
2	no visible damage
3	severe damage to C and C'
4	damaged C line
5	damaged C and C' line
6	no visible damage
8	burned out RFin line segment
9	severe damage to C and C'
10	severe damage to C and C'

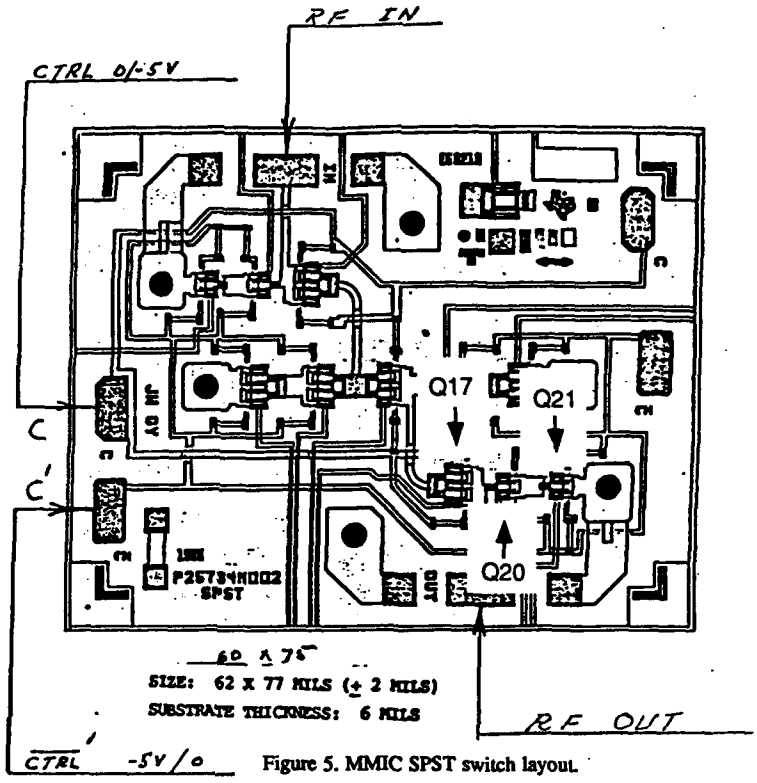


Figure 5. MMIC SPST switch layout.

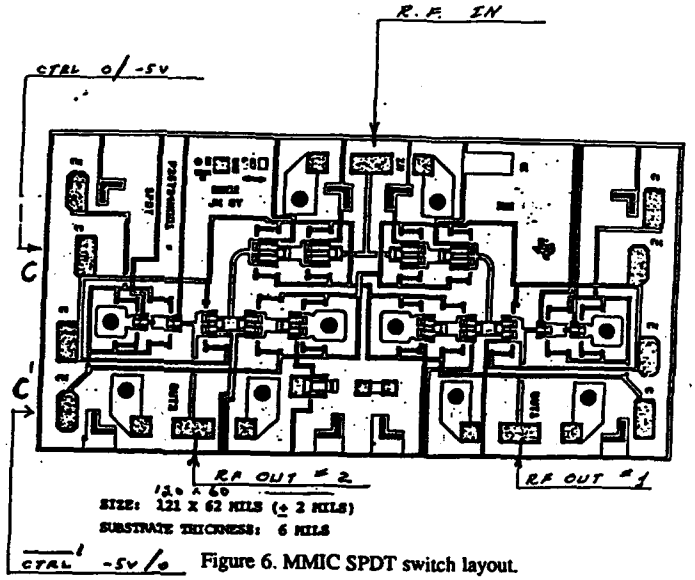


Figure 6. MMIC SPDT switch layout.

PICOSECOND ELECTRO-OPTIC SAMPLING SYSTEM WITH IMPROVED SPATIAL RESOLUTION USING THE TECHNIQUE OF A NEAR FIELD SCANNING OPTICAL MICROSCOPE

C. Böhm, J. Bangert, W. Mertin, and E. Kubalek

Gerhard-Mercator-Universität - GH Duisburg, Sonderforschungsbereich 254
Fachgebiet Werkstoffe der Elektrotechnik
Bismarckstraße 81, 47048 Duisburg, Germany

1. ABSTRACT

The picosecond electro-optic (e.-o.) sampling technique is an established test technique for testing signals in monolithic microwave integrated circuits (MMICs). However, with regard to existing trends in MMICs shrinking feature sizes will restrict its applicability in the future due to its limited spatial resolution in the order of several micrometer. In this paper an e.-o. sampling system which overcomes this limitation will be introduced. By using the technique of a near field scanning optical microscope with its inherent submicrometer spatial resolution combined with a pulsed laser system measurements of circuit internal signals with picosecond temporal resolution will be shown for MMICs based on GaAs-substrate. The extension of this new test technique to Si-based microelectronic devices is discussed.

2. INTRODUCTION

MMICs with submicrometer feature sizes and working frequencies above 100 GHz are already realized (Refs. 1, 2). Therefore, for the characterization of device function as well as for design and simulation verification a test technique with both high temporal and high spatial resolution is required.

The contactless electron beam test technique (Refs. 3, 4) is well established and well suited for probing within MMICs in the submicrometer region. However, its temporal resolution is limited by the transit time effect of the secondary electrons (Refs. 5, 6) and therefore, not longer applicable to higher frequencies.

Another established test technique is the contactless working e.-o. sampling technique based on the Pockels effect (Refs. 7, 8). For this test technique a lot of applications have been presented (Refs. 9 - 12). The e.-o. sampling technique achieves a superior temporal resolution but suffers from limited spatial resolution due to focus restriction based on the wavelength of the used laser beam in the micrometer region.

On the other hand a near field scanning optical microscope (NSOM) offers the potential of nanometer spatial resolution (Refs. 13, 14). By combining an e.-o. sampling system with the technique of a NSOM a picosecond e.-o.-NSOM sampling system with submicrometer spatial resolution seems to be realizable. To our knowledge, no work on this topic has been shown, yet.

Therefore, we present a new contactless test system based on the technique of a NSOM that combines the submicrometer resolution from scanning probe microscopy techniques with the gigahertz bandwidth of an e.-o. sampling system. In order to demonstrate the usability of the new test technique we use the e.-o. effect in a MMIC based on GaAs-substrate to measure circuit internal high speed signals. Due to the fact, that most of the ICs are based on SI we discuss the extension of this new test technique to Si-based microelectronic devices.

THEORY OF OPERATION

The e.-o. sampling technique is based on the linear e.-o. effect or the Pockels-effect in noncentric crystals (Fig. 1). Hereby, the local electric field in the substrate changes the optical properties of the GaAs in such a manner that the orthogonal components of a laser beam polarized in the x-y-plane and traveling in the z-direction get a phase retardation of $\Delta\Gamma$. This phase retardation is solely caused by the influence of the electric field component E_z being normal to the device surface. For a trip ($z = 0$ to $z = h$) of the laser beam from the back side of the MMIC-substrate to the substrate front side (test point) the phase retardation is:

$$\Delta\Gamma = \frac{2\cdot\pi}{\lambda}\cdot n_0^3\cdot r_{41}\cdot\int_{z=0}^{z=h} E_z dz \qquad (1)$$

$(2\cdot\pi/\lambda)\cdot n_0^3\cdot r_{41}$ is a constant with n_0 the index of refraction in absence of an electric field, r_{41} the linear e.-o. coefficient, and λ the wavelength of the laser beam. The phase retardation can be converted into an intensity variation of the detected laser beam after having passed a Pockels-cell. The intensity I detected by the photodiode is given by:

$$I \propto \int_{z=0}^{z=h} E_z dz \propto V, \qquad (2)$$

where V is the voltage between the back side of the substrate and the test point.

The principle of a NSOM is based on the Fresnel diffraction or optical near field effect (Ref. 16) where the optical wavefront at a diffracting aperture cannot be longer considered as planar but the wavefront curvature must be taken into acount. By positioning a fiber with a small aperture d $(d << \lambda)$ into the optical near field of a light source (distance light source/fiber $\delta << \lambda$) and collecting the light with the fiber a nanometer resolution is achievable limited by the aperture, only (Refs. 13, 14).

In the e.-o.-NSOM an optical fiber is immersed into the optical near field of the illuminating laser beam above the test point (see inset of Fig. 1). Hereby, the laser beam is focussed with a microscope objective lense through the substrate onto the test point and is partly reflected from the GaAs/air-interface. The remaining part of the laser beam is transmitted through the substrate and partly collected by the fiber. The polarization of the illuminating laser beam is changed by the electric stray field component E_z inside the GaAs-substrate which can be converted into a change of the detected laser beam intensity resulting in a measure which is proportional to the test point signal.

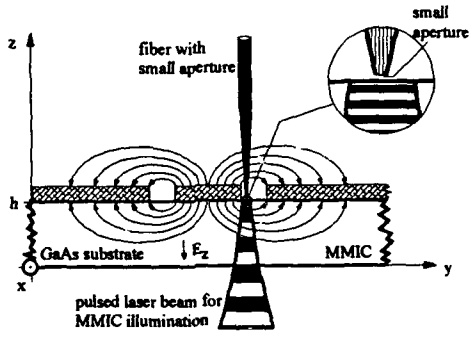


Fig. 1: Test geometry for probing coplanar MMICs with pulsed illumination laser beam and a fiber

TEST STRUCTURE AND EXPERIMENTAL SET-UP

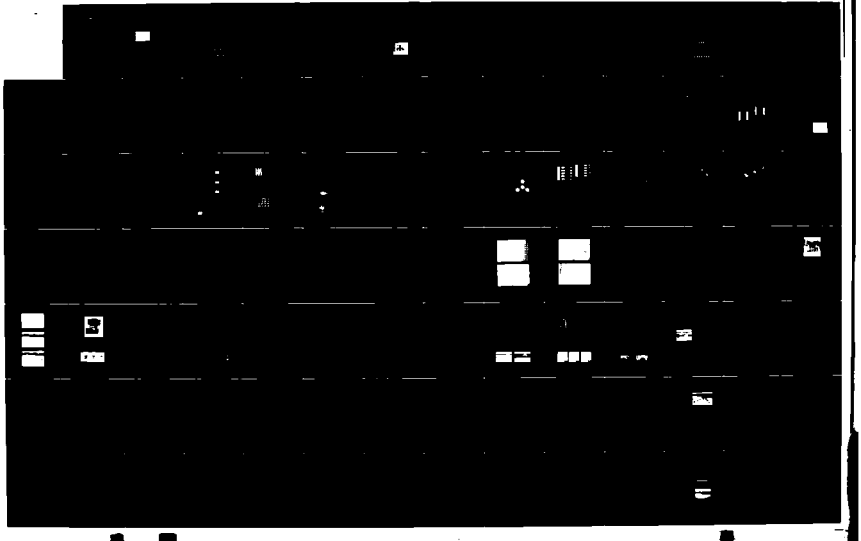
The MMIC used for the measurements is a test structure built up on a 2 x 3 mm^2 wide and 510 µm thick semi-insulating GaAs-chip (Fig. 2a). It consists of a set of coplanar

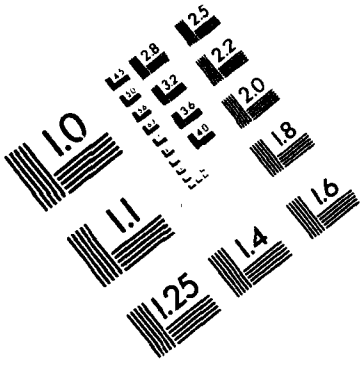
AD-A286 614 EUROPEAN SYMPOSIUM ON RELIABILITY OF ELECTRON DEVICES FAILURE PHYSICS AND..(U) GAREL ASSOCIATES NORTHAMPTON (UNITED KINGDOM) C M BRYDON 7 OCT 94 XB-ONREUR

4/7

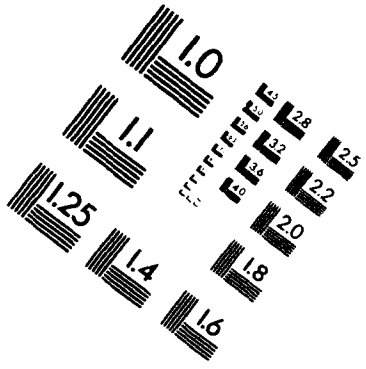
UNCLASSIFIED

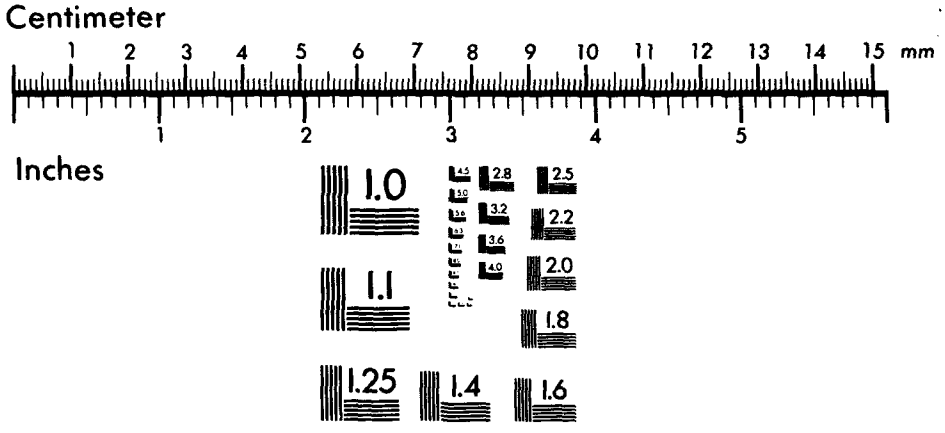
NL

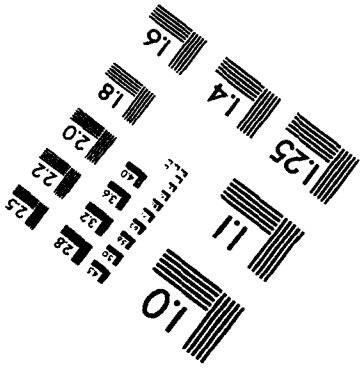


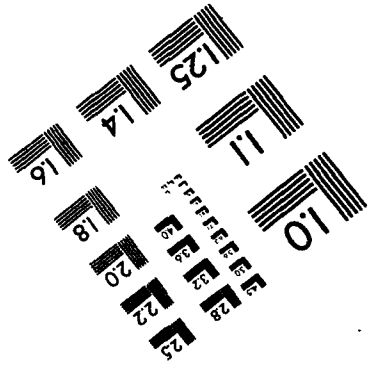


Association for Information and Image Management
1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202



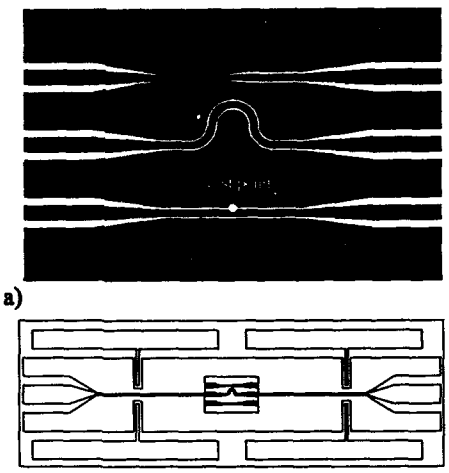






MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

waveguides with a line width of 40 μm and spacing of 26 μm to the outer ground planes. The electrodes are of gold, 3 μm thick, and have an impedance of 50 Ω. For the measurements only the lowest coplanar waveguide within the test structure is used. The test structure itself is mounted on a glass carrier (Fig. 2b) on which there are also two 50 Ω coplanar transmission lines. Electric connections to the GaAs-chip are made by bondwires. Finally, at the outer end of the glass carrier there are transitions from coplanar to coaxial transmission lines.



a)

b)

Fig. 2: Layout of the test structure (a) with three different coplanar waveguides based on a 510 mm thick semi-insulating GaAs substrate and glass carrier (b)

The experiments are carried out with a modified commercial NSOM (TopoMetrix Aurora (Ref. 16)) adapted to an e.-o. sampling system (Ref. 11) as shown in Fig. 3. The laser beam enters the MMIC from the back side. The transmitted part of the laser beam is partly collected via the fiber, analysed by a Pockels cell and fed into photodiodes. Two high sensitive photodiodes, optimized for the wavelength of 1064 nm, detect separately the intensity of the two orthogonal polarization components of the polarized laser beam and an oscilloscope or a lock-in amplifier measures the difference signal in magnitude giving a measure of the amplitude of the normal component E_z of the electric field within the MMIC substrate.

The MMIC under test is driven by a microwave synthesizer tuned to exactly the N-th harmonic of the laser pulse repetition frequency with a frequency offset Δf of 4 kHz. By using the harmonic mixing technique (Ref. 8), the frequency spectrum of the laser beam pulse is mixed with the spectrum of the microwave signal resulting in a transformation of the microwave signal frequency f down to the frequency Δf which can easily be processed furtheron.

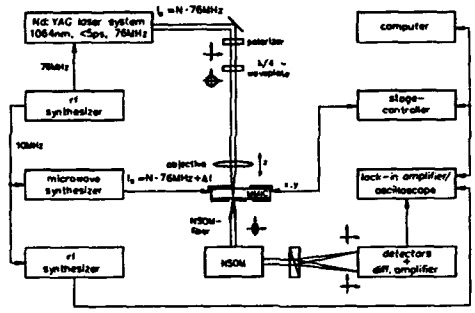


Fig. 3: Experimental set-up for e.-o.-NSOM measurements

5. EXPERIMENTAL RESULTS AND DISCUSSION

In all following experiments we show the possibility of time resolved e.-o. measurements with an e.-o.-NSOM. All experiments have been performed with an average power of the illuminating laser beam of about 200 mW. The test point was always on the lowest waveguide

indicated in Fig. 2. The waveforms have been averaged 100 times.

The spatial resolution of a NSOM is well investigated and 50 nm are still achievable with fiber tips 50 nm (Ref. 17).

For the first two experiments the applied signal has been sinusoidal with a frequency of 4 kHz which is the frequency offset Δf. These experiments have been made in order to show the linearity of our test system and to make a direct comparison between signals measured with the e.-o.-NSOM and an oscilloscope. With the third experiment we demonstrate the high temporal resolution obtainable with the new test system.

Referring to Eq. 2 the e.-o.-NSOM measured signal depends linearly on the applied test point voltage. Fig. 4 shows the measured signal as a function of the applied test point voltage. A very good linearity for typical MMIC voltages can be observed. Measurements below 3 V test point voltage are not performed because of the bad signal to noise ratio.

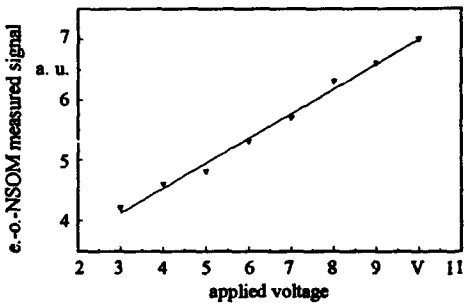


Fig. 4: e.-o.-NSOM measured signal versus applied test point voltage.

In the second experiment the e.-o.-NSOM measured signal is compared with the output of the waveguide measured with an oscilloscope. Apart from some noise the e.-o.-NSOM measured signal (Fig. 5 lower curve) is identical with the oscilloscope measured signal (Fig. 5b upper curve).

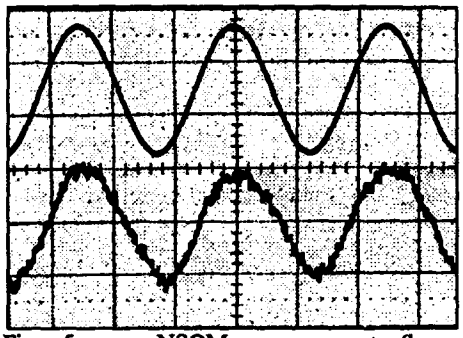


Fig. 5: e.-o.-NSOM measurement (lower curve) of a signal waveform at 4 kHz. For comparison a waveform measured with an oscilloscope is also shown (upper curve). The applied power level was 30 dBm.

In the last experiment a microwave signal with a frequency of 7.59 GHz has been measured (Fig. 6). The applied sinusoidal waveform can be clearly seen. The decrease in the signal to noise ratio results from the specific transmission characteristic of the used glass carrier.

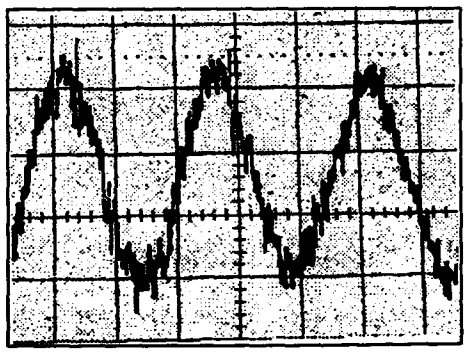


Fig. 6: e.-o.-NSOM measurement of a signal waveform at 7.59 GHz. The applied power level was 25 dBm.

6. APPLICATION TO E.-O. SAMPLING OF Si-BASED DEVICES

Most of the ICs are based on Si-substrate. There is also a trend towards higher operation frequencies and higher bandwidth. In contrast to GaAs Si does not show the e.-o. effect, because it has a centric crystal structure. Therefore, at microelectronic devices based on Si-substrate the substrate cannot be used as an e.-o. modulator. In this case the indirect e.-o. sampling technique has to be used (Refs. 7, 10). Hereby, an e.-o. probe tip is immersed into the electric stray field above the test point. The laser beam is focussed through this tip, is reflected from the test point, and its polarization is modulated within the e.-o. probe tip.

To performe indirect e.-o. sampling with a NSOM, an optimized NSOM-fiber showing the e.-o. effect within the fiber geometry as shown in Fig. 7 has to be built. The standard NSOM fiber is combined with an e.-o. crystal of thickness $t << \lambda$. The transparent electrode between the NSOM-fiber and the e.-o. crystal acts as a reference electrode. By positioning the e.-o. fiber tip in a distance $\delta << \lambda$ above the test point and by grounding the reference electrode a calibrated voltage measurement is possible.

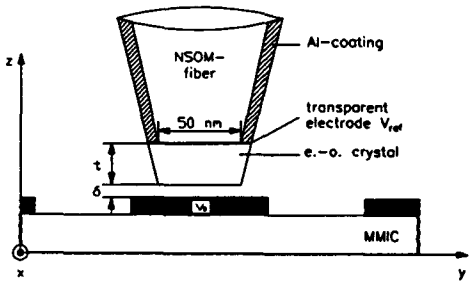


Fig. 7: e.-o.-NSOM fiber geometry for indirect e.o. sampling with voltage measurement.

7. CONCLUSION

The possibility to measure electro-optically ultrafast signals with a e.-o.-NSOM has been demonstrated up to 7.5 GHz. The measurements have been performed on a MMIC based on GaAs-substrate. Additionally, we discussed the extension of this new test technique to e.-o. sampling of Si-based devices.

ACKNOWLEDGMENTS

The authors would like to thank the Deutsche Forschungsgemeinschaft and the Ministry of Science and Research of North-Rhine Westfalia.

8. REFERENCES

1. S. Y. Chou and M. Y. Liu, **"Nanoscale Tera-Hertz metal-semiconductor photo-detectors"**, IEEE J. Quantum. Electron., vol. QE-28, no. 10, 1992, pp. 2358 - 2368

2. R. Majidi-Ahy, C. K. Nishimo, M. Riaziat, M. Glenn, S. Silvermann, S. Weng, Y. Pao, G. Zdasiuk, S. G. Bandy and Z. C. H. Tan, **"5 - 100 GHz InP coplanar waveguide MMIC distributed amplifier"**, IEEE Trans. Microwave Theory Tech., vol. MTT-38, no. 12, 1990, pp. 1986 - 1993

3. E. Menzel and E. Kubalek, **"Fundamentals of electron beam testing of integrated circuits"**, Scanning5(1983), pp. 103 - 122

4. E. Wolfgang, **"Electron beam testing"**, Microelectron. Eng. 4(1986), pp. 77 - 106

5. H. Fujioka, K. Nakamae and K. Ura, **"Analysis of the transit time effect on the stroboscopic voltage contrast in the scanning electron microscope"**, J. Phys. D: Appl. Phys. 18, 1985, S. 1019 - 1027

6. W. Mertin, K. D. Herrmann and E. Kubalek, **"Electron beam testability of monolithic microwave integrated circuits (MMIC)"**, Microelectron. Eng. 12(1990), pp. 287 - 293

7. J. A. Valdmanis and G. A. Mourou, **"Subpicosecond electrical sampling and applications"**, in: Picosecond Optoelectronic Devices, ed.: C. H. Lee, Academic Press, Inc., 1984, pp. 249 - 270

8. B. H. Kolner and D. M. Bloom, **"Electro-optic sampling in GaAs integrated circuits"** IEEE J. Quantum Electron., vol. QE-22, no. 1, 1986, pp. 79 - 92

9. K. J. Weingarten, M. J. W. Rodwell and D. M. Bloom, **"Picosecond optical sampling of GaAs integrated circuits"**, IEEE J. Quantum Electron., vol. QE-24, no. 2, 1988, pp. 198 - 220

10. M. Y. Frankel, J. F. Whitaker, and G. A. Mourou, **"Optoelectronic transient characterization of ultrafast devices"**, IEEE J. Quantum Electron., vol. QE-28, no. 10, 1992, pp. 2313 - 2324

11. W. Mertin, A. Leyk, F. Taenzler, T. Novak, G. David, D. Jäger and E. Kubalek, **"Characterization of a MMIC by direct and indirect electro-optic sampling and by network analyzer measurements"**, Microelectron. Eng. 24(1994), pp. 377 - 384

12. G. David, S. Redlich, W. Mertin, R. M. Bertenburg, S. Koßlowski, F. J. Tegude, E. Kubalek and D. Jäger, **"Two-dimensional direct electro-optic field mapping in a monolithic integrated GaAs amplifier"**, in: Proc. 23rd European Microwave Conference, 6. - 9. September 1993, Madrid, E

13. D. W. Pohl, W. Denk, and M. Lanz, **"Optical stethoscopy: Image recording with resolution $\lambda/20$"**, Appl. Phys. Lett. 44(7), 1984, pp. 651 - 653

14. E. Betzig, P. L. Finn, and J. S. Weiner, **"Combined shear force and near-field scanning optical microscopy"**, Appl. Phys. Lett. 60(20), 1992, pp. 2484 - 2486

15. F. L. Pedrotti and L. S. Pedrotti, **"Introduction to optics"**, Prentice-Hall, Inc., 1987

16. TopoMetrix, Pallaswiesenstr. 180, 64293 Darmstadt, Germany

17. P. Moyer, S. P. Marchese-Ragona, and B. Christie, **"Biological applications of scanning near-field optical microscopy (SNOM)"**, Application Note, TopoMetrix, Santa Clara, CA

PSEUDOMORPHIC HEMTs RELIABILITY WITH SCATTERING AND NOISE PARAMETERS

P. Conti, P. Audagnotto

CSELT S.p.A.
Via G. Reiss Romoli 274, 10148 Torino, Italy
Phone: +39 11 2285014 Fax: +39 2285695

Abstract - The knowledge of the scattering and of the noise parameters gives a complete description of the radio frequency operation of a linear two-port circuit.
In this work, the measurement of these parameters, together with the d.c. parameters, has been performed on low power, low noise pseudomorphic high electron mobility transistors during a reliability test of thermal cycles executed according to MIL-STD883D.

The reliability of HEMTs and of pseudomorphic HEMTs was studied in previous works (Ref.1,2) with storage tests at high temperature with and without bias. The main failure modes detected during these tests were the burn out and the degradation of the Schottky barrier that started around 170 °C. All these failures involve the metallizations in such a way that what has happened in the semiconductor layers is hidden to the electrical measurements.
In this work, to avoid the early degradation of the metallizations, a different test has been performed: a thermal cycles test without bias with an higher temperature of 125 °C. During the test, after a fixed number of cycles, the performances of the PHEMTs were accessed in d.c. as well as in RF; the Id-Vds curves, the Ig reverse, the Ig forward and the transconductance were measured along with the scattering and noise parameters between 2 GHz and 26 GHz, frequencies in the range of employment of the studied components.
After the description of the PHEMTs under test, of the reliability test and of the measurements performed, the measurements of the scattering parameters and of the noise parameters are presented together with the d.c. measurements.

1. PHEMTs AND THERMAL CYCLES TEST

The components under test, shown in figure 1, are AlGaAs/InGaAs/GaAs in chip pseudomorphic HEMTs guaranteed up to 35 GHz with a typical optimum noise figure at 18 GHz of 1.5 dB, a gain of 13 dB at 12 GHz and a maximum power dissipation of 400 mW. To perform the measurements, these devices were attached on a carrier of gold plated brass on allumina with an eutectic die attach (Preform AuSn (80% Au, 20% Sn). Stage temperature: 290 °C ± 5 °C) as specified in the data sheet. The bonding was made by thermocompression in a class 1000 clean room.

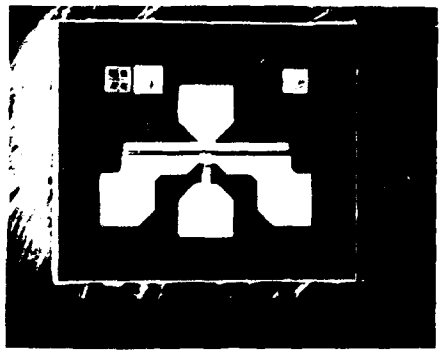


Fig. 1 - SEM photo of a component under test.

The thermal cycles test was performed on more than six PHEMTs according to MIL-STD883D, Test Method 1010.7, test condition B that provides that the components must be placed at -55 °C for ten

minutes and at 125 °C for other ten minutes with a transfer time from the cold to the hot temperature of less than 1 minute. This test is supposed to stress the semiconductor layer of InGaAs which forms a thermodynamically metastable structure (Ref.2) with the upper doped AlGaAs layer and the lower GaAs buffer layer. The InGaAs layer is twisted on its lattice planes due to the lattice constant mismatch between the semiconductors and the thermal cycles, in case of different thermal expansion coefficients, could increase the twisting and produce modifications in the semiconductor layers structure.

2. MEASUREMENTS PERFORMED

After a fixed number of thermal cycles, the scattering parameters S_{11}, S_{21}, S_{12}, S_{22} (Ref.3) and the noise parameters; the noise figure minimum Fmin, the equivalent noise resistance Rn normalized on the characteristic impedance of 50 Ohm and the optimum input reflection coefficient Γ were measured. The experimental set-up built for this purpose with an HP and Cascade instrumentation is presented in figure 2 (Ref.4).

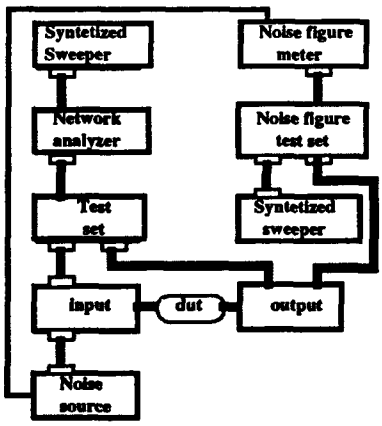


Fig. 2 - Experimental equipment for scattering and noise measurements.

It measures the scattering parameters directly with the network analyzer HP8510, while the measurement of the noise parameters employs a tuner Cascade, named Input in the figure, that applies 9 different impedances to the PHEMT under test. Using the classical measurement approach of IRE (Ref.5), the noise parameters are extracted from the noise figure measurements made by an HP8970B equipped with an HP8971C. This experimental set-up can work between 45 MHz to 26.5 GHz, but the measurements in this work are performed at the frequencies of employment of the PHEMTs.
Together with the radio frequency measurements, also the principal characteristics in d.c.: the curves I versus V, the transconductance at 3Vds, the forward gate current Igf and the reverse gate current Igr were measured.

3. EXPERIMENTAL RESULTS

During the test, variations were detected on the scattering parameters S_{21}, S_{12} and S_{11} and on the noise parameters Γ and Rn. The number of components on which these variations were detected and the number of thermal cycles at which they were measured are shown in Table I.

Parameter	n. devices / n. devices under test	Cycles number of first detection
S_{21}	4 / 6	5 c. on 2 dev. 20 c. on 2 dev.
S_{12}	2 / 6	5 c. on 1 dev. 20 c. on 1 dev.
S_{11}	1 / 6	5 c.
Γ	2 / 6	5 c.
Rn	6 / 6	5 c. on 1 dev. 20 c. on 5 dev.

Tab. I - Parameters that varied during the test, number of components on which the variations were detected and number of thermal cycles executed before them.

3.1 Scattering parameters measurements

The typical trend of the modulus of the scattering parameter S_{21} versus frequency during the test is shown in figure 3 that shows that, after 20 cycles, a decrease of the forward gain, to which S_{21} is linked, has occured. S_{21}, like other parameters shown in this paper, is drawn on a

Cartesian graph instead that on a polar diagram because from a Cartesian graph is sometimes easier to appreciate the variations of the measured parameters.

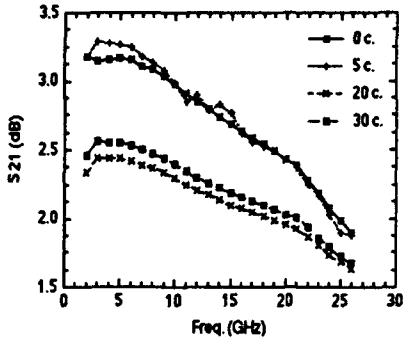


Fig. 3 - S_{21} modulus measured on a PHEMT under test.

The percentage variations of $|S_{21}|$ as to the initials values in dB have been calculated to be of about -20%, with the exception of one device in which it strangely varied of +20%.

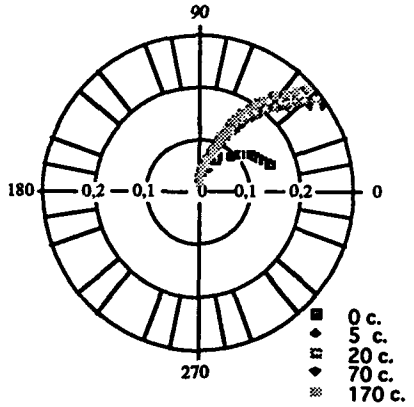


Fig. 4 - S_{12} measured in dB on a PHEMT under test.

As the other parameter S_{12} is concerned, it showed a variation well evident also on the polar diagram (Fig.4). On this device the modulus of S_{12} (dB) varied versus frequency from 0 to +100% at 26 GHz and on another device its variation was between 5% at 2 GHz and 16% at 26 GHz.

The scattering parameter S_{12} is linked to the reverse gain, and in this case indicates that the insulation between the output and the input of the device is getting worse.
Also a great variation of the modulus of the other parameter S_{11}, from 20% at 2 GHz (Initial value = 1) to -70% at 22 GHz (Initial value = 0.5), was detected on one device.

3.2 Noise parameters measurements

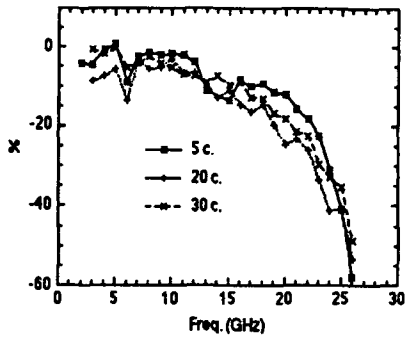


Fig. 5 - Per cent variations of Γ modulus versus frequency.

Among the noise parameters, the modulus of the optimum reflection coefficient Γ had great variations on two devices, even of about -55% at 25 GHz (Initial value = 0.6). The trend of the variations of this parameter versus frequency (Fig. 5) is similar for the two devices on which it happened.

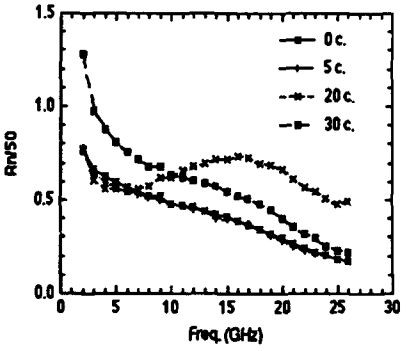


Fig. 6 - Normalized noise resistance measured on a PHEMT under test.

On the other hand, the equivalent noise resistance Rn increased between +30% and +80% on four devices out of the six tested. On the remained two devices, Rn had a different behaviour that must be studied further. The typical trend of Rn versus frequency is shown in figure 6.
Probably, after 20 cycles, the devices were put in an unstable and noisier configuration (Fig. 6, curve at 20 c.) from which they decaded after 30 cycles, founding a new state still noisier than that before the starting of the test.

3.3 D.c. measurements

During the test, no remarkable variations in the d. c. measurements Ig forward, Ig reverse, transconductance and I-V characteristic were measured. Besides, the examination of the components with the optical and SEM microscopes, that is in course, till now has not shown any components degradation.

4. CONCLUSIONS

The measurements of the scattering and of the noise parameters together with the d.c. measurements is a complete procedure to verify the functioning of the PHEMTs under test and to detect variations in their characteristic that can cause in circuit misfunctioning.
The RF parameters have shown an higher sensitivity than the d.c. parameters to the modifications induced in the PHEMT structure by the thermal cycles test but the interpretation of the many results obtained is not easy. The variations of the parameters are abrupt, and happen after a number of thermal cycles that make the devices to assume structural configurations with different RF properties. Probably some of these configurations can even improve the initial scattering and noise characteristics of the components, so that the anomalous variations seen on some measurements can be explained.
A further work with another more controllable test, for example a storage test under 170°C, not to stress the metallizations, with applied bias and with use of analytical techniques, could be useful to give a deeper interpretation of these aspects of PHEMTs reliability.

ACKNOWLEDGMENTS

The authors wish to thank Eng. R. Menozzi of the University of Parma who gave a great contribution to our work with helpful discussions and suggestions, together with Prof. F. Fantini. Thanks also to Mr. A. Valente of the Laboratories of SIP, that made possible the execution of the thermal cycles test, Eng. A. Angelucci, Dr. P. Montangero and Eng. M. Melgara for their encouragement.

REFERENCES

1. E. Pollino, "Microelectronics Reliability", Artech House, Vol. II, 1989.

2. A. Christou, "Gallium Arsenide MMICs", John Wiley and Sons, p. 169, 1992.

3. G. D. Vendelin, "Design of Amplifiers and Oscillators by the S-parameter Method", John Wiley and Sons, 1981.

4. P. Conti, P. Audagnotto, "Scattering and noise measurements of pseudomorphic high electron mobility transistors", Proc. GAAS, Torino, 1994.

5. 59 IRE S1 Committee Personnel, "IRE standards on methods of measuring noise in linear twoports", 1959.

RELIABILITY IN AUTOMOTIVE ELECTRONICS - A REVIEW.

R. Rathbone B.Sc.- Elmos GmbH, Dortmund

1. INTRODUCTION

No branch of industry, besides the aero-space industry, has ever been placed in a more stringent requirement for quality and reliability than todays automotive industry; and this using a cost structure for packaging and assembly techniques appropriate to consumer products. Failures of components in normal consumer products can cause discomfort, but because of the trust we place in the motorcar, failures in automotive applications can be a matter of life and death. Thus safety takes prime position in the list of desired improvements; safety, environment friendly performance, comfort, and reduced cost.

Formerly, the reliability of automotive electrical systems was dominated by failure in contacts and connectors, but the drive towards more and better functionality at lower cost is clearly replacing kilograms of copper wiring (up to 2000 m) with micrograms (or mm^2) of silicon chip, as illustrated in the modern Smart Power Bus Systems, such that the growth in the electronic content of automobile manufacturing costs can truly be described as explosive.

We must be sure that this rapid growth in the use of microelectronics is accompanied by attention to reliability, especially when safety is involved, in a field where somewhat different rules dominate, compared with the previous main body of electronic reliability studies.

Although automotive system suppliers are proud to announce electronic modules with lifetimes up to 30 years which far outlive the other vehicle components, the typical vehicle life is about 17 years with a variety of overstress conditions punctuated by 10 to 20 thousand switching disturbances and temperature cycles, for a functional life as short as 5000 hrs.

2. EVOLUTION

In the last three decades there have been three distinct, but overlapping, phases of automotive electronic development as described in the literature (Ref.1.) and illustrated in Fig.1.
Initially, through the 70's, many stand-alone functions were developed as simple add-on components, such as electronic ignition or digital clocks; the motivation being to increase functionality, to improve reliability by replacing mechanical functions, or to reduce manufacturing costs.
Standard discrete components and SSI standard ICs were mostly used on single level PC boards because of their availability.
The second phase, through the 80's, saw a strong increase in the level of digital electronic complexity and a greater degree of integration at the component level (MSI and LSI), which was accompanied by a higher degree of communication between the components themselves and also with sensors and actuators. Attention was given to fuel injection and engine control management because of the laws initiated in the US on exhaust emission control, and more powerful cars were demanding better safety features like ABS braking systems. This increased complexity raised questions of reliability, in particular because of the large number of electrical connections of all types (mechanical, solder joint, bond wires) involved, and because of the known inadequate reliability of some connections. For example since it was known that solder joints have a potential failure rate in excess of 200 ppm over the predicted lifetime of the car, concepts such as the 'Vital Core' electronics were developed with important safety and performance functions fabricated using either early ASICs or hybrid modules, mounted on PCBs, but containing a minimum number of solder cobnnections. But the cost of the hybrid approach was not acceptable for wider application.

In the third of the evolutionary phases, into the 90's, the automotive electronic designers have been able to design more completely integrated systems from the ground up, with the purpose of optimizing the vehicle electronic system for reliability, flexibility, adaptability, and lower cost.
Several trends have become established in this optimization process. Standard integrated circuits have been replaced by mixed signal ASICs, not only because they provide optimum technical and cost partitioning of the electronic sub-systems, but also because a high density of reliable interconnections can be placed within the chip metalization and at the bond wires to the ASIC, where higher reliability is obtained than at those connections external to the ASIC. The number of solder joints in the whole system can be further minimized by adoption of multiplexed single or dual wire Bus systems which are gaining in popularity not only because of the reduced costs of wiring and connectors, but also because of their increased fault diagnostic. A generalized Bus system with distributed control architecture is shown in Fig.2. Presently such systems with 30 to 40 interfaces per bus line are being introduced in high end vehicles like the new 700 series of BMW.

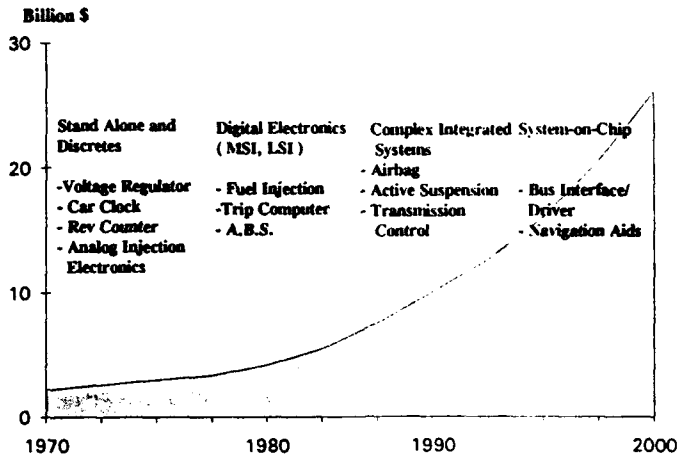


Fig. 1 Three Decades of Automotive Electronics

Future Bus systems will not have a large centralized computer, but will rely on distributed intelligence. The implications of the bus system feed forward into future ASIC chip design as illustrated in fig.3.
Such ASICs will include the following functional blocks:

- Sensor Interface Electronics
- Microprocessor Core (8 or 16 Bit)
- Memory (Both RAM, ROM and EEPROM)
- Internal Regulated Power Supply
- 2-Way Bus Interface (Communications Port)
- Smart Power Output to drive Actuators.

A mixed signal Smart Power ASIC technology is called for in a high level of complexity together with the need of a package to dissipate the heat generated by the output drivers. These features in themselves indicate that in future even higher reliability will be demanded from the Smart Power ASICs because the weight of functionality and connectivity is concentrated in them. To date microelectronic design and technology development has kept pace with the demands of complexity and reliability placed on them. Fig.4 demonstrates the dramatic improvement in ASIC reliability, expressed as AOQE, achieved in the last few years, and by the established principles of continuous improvement we expect the trend to continue.

In parallel with the improvements indicated for components, a wide range of automotive system improvements have been introduced:

- Suitable components are now used in well located positions.
- More exact control of incoming/outgoing quality (AOQ).

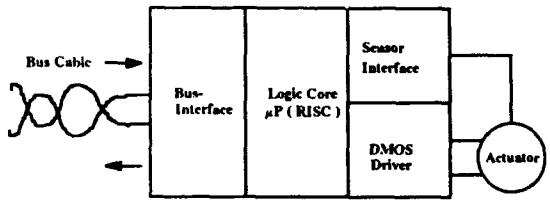


Fig.2 Bus System

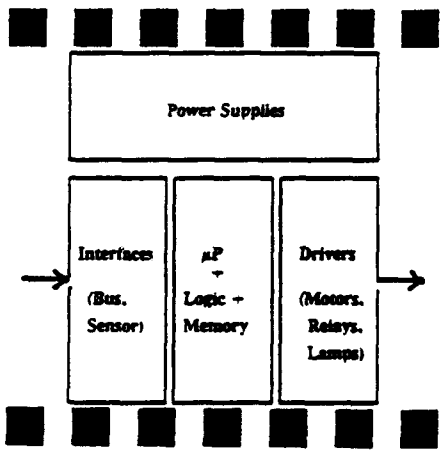


Fig. 3 Automotive ASIC System I.C

- Fewer system components used, leading to less connectors and contacts.
- Robust system design (built-in quality/design for reliability).

- Improved materials and design for connectors, fault detection and self test methods.
- Optimized reliability tests for temperature extremes and shock.
- Improved development methods and tools (simulation, CAD).
- Return and analysis of failed customer parts to continuously feed back improvements.
- Better training of service personel in electronic systems, diagnostics and repair.

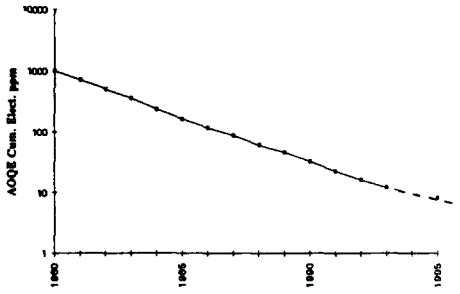


Fig 4. ASIC Reliability continuous improvement

Another feature of the ASIC concept outlined above is that the availability of fairly massive amounts of memory capacity, either on-chip or externally connected, enables comprehensive data collection of system parameters for the improvement of performance and reliability, via measured maintenance requirement or failure rate reduction. Such systems will require a large future investment in software protocols and tools.
In the latter part of the indicated third phase of automotive electronic evolution there will be further innovations in the areas of safety, environmental friendliness, performance and comfort. We expect to see:

- More common use of Airbags and other crash protection.
- Electronic suspension control.
- Radar distance measurement and warning.
- Navigational aids.
- Foolproof anti-theft devices.

And many other advanced features.
The electronic content of motor cars has increased by a factor of ten in the last ten years. Fig.5 shows that by the year 2000 high-end motor cars will embody up to $20,000 of their manufacturing costs in electronic systems and low-end cars over $6,000 Whereas 6% of the world sales of semiconductors (total value $ 4,8Bio) in 1993 went into motor cars, by the year 2000 automotive electronics will become one of the biggest single markets for semiconductors at greater than $25 Bio.

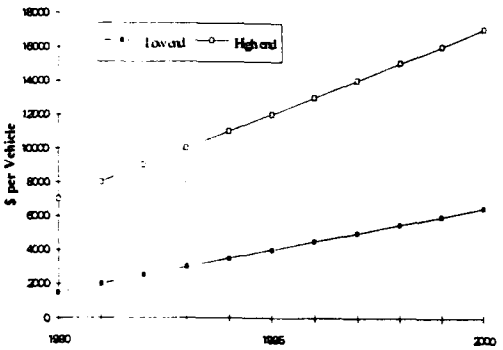


Fig. 5 Average Vehicle Semiconductor Content

3. ENVIRONMENT

The extreme physical environment encountered by automotive electronic components is shown in the table in Fig.6, from Ref.1, for the three main component locations. Engine compartment applications combine the most extreme combination of temperature range (-40 to +200°C), humidity (95% at 40°C), salt spray, mechanical shock (20g at 1 KHz), and thermal shock (100°C/min.); conditions which are not easily avoided because relocation would run contrary to the requirement to reduce the wiring harness. Generally todays ICs are being specified at -40°C to +150°C (junction temperature) which requires a lower environment temperature according to chip power dissipation, and means that system layout compromises must be made with regard to spacing and thermal insulation from the highest temperature sources. Technology developments, like SOI technology, may enable future ASIC generations to operate above 200°C, but interconnect and packaging technology developments must keep pace with the basic silicon technology to withstand this high temperature.
Apart from the locational subdivisions shown above, two other main groups are often defined:

- Devices which control vital functions (Ignition Control, ABS). These are often refered to as the "Vital Core".
- Devices for information, diagnostics and comfort (e.g. instrumentation, radio etc.)

Unfortunately, Murphy's Law decided that the vital functions must be located in the toughest environment, which has had an appropriate effect on the assembly technology, such that we will see in a later section how thick film hybrid techniques are often used for vital functions while the older and less reliable PCB technology is still applied to modules in the passenger compartment.

Position	Temperature. Range	Humidity	Shock/ Vibration
Engine			
• Choke	-40 to 200°C	95%(40°C)	20g,1KHz
• Carb/ Air Intake	-40 to 125°C	95%(40°C)	>100g
Chassis			
• Isolated	-40 to 85°C	98%(40°C)	2g, 2KHz
• Extreme	-40 to 175°C	98%(40°C)	>100g
Interior			
• Front/Dashboard	-40 to 85°C	98%(40°C)	1g, 20Hz
• Rear	-40 to 105°C	98%(40°C)	1g, 20 Hz

Fig.6 Extreme automotive environment

Design of ICs must also match the environment. ASICs designed specifically for automotive applications are designed for continuous operation up to 40V. Although normal car batteries supply 6 to 14V depending on their condition, this 40V requirement arises as a requirement for robust function against noise conditions which can give voltage spikes up to 150V. The electronic subsystem in which the IC must function is then protected from spikes by a zener clamp diode at 37V

In much the same way that an ASIC chip must be protected against ESD damage, it must also be protected at chip level against damage from noise spikes. As with ESD damage, noise induced damage can produce partial or latent damage which appears as a real defect or reliability hazard at a later date.

An important and completely different environmental condition is that created by competition in the market-place. Intense international competition in the last ten years has had a remarkably positive effect on improving quality and reliability of electronic components in the market environment. Abbreviations such as TQM, JIT, FMEA and QFD are today the stuff of newspaper articles, and the awareness of all involved managements that partnership and the closest cooperation between suppliers and customers lead to the best qualitative and financial results is alive and growing

Establishment of quality Management Systems, and obtaining ISO 9000 certification provide an organizational framework and incentive for financial support on which the technical work for improved Quality and Reliability can be hung.

Electronic component suppliers to the automotive industry have been among the first to obtain ISO 9000 approval. Previously the Quality Assurance department of a company assumed nearly complete responsibility, rubber stamped by a managing director, for the quality of the components or products delivered by the company. Quality was inspected and controlled into the product. Today ISO 9000 certification calls for and examines the active role of company management and staff in attaining high and continuously improving quality and hence reliability for its products. Quality and reliability must now be 'built-in' to a product as never before.

Changes in procedures and processes must be documented, and major changes reported to the customer for necessary requalification.

Partnership and cooperation between vendor and user today covers a much wider range than just the fabrication steps of a product. Total Quality Management (TQM) ensures that planning, design, development, fabrication, test and JIT delivery are more important than inspection in the QA department for the delivery of the product to the customer.

Quality Function Deployment (QFD) is used in the planning and set up phase of a project in an effort to see that no important element has been overlooked. Flow charts of product life cycle from concept through to delivery in high volume production are set up such that regular check points are established under an FMEA scheme to ensure that planned design and development steps have been properly performed on time, and foreseeable future errors are avoided. Regular audits are performed on all elements of the Quality System to ensure proper function of the system and to correct weaknesses.

Mechanisms for customer feedback with regard to performance and reliability are being installed. Not all components which fail in the field are today recovered for analysis by the manufacturer, but the situation is remarkably better than previously and steadily improving because of the pervading mentality which has accompanied the introduction of ISO 9000. Corrective measures enabled by the feedback of failed customer parts must be heavily supplemented by the failure analysis of components which failed normal production test, production or sample Burn-in or accelerated life test. Such analysis forms the single most effective path to product quality and reliability improvement.

4. AUTOMOTIVE SUB-SYSTEMS

A chain is as strong as its weakest link, so there seems little point in perfecting ASIC chip fabrication technology to sub-ppm reliability levels, at high cost,

if the surrounding sub-systems are reliable only to the many hundreds of ppm level. We must thus look at the interconnect technology in general and the chip packaging technology in particular.

4.1. Interconnection

Automotive electronic sub-systems have been traditionally fabricated using either solder connection of IC packages to printed circuit boards (PDIP or SOIC) or for reduced contacts ("critical core") application using wire bonded chips on thick film hybrids mounted on PCBs.

Examination of the production process flows for both processes reveals that the hybrid approach can be shorter than that of ICs on PCBs but because of material costs and available volumes it is more expensive. The hybrid approach undoubtedly improved realiability in the early stages of the development of automobile electronics, simply by replacing solder contacts with wires bonded to the hybrid substrate. There was thus a tendency to use hybrids for critical under-bonnet, so called "critical core" applications and standard ICs on PCBs in the passenger compartment. "The jury is still out" on this issue because the availability of large scale mixed signal ASIC chips may be replacing hybrids because wire bonding to the optimal metallization of ASIC chip and leadframe with much lower than 1 ppm failure rates may be even more reliable than wire bonds made to a somewhat rougher hybrid metallization. Also ASIC asembly tends to be more automated than hybrid assembly which tends to improve reliability. In either case the lowest number of solder contacts will be used because of their known lack of reliability relative to the wire bonds.

There are a number of characteristics of solder joints which are relevant to their performance and reliability:

-

- Microstructure of a solder joint depends on cooling rate from the molten condition, thus although joints may have the same physical shape they will be of different microstructure because they are connected to different thermal masses.
- Stress is caused at the solder joint because it is the meeting point of materials of very diverse coefficients of thermal expansion; moulding plastic, leadframe metal and epoxy (FR-4) PCB. Stress induced creep can result under typical lifetest conditions.
- Intermetallics are formed between the common constituents of solder; namely tin, copper and nickel; and intermetallic growth of several microns in thickness at modest operating temperatures (50°C) is reported to take place, accompanied by resistance changes, during 10 year operating life.
- Solderability of the initially solder plated (or dipped) leads is an important indication of future operating performance. It has been shown that poor solderability leads to high initial electrical resistance accompanied by abnormal increases on accelerated lifetest.
- Voids in the solder joints are not infrequent in the soldering process, and besides constituting a reliability risk they are mostly indetectable by optical inspection for screening.
- Temperature Cycling between -40°C and +125°C is known to be a difficult test for solder connections such that 100% failure rates have been reported in less than 1000 cycles.

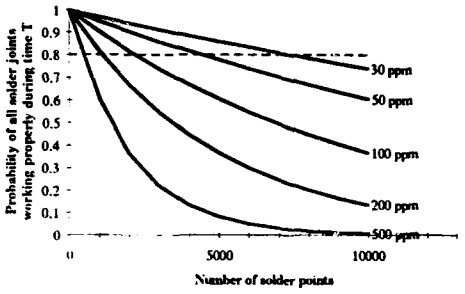


Fig. 7 Probability of complete function of solder joints over 17 year car lifetime

Armed with this detailed knowledge of the various failure mechanisms mentioned above engineers from Saab-Scania (Ref.2.) have shown, Fig.7, that less than 30ppm failure rates of solder connections in 17 years of operationg life would allow a 'vital core' electronic subsystem with 7000 solder joints to function without failure with a probability of 80%. This is the present state-of-the-art for solder connections.

4.2. Plastic Packages

Cost pressures dictate that plastic packaged components will be used as frequently as possible in automotive electronics within the limitations of the reliability constraints.

Formerly "High-Rel" components were always assembled in ceramic or cerdip packages and aluminium wire bonding was frequently used to prevent intermetalic formation on the aluminium chip pad. Such assemblies demonstrated failure rates over operating lifetime of less than 1ppm.

Plastic IC packaging was originally set up more than 20 years ago for the manufacture of low cost consumer products, which were not troubled by component failure rates of 200ppm. Over the years materials and methods for plastic assembly have improved such that failure rates of 1ppm over operating life are now conceivable although difficult to achieve. Fortunately automotive electronics is not the only electronics sector which desires low cost but high reliability plastic packages, and the experience of the

telecommunications sector can be passed on to good effect.

We will mention briefly here some of the recent improvements in methods, materials, analytical techniques and handling which have made quality and reliability improvement in plastic packages possible: -

- Methods such as the use of multi-plunger and top-gate mould tools have done much to reduce wire-sweep which can cause bond wires to touch one another and also chip edges. Voids and delmaination are also reduced. Designers of ICs are today more concious than ever before of the need to observe design rules for bond pad layout to facilitate reliable automated assembly.
Solder plating of leads has replaced solder dip coating for better uniformity control.

- Materials such as low expansion epoxy plastics with higher (165°C) glass transition point than previously (150°C) lead to less stress between chip, leadframe and plastic moulding. Even today, however, great care must be taken to minimize different stress created by the 'wave flow' solder process. Modern mould compounds contain significantly lower chlorine residues for reduced corrosion failures under temperature/humidity/bias conditions. Low cost oxidised copper leadframes have better thermal conductivity to remove heat dissipation and better adhesion to the mould compound than previously used Alloy 42 leadframes.

- Analytical Techniques such as the use of high resolution X-ray inspection of bond wires and postmould mechanical defects, together with the Scanning Acoustic Microscope (SAM) to check for poor adhesion and delamination of the plastic at the chip surface have enabled significant improvement in quality.

- Manual Handling of the components during assembly is the cause of many of the defects which effect reliability, such that replacement of present semi-automatic machines with complete automation would bring further progress for improved quality and reliability in the future. Careless manual handling during assembly can also lead to E.S.D. defects and soft E.S.D. defects, even when surface coated antistatic device tubes have been replaced by volume conducting tubes to protect known weakly ESD protected circuit I/Os.

This is but a brief summary of the interconnection and packaging methods being attempted for low cost but high reliability automotive electronic modules, but we hope it is sufficient to show that good progress has been made and further scope exists to enable yet higher reliability levels in the future.

5. COMPONENT RELIABILITY REQUIREMENTS.

Component reliability can be described in qualitative terms where the time dependence of failure rate (λ) follows the familiar bath tub shaped curve. The beginning of the curve shows the higher than average failure rates of so called 'infant mortality', the length of which is between minutes and 100's of hours depending on operating temperature. Such failures are generally related to chip package or board manufacturing defects. These defects can be eliminated or greatly reduced by better and tighter manufacturing process; by screening using Burn-in (under elevated temperature and dynamic electrical stress) to accelerate the failure mechanism, or by better understanding of the failure cause.
The first unstable zone is followed by a lengthy, more stable zone generally of many thousands hours (temp., bias and complexity dependent) where failure rates are low and random and generally caused by latent manufacturing imperfections. After many thousands to many tens of thousands of operating hours we enter a region of rising failure rate caused by so called 'wear-out' mechanisms, a typical example of which is Electromigration in chip level metalization. Although it is the guiding principle in automotive electronic reliability that the electronics should not fail in 30 years of life, experience shows that the actual operating life of the electronic components during this 30 years is probably less than 2000 hours, such that most automotive electronics never enter the wear-out mechanism region. Automotive electronics is much more plagued by the variety of overstress conditions which can occur starting with the 2000 to 4000 switch on/off cycles which occur during this short operational life. A list of known disturbance signals, fig.8, has been quantified under ISO 7637 and SAEJ1113 and is led in importance by the 'load dump' voltage spike, followed by the voltage stress condition caused by the back-e.m.f. from inductive loads. We shall see later how system design tries to deal with these voltage overstresses. We must also not forget that some of the 2000 switching cycles mentioned in total lifetime above can be accompanied by extreme temperature cycles (-40°C to 150°C); that 5000 hours of salt spray and high vibration/shock levels are possible during operation, and 150 hours of high level humidity under non-operating conditions could also be encountered, under extreme circumstances.
Basically the requirement of any High-Reliability system is to have 'Zero Failures' upon delivery of the component and the minimum number thereafter during normal operating lifetime. The 'upon delivery' requirement is more easily met, particularly if the infant mortality devices have been eliminated by Burn-in screening, and reasonable guard-banding of these tested electrical parameters is performed.

Pulse Type	Parameter (ISO DP 7637/1) V_s	$V_s > V_3$	t_r	t_d	t_T	t_2	t_3	t_f	R_i
1	to 100V	-	1µs	2ms	0.5-5s	200ms	-	-	10
2	to 100V	-	1µs	50µs	0.5-5s	200ms	-	-	10
3 Pos.	to 150V	-	5µs	100ns	100µs	10ms	90ms	-	50
3 Neg.	to 100V	-	5ns	100ns	100µs	10ms	90ms	-	10
4	4-7V	2.5	<10ms	-	100ms	<1ms	0.5-20s	<10ms	
	0.01	-6V							
5	4-7V	26.5 -86.5V	<10ms	40-400ms	-	-	-	-	0.5-4
6	to 300V	-	60µs	300µs	15s	-	-	-	30
7	0-80V	-	5-10ms	100µs	-	-	-	-	10

Fig.8 Voltage/timing of pulses for test standard (automotive)

We have seen previously in Fig.4 how the Average Outgoing Quality Estimator for cumulative electrical defects (AOQE) has been continuously improved by the automotive component suppliers. Between 1985 and 1993 the cumulative electrical AOQ has been reduced from 200 ppm to the region of 4-8 ppm. At the same time the AOQ for visual reject rate has also been reduced to 4 ppm. By 1995 it is required that both of these defect levels are defined as 'Zero-Defects'. However, the more difficult task is to eliminate random defects which occur in the field, where often the first year under guarantee is used to define the time period. We see below in Fig.9 best results from several suppliers for 1993 together with targets now being set for 1995: Strictly speaking, hybrids and PCBs should not be directly compared with ICs, but they are included for completeness.

	1993 Result	1995 Requirem.
(a) Cum. Electr. AOQ:		
- All components, subsystems	<4ppm	0ppm defects
(b) Failures in Field, first Year:		
- Passive Components, Discretes	<2ppm	<1ppm
- Standard ICs	<4ppm	<1ppm
- ASICs (Volume Production)	<8ppm	<2ppm
- Standard Hybrid	<10ppm	<2ppm
- Custom Hybrid	<10ppm	<5ppm
- PCB Subsystem	<50ppm	<30ppm

Fig.9 Ongoing Reliability Improvement.

Presently the 1995 target results can only be met using extensive Burn-in programmes, but most suppliers are well under way with a newer methodology to prevent the process related defects at source by extensive use of SPC and Wafer Level Reliability Testing. We shall address both of these points in detail in later sections, but they are mentioned here to illustrate the growing tendency to fix problems at source rather than screen out defects after manufacture. The same customer cooperative and open mentality can also be seen in the fact that the leading component suppliers now publish present and target quality standards, as example in Fig.10, which specify all aspects of the manufacturing process from incoming materials to shipped product and include:

- Lot Reject Rate, LRR (%), Incoming Materials.
- Defect Density (def./cm^2/level) during wafer manufacture.
- Quality Levels on Final Product (LPR, AIQ, AOQ-Electrical/Visual).
- Reliability/ Life-Test Results on sampled final product.
- SPC results for Waferfab and Back-End Assembly.

In the last 10 years the following progress has been demonstrated in these areas:

- Lot Reject Rates of both incoming material and material in process have been reduced from 10% to less than 1%.
- Defect Densities per layer in waferfab have been reduced from 1 def/cm^2 at 0.5µm particle size to below 0.1 def/cm^2 at 0.3µm particle size, not withstanding that there exists an exponential increase in particle number with reduced particle size for both air-borne and liquid-borne particles.
- 1000 hour (at T_{jmax}) Lifetest failures have been reduced from 1% (10k ppm) to 400 ppm and Temperature Humidity Bias (THB) at 85°C/85° R.H. have been reduced to about the same levels.

	1993	1995
1. MATERIAL QUALITY		
Incoming LRR (%)	< 0.8	< 0.5
2. PRODUCT AVERAGE QUALITY		
LRR, Electrical & Visual (%)	≤ 1.5	≤ 1.0
AIQ, Electrical (ppm)	≤ 70	≤ 50
AOQE, Electrical (ppm)	< 4	Zero
AOQE, Visual (ppm)	< 4	Zero
3. PRODUCT RELIABILITY (Failure rate)		
Short term (@ 300 hours)	0.3	0.03
Long term (FIT)	40	10
4. WAFER FABRICATION & ASSEMBLY		
Crit. param. Cpk ≥ 2	≥ 50%	≥ 75%
Crit. param. Cpk ≥ 1.33	≥ 85%	≥ 95%
Crit. param. Cpk < 1.33	≥ 2%	0%

Fig. 10 Quality & Reliability Standards 1993-1995

- Statistical Process Control, not applied 10 years ago, has been introduced and C_{pk} values have been continually improved from 1.0 to greater than 2.0, over the critical definition points 1.33 (3 sigma) and 1.66 (5 sigma). Normally 1994 waferfabs have 100% of all critical process steps at $C_{pk} > 1.33$ and 60% with $C_{pk} > 2.0$. Similar values are reported for machine capability, C_{mk}. The SGS-Thomson SURE 6 handbook (Ref.3) is a good example of the wealth of reliability data available from IC vendors.

The reliability requirement and method outlined above are under continuous review, both internally in component manufacturers, and also by means of quality audits performed by the automotive customers. By setting new targets on a yearly basis, continuous improvement is guaranteed.

Under the title 'Qualification', the automotive customer today has a relatively broad view of what he expects from a product, and the purpose of qualification can be stated as:

- Safeguarding the Quality and Reliability of the product through all steps; Concept, Design Development, Process Development, Production and Testing.
- Establishing and mutually agreeing Quality and Screening Levels and Procedures, especially during the ramp-up phase from development into full production.
- Establishing the characteristics and application features of the component especially with regard to its suitability for the planned application. It is also essential to characterize a good number (3-5) of process batches for this last exercise.

6. AUTOMOTIVE TEST/SCREENING/BURN-IN

Testing of automotive components in plastic packages is a lengthy and costly process. Whereas consumer ICs can use extrapolation methods to guarantee specification across the operating temperature range, automotive ASICs are tested at -40°C, 25°C and 85°C before being burned-in and then retested at 25°C.
Most IC manufacturers are working towards reduction or elimination of Burn-in while simultaneously improving both quality and reliability of the products.
Test is that sorting procedure which establishes if an electronic component functions (or not) according to specification at a fixed point in time. Screening by Burn-in stimulates a component with the objective of making testable or visible those latent failures which become infant mortality in the application and which were not recognisable at the original test. Screening can involve a number of pre-stress conditions such as high temperature storage, Burn-in, temp.cycle, temp. shock, vibration, centrifuge, or mechanical shock, or it can also include visual or X-ray inspection.
Burn-in/Screening can be used in the combination of pre-conditioning and test to elliminate those devices which would otherwise have failed early in life. Typically Burn-in is conducted under dynamic conditions (i.e. the device is exercised with a test pattern) at 125°C for 48hours (MIL-STD-883 or CECC 90000 standards) before a final electrical test is performed.
Failure analysis, failure documentation, feedback and corrective action are an essential part of the procedure, in a quest for continuous product improvement.

We have seen in our definition of component reliability requirements that ASIC failure rates in the field can presently only be brought into the 1ppm level with the aid of Burn-in, but Burn-in is a costly and time consuming process. We want therefore to examine the Burn-in process and to see if we can find sensible measures by which it can either be reduced or eliminated. Without degrading product quality and reliability.

Since our aim is to fix the problem at source we should ask how the product quality is related to the manufacturing process. In general we can observe that both yield and quality problems are most pronounced either at the start-up phase of a new manufacturing process, or when a new device or application is run on an existing technology. It is reasonable to assume that in the former case improved AOQ can be achieved by optimising the manufacturing process using SPC methods, and in the latter case we have either insufficient test cover, electrical guardbanding, or weak design. The latter can of course be much improved by increased use of Design for Testability and Built-in Self Test.
If the problem is a reliability problem of infant mortality or lifetest failures, we should first look for a design or technological weakness, fix this by redesign or process change and check improvement by performing 48hr Burn-in. If we suspect a fabrication problem, both Burn-in and improvement of the manufacturing process must be performed.

Before we try to replace Burn-in by other methods we should first examine its advantages and disadvantages:

Advantages:

- Infant mortality failures can be reduced (Established method).
- It can be performed independently of manufacturer by device user.
- It can be reduced or eliminated when quality improves with improved or prolonged manufacturing.
- 100% Screening of small quantities of components improves quality.

- Indications of reliabiiity in the field is gained from Burn-in results.

Disadvantages:

- Burn-in/Screening cannot solve all reliability-problems.
- It is costly to perform.
- A longer manufacturing chain results.
- Extra and incorrect handling in test can lead to yield loss, latent ESD failures or bent lead, coplanarity, soldering problems (SOICs).

Since Burn-in is such a time consuming and costly business, much thought is being given to forcing the learning curve such that 100% Burn-in can quickly give way to sample Burn-in and eventually Burn-in can be eliminated completely. Avenues of approach are:-

- Optimise the whole frontend and backend manufacturing processes using SPC methods, including the part process Burn-in.
- Analysis of both Burn-in failures and field failures.
- Tracing failure causes to the real origin in either waferfab, assembly, or test area and taking appropriate corrective action. It is not sufficient merely to quote the fail binning from electrical test.
- Developing a strategy to eliminate Burn-in altogether.

Generally a close cooperation between IC manufacturer and user is essential to developing both a beneficial strategy and a positive "mind set". On no account must the user allow the supplier to simply replace defective parts.

An effective strategy involves replacing Burn-in investment with investment in quality and reliablity methods in waferfab and test (e.g. SPC and Wafer level Reliability Methods). It must also be realised that quality costs reduce with increased volume, so a user is unwise to split volume between too many suppliers.

A typical strategy for the start-up phase of a new ASIC device is as follows:-

- The user bears some part of the extra cost involved in obtaining higher defined and guaranteed quality in the start-up phase only. The user does not pay for extra Burn-in, nor does he dictate which tools or methods are used.
- The ASIC manufacturer invests in failure analysis and feedback of results to continuously improve quality, such that the number of batches requiring Burn-in is minimized.
- The target for Burn-in failure rate is $< 0.3\%$.
- A quality plan with the following features is installed: -

Every batch is sampled to LTPD 0.3, and batches not reaching this quality are 100% burned-in and tested.

The first few batches of a new ASIC product are in any case completely burned-in.

Subesequent batches receive 100% burn-in on a reducing proportion of the whole batch.

When a low and stable failure rate from burn-in is achieved and sampling plan of LTPD 2 is installed, purely as a monitor.

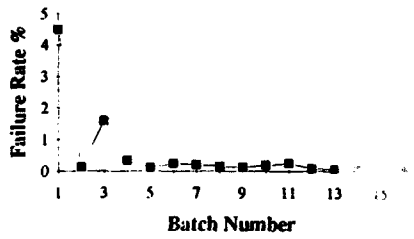


Fig. 11 Failure rate improvement by systematic Burn in

Fig.11 shows an example of the improvement of a typical automotive product subjected to the scheme outlined above.

7. IN PROCESS RELIABILITY

It is modern practice in qualifying a new technology for practically any new application, to examine and qualify the individual critical process steps. The same methodology is then employed for continuous process monitoring and improvement.

Main elements for examination are: -

- Active Gate Oxide Quality.
- Hot Carrier Injection Effects.
- Electromigration in Metalization.
- Other Wafer Level Reliability Measures.
- E.S.D. Protection.

This is but a short list of the many factors which are actually examined to create and maintain reliability. Many of the tests can be performed at wafer level without encapsulating the test device.

Gate Oxide Quality is generally tested on a statistical basis by applying electric field stress to gate oxide capacitors in the form of a time dependent voltage ramp of 1MV/cm.sec., and measuring the distribution of electric fields at which the capacitors break down. For silicon dioxide the breakdown field , E_{bd} peaks at about 12MV/cm, but apart from this we can establish from the test the number and distribution of defects causing early breakdown and make extrapolation of the Time Dependent Dielectric Breakdown (TDDB). TDDB which is used to give an indication of the long term reliablity of the device, is extracted from the E_{bd} distribution using the well established Berman method.

Correlation is made between TDDB and lifetest results on final devices at elevated temperature, see Fig.12, such that a method is established for the fast and routine checking of potential reliability problems, and a route is created for monitoring continuous improvement of the process.

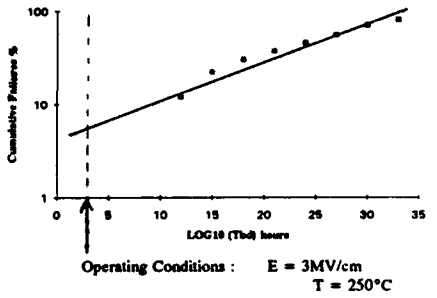


Fig. 12 Correlation of TDDB with lifetest failures (Berman method)

We have described here one of the tests performed, but in practice this is supplemented by Current Stepped Charge to Breakdown (Qbd) and Constant Voltage tests. The present status in VLSI technology is that oxide quality of 10nm gate oxides for 0.5μm CMOS can guarantee greater than 15 year lifetime at 5V and 10 FIT at 150°C.

Hot Carrier Injection into the gate oxide of sub-micron channel length ULSI transistors can cause change and degradation of the transistor parameters leading to reliability problems. Because of the higher mobility of electrons compared with holes this is primarily an N-channel transistor effect, although degradation of very short channel P-channel can also be measured.

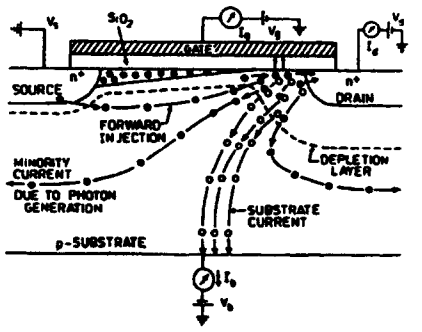


Fig 13 Hot carrier effects in N-channel transistor

Fig.13 illustrates the situation where exponential increase of lateral electric field in the depletion region at the drain corner initially imparts sufficient energy (ca. 1.5 eV) for these so called 'hot electrons' to cause impact ionization or the creation of electron/hole pairs which can be detected by the onset of the substrate current. If the carriers gain 3-4 eV from the field or by collision, they can be injected into the gate oxide above the channel region to degrade the transistor parameters in a number of ways. Changes occur in the threshold voltage, V_t, and degradation in current drive capability occurs in both linear (I_{lin}) and saturation (I_{Dsat}) regions of transistor operations. Furthermore channel shortening effects from the injected charge can change both transistor transconductance and sub-threshold slope. Transistors are designed with spacers and processed with low doped drain regions (LDD) to minimize hot carrier degradation and reliability studies are performed to relate measured transistor parameters under normal and stress conditions to the operating lifetime. Generally maximum substrate curren, I_b (μA/μm width), which occurs at about $V_G=V_D/2$, indicates the maximum stress condition which can be correlated with operating lifetime, the limit of which is often defined by a 10% change in I_{Dsat}. Using the method and model of Takeda, operating lifetime can be extrapolated from the measurements taken under stress. Fig.14 shows typical Takeda plot for 0.7μm N- and P-channel transistors. Properly designed and fabricated sub-micron transistors have expected lifetimes well in excess of 10 years, and we should mention that the effects described are only weakly temperature dependent.

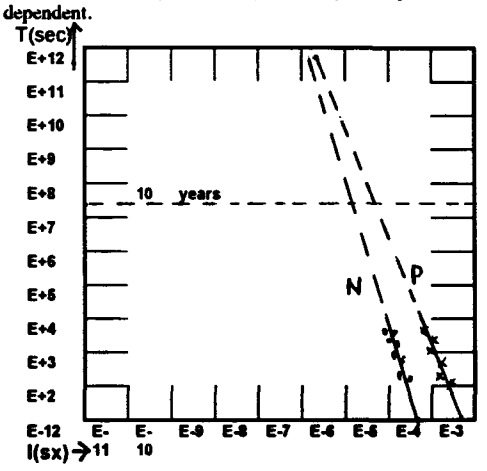


Fig 14 Hot carrier lifetime degradation correlated to substrate current (Isx) (Takeda)

Much interesting work is being conducted in this area to explain such effects as the deleterious effects of multi-level metalization and worse degradation under AC rather than DC operation of circuits.

Electromigration Effects in IC Metalization, have been the subject of intensive study since the 1960's. Basically when current flows at elevated temperature (>150°C) in thin films of metal (Aluminium) with grain structure, material transport which occurs at an atomic level can create voids and breaks (open circuits) in the metal tracks. The effect called 'Electromigration' is strongly dependent on current density, temperature, film stress and composition. Black's famous equation for Mean Time to Fail:

$$MTF = A\ J^{n}\ Exp\ (E_a/kT)$$

allows us to establish to coefficients A, n and E_a experimentally and to some degree to correlate expected performance (MTF) in the field with the results of accelerated lifetest performance at elevated temperature.

Automotive electronics represents 'life-in-the-fast-lane' because IC surface temperatures of 150°C, which were formerly considered appropriate for accelerated lifetest, are now encountered in underbonnet automotive applications. Normally aluminium metalization in ULSI technology design rules is specified for a current density of 0.2E6 A/cm^2 at 100°C to give an operating life (defined by 10% resistance change) in excess of 30 years (for 10 FIT).

Fast test methods for electromigration like BEM (Breakdown Energy of Metals) and SWEAT (Standard Waferlevel E-M Acceleration Test) are applying current densities of 1.0 to 4.0 E6 A/cm^2 at 200 and 250°C, to accelerate MTF from years to minutes and hours. Because of Joule heating effects and the difficulty in estimating the actual temperature it is not easy to extrapolate real life performance from the fast test methods, even when the Black coefficients E_a and n have been established previously by normal accelerated lifetests, but the fast test methods have the advantage that they can be used at waferlevel, sometimes as in-process monitors to check the consistency of the metalization fabrication process. Recent fast test results would suggest that current densities as high as 0.8 E6 A/cm^2 at 150°C operation for 10 year MTF (at 10 FIT) are possible for Al/1%-silicon metalization sandwiched between TiN barrier at bottom and anti-reflection layer at top.

Some interesting facts about aluminium metalization performance are given below:

- Passivated metal lines (real situation) show an order of magnitude better EM properties than unpassivated ones, which thus should not be used for lifetest, SWEAT or BEM.
- Because of 'Bamboo Effect' narrow (1.5-2.5µm) metal lines show 4 times better EM performance than 4 to 8µm lines. The latter should be used for lifetests.
- 0.5% Copper in Alu. or Alu/Si metal lines improve EM resistance by an order of magnitude, although detractors discuss longterm corrosion failures with the copper inclusion.
- Stress in metal lines causes both 'hillocks' and poor EM performance, so the former is an indication of the latter.
- Titanium nitride, TiN, used in metalization layering has excellent EM hardness.

Tests for life performance of contacts and vias follows similar lines to the details given above. Great care is taken that metalization layers adequately contour contact/ via edges and that barrier layers to prevent 'spiking' of underlying junctions remain effective.

Other Wafer Level Reliability (WLR) Measures are too numerous to be covered adequately in a review paper, and although the topics covered so far in this section can fall into the category Wafer Level Reliability it is useful to look at some of the advantages of WLR method since they are becoming increasingly practiced.

- WLR encourages built-in reliability as a replacement for the more traditional tested-in methods. It is possible that this is the only way to achieve 10 FIT (0.001% per 1000 Hr.) reliability levels.
- WLR is cost effective by aiming at short, typically 30 sec., test times. And since it is less costly it can be more widely and more uniformly applied for uniformly high reliability levels.
- Real time measurement allows a short feedback loop for corrective measures, compared with the weeks and months of lifetest on finished parts.
- Defective material can be removed from process immediately to save further processing costs.
- Wafer pattern related problems can be identified before the pattern disappears in assembly.
- Practically all design rule and wafer process steps can be examined with the use of suitable patterns.
- Local heating and temperature measurement can be performed so that many hundreds of temperature cycles can be performed without preconditioning neighboring devices to be subsequently measured.

Electrostatic Damage Effects (ESD) or protection capability are not generally measured at waferlevel, but since handling in both waferfab and assembly can cause problems they are mentioned in this section. In wafer fabrication silicon wafers can become charged by the movement of dry air in spinning rinser-dryers or in laminar flow cabinets. In both cases air ionizing protection devices are usually provided, but these must be regularly checked for functionality. Discharge sparking between wafers and with the carrier can cause latent damage to both gate oxides and junctions.

In assembly, and particularly after trim-form singulation ICs are particularly susceptible to ESD damage. Stringent grounding of machines, use of grounded operator wrist straps, and the use of anti-static and volume conducting carrier tubes are some of the measures taken to prevent ESD damage, but there are several weaknesses in the system which often depends to a high degree on manual handling and thus operator discipline. For example it is advised by the suppliers of surface coated anti-static carrier tubes that

such tubes become ineffective against ESD after one time and maximum double handling operations, but in Far-East subcontract assemblers the superior volume conducting tubes are hardly in evidence.

The greatest reliability problem with ESD at wafer level or assembled part level is the so called 'Walking Wounded', or those devices which have been only partially damaged by ESD, partially healed at room temperature to avoid detection at electrical test, but able to reappear as defective devices in the application; even in the ppm level they are sufficient to ruin reliability records, statistics and reputations.

ESD protection circuits are built in to CMOS ASIC circuit I/Os to protect devices against improper handling. MIL STD. 883C Method 3015 describes characterization using the 'human-body-model' (HBM) which discharges a 100pF capacitor rapidly through a 1.5KOhm resistor at the device pin. Proper testing demands inputs to be tested with respect to other circuit pins, to ground and power rail. Most suppliers guarantee a conservative test voltage of 2.5KV for this test, but it is generally known that good protection can withstand 4KV, and special devices are now appearing on the market with 10KV claimed protection. It may however be more interesting in the cause of continued inprovement, in determining the ESD hardness of a device, to use the combination 200pF, 200 Ohms in the HBM.

8. FUTURE DIRECTIONS

The Prometheus study of 1988 (Ref.4) predicted the following picture for automotive electronic system in 1995:

- 100 Intelligent Sensors
- 80 Intelligent Actuators
- 45 Electric Motors
- 5 Displays
- 1000 ICs (Equivalent 1987 Complexity)
- Operation at 100-500 MHz for Real time Communication

Six years later, and one year from the target, we can report that at least in high end motor cars the number of predicted ICs (equivalent) and electric motors will be met, but that sensors and actuators will be less intelligent than expected, displays less in number, and maximum operating frequency at 200 MHz (fastest microprocessor).

Looking at the present situation and its extension into the future we can say that reliability considerations will surely dominate the form and complexity of future automotive electronic systems. Because of the desirability to reduce wiring and connectors, the multiplexed one or two wire BUS System will gain popularity. This in turn will encourage the use of Smart Power ASIC systems-on-chip with integrated sensor and actuator interface and with local inteligence provided by a microprocessor core. Certainly the combination of ASICs and microprocessors will dominate. The reduced number of contacts will enable those that are still being used to cost more and be of improved reliability.

Solder contacts have improved remarkably in recent years, such that some 'Motronic' and 'Tempomat' applications use PCBs with a limited number of solder contacts, and further improvement in reliability will certainly be achieved, but the multiplexed Bus System shows the way, in many parts of the motor car, to eliminate solder contact completely. This will be achieved by using a metal frame which acts simultaneously as pins to the connector and as leadframe for wire bonded silicon chip assembly. This simple form of 'mechatronics' is already in development.

Looking to the semiconductor components themselves it is clear that structural dimensions will continue to decrease with rates predicted by 'Moore`s Law', and complexity will increase beyond todays 2 million transistors per IC. Quarter micron structures are already available as development samples. This high complexity level encourages the use of more RAM and EEPROM memory which can be used to store data on system performance for preventive and corrective maintenance and ensure greater reliability for the vehicle.

Although built-in self test (BIST) and design for testability (DfT) are common features of todays ICs, use of these features and improvement in test coverage will increase still further. It is also possible to orient testing philosophy much more towards failure analysis and reliability improvements. Intelligent testing of abnormal parameter distributions would fail devices from the abnormal distribution even though they officially meet the parameter specification. Certainly leakage current and standby current distributions could be screened in this way. Analysis of the reverse bias characteristics of ICs (I_{DS} Analysis) will be used to fingerprint weak devices. It is also possible that irregularities in delay time distributions could be used to 'weed out' weaknesses which control device and process parasitics.

There will be an overwhelming trend to Wafer Level reliability studies in waferfabs as a means of cost effectively building-in quality and reliability and reducing the amount of costly burn-in and routine lifetesting on finished parts. Where SPC is used today to show that manufacturing processes are under the the tightest control, this will increase to cover more parameters better, and WLR will be used to detect potential instabilities and drift, reducing and possibly replacing burn-in at the end of the process.

Communication and 'openess' between vendor and customer will increase so that each understands better the requirements of the other, and 'open-book' will be practiced with respect to the customer access in-fab. SPC and WLR results as well as reliability and failure analysis details. Vendors will be expected to predict failure rates in the field.

Such changes will not come about overnight, and will probably not be forced by contracts, but rather will be the result of the realisation between vendor and customer that this is the only way to achieve and maintain 'zero-defect' delivered quality, sub 1ppm failure rates in the life of the application and less than 10 FIT.

9. REFERENCES

1. R.Dell'Acqua and F.Forlani (Marelli SpA): "Automotive Electronics: Where Are We Going?", Hybrid Circuits, No.15, Jan.88, pp 8-12.
2. H. Danielson (Saab-Scania): "Will ASIC Technology Demand a New Interconnection Technology instead of Soldering in Automotive Electronics?", Hybrid Circuits, No.19, May 1989, pp 32-37.
3. SURE 6 - The SGS-Thomson Microelectronics Quality and Reliability Program - Handbook (July1991).
4. Prometheus 'Pro-Chip White Book' (1988).

10. ACKNOWLEDGEMENTS

The author wishes to thank members of the automotive group of the ESPRIT project 'STATIC', No.6485, for encouragement to write this review and especially Ms. Finetti in the CEC ESPRIT office. ELMOS colleagues J. Smith and S. Seliger together with R. Maier, BMW Quality Department, provided valuable information for the content.

IN-SITU FAILURE DETECTION IN THICK FILM MULTILAYER SYSTEMS

J. Manca[a], L. De Schepper[a,b], W. De Ceuninck[a,b], M. D'Olieslaeger[a] and L.M. Stals[a]
[a]Limburgs Universitair Centrum, Institute for Materials Research, Materials Physics Division, Wetenschapspark 1, B-3590 Diepenbeek (Belgium)
[b]DESTIN N.V., Wetenschapspark, B-3590 Diepenbeek (Belgium)

ABSTRACT

By means of in-situ e.m.f.-measurements, leakage current measurements and impedance spectroscopy, it has been possible for the first time to detect spontaneous and forced blistering in thick film multilayers during formation at high temperatures. Also the occurrence of high temperature shorts in Ag-dielectric-Ag multilayers under DC-bias resulted detectable.

1. INTRODUCTION

Important failure mechanisms in thick film multilayer hybrids are shorting and blistering of the dielectric between two metal layers. Spontaneous blistering has been observed in mixed metallurgy multilayers, i.e. multilayers including different metals, e.g. Au and Ag (Refs.1-3). High temperature shorting can occur spontaneously in multilayers with Ag-electrodes (Ref.4). To build a complex mixed metallurgy multilayer circuit succesfully, a very robust dielectric material, able to withstand multiple (up to 30) refires to 850°C without occurrence of short circuits or blisters, is required.

The basic physico-chemical processes causing these failures are insufficiently understood. In particular, the correlation between physical and chemical properties of the dielectric material and the occurrence of these failures is not clear.

An important step towards a better understanding of the driving mechanisms is to mark out the critical temperature and time region for possible failure formation. Since failures such as shorting and dielectric breakdown give rise to a sudden decrease of the electrical resistance they can be characterized by a well defined time to failure. For blistering the situation is probably less simple, but still it should be possible to indicate by experimental means the time region of blister formation.

In this paper some experimental techniques are presented that allow the detection of blisters and shorts during formation at elevated temperatures.

2. IN-SITU ELECTRICAL MEASUREMENTS

A new approach to the study of thick film failures is given by the in-situ electrical measurement technique. With this technique it is possible to monitor continuously the electrical behaviour of thick film test structures during thermal treatments of choice. This approach is justified by the facts that the failures occur during the thermal treatments and that they are related to electrical parameters such as resistance (shorting => zero resistance ; blister => local infinite resistance).

A significant result of this approach is the observation of a close correlation between the leakage current measured through the dielectric at elevated temperatures, and the ability of the dielectric to resist shorting and blistering effects during the preparation of circuits (Ref.3). The materials with the lowest electrical resistance are also the materials with the largest probability to blister when used in a mixed metallurgy multilayer, and with the largest probability to give shorts when used in multilayers with Ag-electrodes. The in-situ leakage current measurement technique has been shown to be a rapid evaluation method for these materials.

Since the in-situ measurement technique allows to measure continuously the evolution of an electrical property of interest, it should also be possible to determine the moment or time region for failure formation. As will be reported further on, this has been achieved with in-situ e.m.f. measurements, leakage current measurements and impedance spectroscopy.

3. TEST STRUCTURES

In-situ electrical measurements require test structures with two (leakage current;e.m.f) or four (impedance spectroscopy) contact terminals. Parallel plate capacitor test structures have been designed with overlap areas (A) of 25, 50, 75 and

100 mm². The multilayers consisted of 96% Al_2O_3 substrates; Au-, Ag- and AgPd-electrode material and a commercial dielectric with low blister resistance.

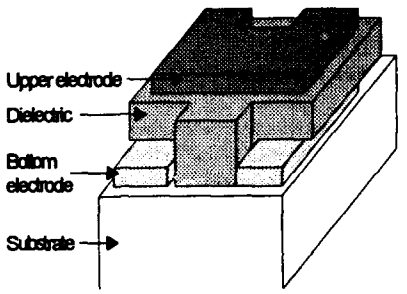


Figure 1 : *Lay-out of the studied thick film multilayer system*

4. SPONTANEOUS BLISTERING

With in-situ e.m.f.'s meauserements it has been shown that the mixed metallurgy system Au(bottom)-dielectric-Ag(top) acts at 850°C as a spontaneous battery, and the battery voltage (i.e. the spontaneous electromotive force) was measured (Ref.3). During the heating step to 850°C with a rate of 10°C/min, the magnitude of the e.m.f increased continuously. At the peak temperature a battery voltage up to 200 mV between the two metal layers was observed. In systems with an unmixed metallurgy, Ag-dielectric-Ag or Au-dielectric-Au, much smaller e.m.f.-values have been measured at the peak temperature.

During the applied temperature treatment, blisters were only formed on the test structures with a Au(bottom)-dielectric-Ag(top) configuration. For these test structures a corresponding voltage drop has been observed after annealing for a few minutes at 850°C.

With a measurement system that allows to inspect electrically and visually a test structure of this type during a given thermal treatment, it has been found that the moment of maximum voltage, i.e. the moment previous to the voltage drop, corresponds to the moment where the first blister can be seen. During further annealing new blisters have been observed, while the e.m.f. continuously decreased.

From the in-situ voltage measurements it could thus be concluded that the Au-dielectric-Ag configuration behaves at 850°C as a spontaneous battery with a positive Au-electrode, a negative Ag-electrode and the dielectric as electrolyte (see figure 3). This behaviour can be qualitatively understood by looking to the electrochemical series listing standard reduction half-reactions in order of increasing standard potential. For the case of Ag and Au the corresponding half-reactions and standard potentials ($\mathcal{E}^\circ$) are given by (Ref.5) :

$Ag^+(aq) + e^- \Rightarrow Ag(s)$ $\quad \mathcal{E}^\circ = 0.7996$ V
$Au^{3+}(aq) + 3\ e^- \Rightarrow Au(s)$ $\quad \mathcal{E}^\circ = 1.42$ V

Since the Ag^+/Ag half-reaction has a smaller standard potential than the Au^{3+}/Au half-reaction, it has a greater tendency to reverse. The galvanic cell obtained with these two half-cells thus results in a positive Au-electrode and a negative Ag-electrode. Quantitative information is much more difficult to obtain. In principle it is possible to calculate the potential difference between the electrodes of a galvanic cell as a function of temperature by using the Nernst equation. However, a practical constraint in doing so is the lack of available information concerning the electrochemical activities of dissolved Ag^+ and Au^{3+} ions.

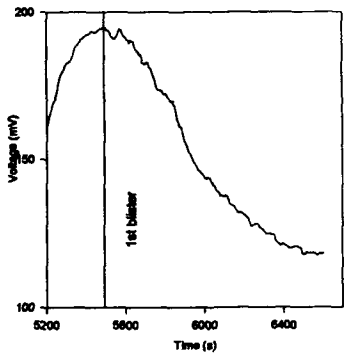


Figure 2 : *E.m.f.-measurement at 870°C for a Au(bottom)-dielectric-Ag(top) multilayer.*

A first feature of figure 2, the occurrence of a meaningful potential difference between the electrodes of the Au-dielectric-Ag multilayer can thus be attributed to a difference in electrochemical behaviour of the Au^{3+}/Au and Ag^+/Ag half-cells. The potential drop observed in figure 2 can also be explained qualitatively by starting from the relatively similar behaviour often encountered in common batteries such as dry cells.

A drop in emf is known to occur in many galvanic cells during operation at room temperature due to a number of effects in which the chemical conditions around the electrodes are modified. In electrochemistry this phenomenon has been termed polarization, not to be confused with the dielectric polarization known in electromagnetism. The simplest kind of polarization occurs when the ionic concentration around an electrode is altered by the passage of current. One of the most common cases where polarization seriously limits the operation of the cell, is encountered when hydrogen ions are discharged, forming bubbles of gaseous hydrogen. One effect of the hydrogen gas is to increase the resistance between the electrode and the solution, increasing the internal resistance of the cell. In some types of cells, a substance called a depolarizer is added to the electrolyte which reacts with the hydrogen to prevent the formation of free gas (Ref.6).

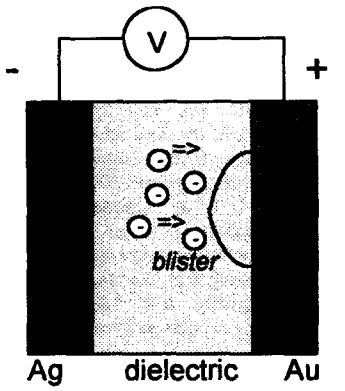


Figure 3 : *Spontaneous blistering in a mixed metallurgy mutlilayer.*

It has to be pointed out that the described polarization phenomenon occurs during operation of the cell and thus while the electrodes are electrically connected. For the high temperature spontaneous blistering however no external connections are present between the electrodes involving a polarization mechanism of an other type.

We believe the occurring polarization to originate from a spontaneous electrolysis. From the measured potential difference and the small spacing between the electrodes a relatively high electric field arises in the dielectric bulk material. Since at the given elevated temperatures the dielectric aquires a high conductivity, one can presume that the electric field inside tl.. dielectric leads to the directional motion of mobile, charged species that discharge on the electrodes. The observed voltage drop is thus interpreted in this context as an electrical discharge.

5. FORCED BLISTERING

When an external voltage of 300 mV is applied to unmixed metallurgy systems (Ag-Ag; Au-Au) during a temperature profile up to 850°C, blisters are also formed (Ref.3). This shows unambiguously that blistering is a voltage driven effect.

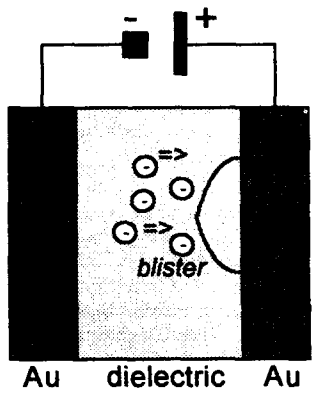


Figure 4 : *Forced blistering in a Au-dielectric-Au multilayer.*

During blistering electrode material delaminates from the dielectric causing an important increase of the internal electrical resistance. The sudden increase of electrical resistance during a thermal treatment with an external voltage stress therefore determines the moment of the blister formation. A sudden increase in resistance has been observed both with in-situ leakage current measurements and in-situ impedance spectroscopy.

Figure 5 shows the result of a voltage ramp on a multilayer structure at 850°C. It can be seen that for voltage stresses below 200 mV the resistance remains rather constant while higher stresses give rise to an important increase of the electrical resistance. In-situ optical inspection of the given test structure confirmed that blistering starts from the moment that the magnitude of the applied voltage exceeds 200 mV with the corresponding increase of the measured resistance.

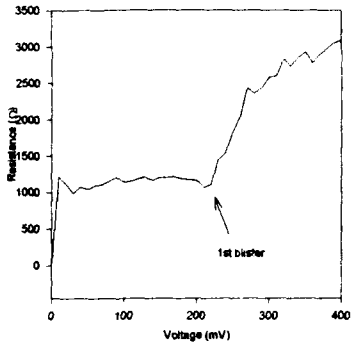


Figure 5 : *Resistance for a Ag-dielectric-Ag multilayer at 850°C submitted to a voltage ramp.*

Also impedance spectroscopy measurements have been performed for a multilayer heated from room temperature to 850°C with and without a permanent DC-bias of 300 mV in order to detect the eventual blister formation. During the thermal treatment rapid frequency sweeps have been continuously performed. Each sweep consisted of 40 equidistant frequencies on logarithmic scale between 100 Hz and 10 MHz. In the high temperature region of the stressed test structure an anomalous increase of the impedance can be observed, indicating the blister formation. This increase is not present in the results for the unstressed material.

6. HIGH TEMPERATURE SHORTING

The occurrence of shorting in thick film multilayer and crossover systems has been extensively studied in the temperature region between room temperature and 200°C. A high number of papers have been published concerning especially the formation of Ag-shorts in this temperature region mainly attributed to an electrochemical migration mechanism. It is in fact known that in uncovered conductor-dielectric systems confined by a humid environment, often conductive pathways are built up due to a transport of metal ions from an anodic site through an aqeous electrolyte to a cathodic site where electrodeposition occurs, forming a dendrite or dendrite-like deposit (Ref.7).

High temperature, electrical shorting has been treaten by Needes et al (Ref.4). They observed that the manufacturing yield of silver multilayer circuits can be adversely influenced by the development of interlayer electrical shorts due to migration of silver ions during firing.
Silver diffuses into the dielectric from both the top and bottom conductors, whereas it is observed that silver migration (shorting) originates from the bottom conductor. A distinction is drawn here between diffusion which results from differences in chemical potential, and migration (the fundamental mechanism underlying the current phenomenon) which in its basic form, is diffusion promoted by the presence of an electric field. During firing and subsequent refiring of a crossover, a potential difference is developed between the top and bottom conductor layers, leading the crossover to fail.
According to Hang (Ref.8) this conductive ion migration is enhanced by the presence of a stable liquid glass phase in the dielectric microstructure during subsequent refirings. The reliability is therefore largely determined by the remanent vitreous glass in each dielectric.

In this study the electrical conductivity has been measured in-situ, not only for the previously discussed dielectric, but for a set of dielectrics with a different tendency to form blisters and shorts. For the dielectrics with the lowest (D2) and the highest conductivity (D7) at elevated temperatures, Ag-diffusion profiles have been measured in a scanning electron microscope with EDAX-analysis It has been found that the dielectrics with the highest conductivity also allow the highest Ag-diffusion into the bulk as shown in figure 6.

Figure 6 : *Ag-diffusion profiles into dielectrics D7 (upper curve) and D2 (lower curve). (Full screen : 100 weight% Ag)*

Since there is a direct relation between conductivity and diffusion coefficient (Nernst-Einstein relation) for the mobile species that participate in the conduction process, this result can easily be understood.

As illustrated by the former diffusion profiles an important amount of Ag diffuses and migrates into the dielectric bulk. The kinetics of the Ag-migration can be studied in-situ by measuring the leakage current during a continuously applied voltage stress and determing the time to failure. The criterium for shorting is the sudden increase of leakage current exceeding the value of 2 mA, the upper limit of the used picoammeter.

For dielectrics D7 and D2 a Ag-dielectric-Ag test structure has been heated from room temperature up to 850°C at a rate of 10°C/min under a DC-bias of 100 mV. It can be seen from figure 7 that in the high temperature region a clear current increase occurs due to the formation of a short for dielectric D7, while dielectric D2 only shows a much smaller leakage current.

With the presented measurement method it is therefore possible to perform additional measurements in order to investigate closer the role of parameters such as dielectric thickness, dielectric composition, temperature and applied voltage in the formation of high temperature shorts.

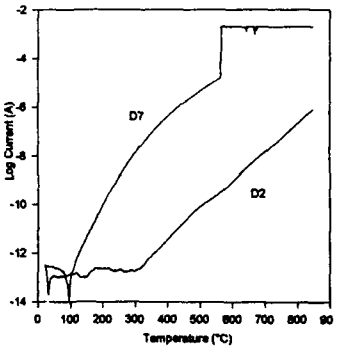


Figure 7 : *Leakage current for two Ag-dielectric-Ag multilayers with different dielectric composition continuously stressed by a 100 mV DC-bias.*

7. CONCLUSIONS

With a measurement system that allows to inspect electrically and visually a mixed metallurgy test structure blistering during a given thermal treatment, it has been found that the start of blister formation is accompanied by the occurrence of an important voltage drop between the electrodes. During further annealing new blisters have been observed, while the e.m.f. continuously decreased. The formation of forced blisters can be detected in-situ by measuring the corresponding increase in electrical resistance with leakage-current measurements or impedance spectroscopy.
The discussed in-situ electrical measurement methods also allow to observe the high temperature formation of shorts.

REFERENCES

1. Gilles T., Reynolds Q. and Steinberg J.,'A New Generation Multilayer System : Low Cost and High Reliability', Hybrid Circuits, No. 21, January 1990.

2. C.R. Pickering, W.A. Craig, M.F. Barker, J. Cocker, P.C. Donohue and G. Vanrietvelde, 'New high yield, high reliability materials system for silver multilayer construction', Hybrid Circuits, No. 30, January 1993.

3. J.V. Manca, L. De Schepper, W. De Ceuninck, M. D'Olieslager, L.M. Stals, M.F. Barker, C.R. Pickering, W.A. Craig, E. Beyne and J. Roggen, 'In-situ electrical measurements on thick film dielectrics', Hybrid Circuits. No. 31, May 1993.

4. Needes, C.R.S., Hang, K.W. and Knaak, J.F., 'Manufacturing Yield of Pure-silver Thick Film Multilayer Circuits', in Proceedings IMC, Tokyo, May-June 1990.

5. R.S. Boikes and E. Edelson, 'Chemical principles', Harper & Row, Publishers, New York. 1981.

6. F.W. Sears, 'Electricity and magnetism', Addison-Wesley Publishing Company, inc., 1946.

7. G. Hársany, 'Dendritic growth from dielectric constituents in thick film IC's', Hybrid Circuits, No. 27, p. 54-59, January 1992.

8. Hang, K.W., Jaun, M.J., Needes, C.R.S., Halliyal, A., Donohue, P.C. and Drozdyk, L., 'Hermetic Devitrifying Glass Multilayer Dielectric for Advanced Hybrids', Proceedings IMC, Japan, 1990.

OPTICAL AMMETER FOR INTEGRATED CIRCUIT CHARACTERISATION AND FAILURE ANALYSIS.

W. Claeys+, S. Dilhaire+, D. Lewis+, V. Quintard+, T. Phan+ and J.L. Aucouturier*.

+Centre de Physique Moléculaire Optique et Hertzienne (CPMOH)
URA-CNRS 283
*Laboratoire d'Etude de l'Intergration des Composants et Systèmes Electroniques(IXL)
URA-CNRS 846
University Bordeaux I and CNRS
351 cours de la Libération, F-33405-TALENCE Cedex France
Tel : 56846200 - Fax : 56846970

Abstract: The current which flows through the metal semiconductor interface of an ohmic contact produces a Peltier effect. We optically detect this thermal effect and use it for the development of an optical ammeter, the determination of doping type of semiconductors and the homogeneity scanning upon integrated circuits.

1. INTRODUCTION

Every semiconductor device is inevitably connected to metallic lines for current transport. In planar geometry integrated circuits, the electrical connections to the components and the interconnections between components, necessarily proceed through ohmic contacts. These ohmic contacts, which link the metal to the semiconductor, are designed to have the lowest possible resistance in order to have a negligible influence upon the device characteristics. This work presents a new method for integrated circuit characterisation, based upon the detection of a thermoelectric effect upon ohmic contacts. This effect is measured with a high resolution laser probe (Refs. 1, 2) upon the ohmic contact. Ohmic contacts are spread all over the surface of integrated circuits and are of sufficient large size to be probed without perturbation with a laser beam of 1 μm in diameter. The thermoelectric effects that can take place in functioning semiconductor devices are the Joule, the Peltier and the Thomson effects. The first is highly reduced in ohmic contacts as the resistance is made as low as possible. The Thomson effect needs high temperature gradients to be significant, this is strictly avoided in integrated circuit design. Our experimental approach is therefore focused upon the detection of the Peltier effect in ohmic contacts. When a current flows between the metal and the semiconductor in an ohmic contact, a heating or cooling occurs due to a Peltier effect, and the corresponding thermal expansion or shrinking of the contact surface is observed by measuring the absolute surface displacement with our interferometric laser probe. The surface displacement due to the Peltier effect can selectively be detected and the response can be related to the device characteristics. We show, for example, how the current intensity in an ohmic contact can be measured through the surface displacement, showing the laser probe to be an optical contactless ammeter. This experimental approach can also be used for the determination of the doping type of the semiconductor under the metal contact.

The periodic heat exchange at an ohmic contact, when a sine wave current is flown through it, generates a thermal wave that propagates through the surrounding material. We show how the corresponding surface displacement wave propagates around the ohmic contact. An interesting situation is found when the surface waves from two identical ohmic contacts linked through a surface resistance interfere. The study of these Peltier effect generated surface waves has been the starting point for the development of a new imaging method of integrated circuits. We show phase and amplitude images of a simple test structure upon an integrated circuit obtained from ohmic contact thermal waves.

2. THERMOELECTRIC EFFECTS IN OHMIC CONTACTS.

An ohmic contact is defined as a metal-semiconductor connection that has negligible resistance relative to the bulk or spreading resitance of the semiconductor. The current conduction in these contacts can take place via two mechanisms. The first one is thermionic conduction and is illustrated in figure 1 for a n type semiconductor ohmic contact. The energy band diagram is shown and the conduction is illustrated for a given current direction. Majority carriers from the conduction band jump the contact barrier and reach the metal at the Fermi level energy. It is clearly seen that the carriers undergo energy shifts during this move, the net energy balance is $E_C - E_F$. The second mechanism is conduction through tunneling and is effective when a highly doped n^{++} region is added between the metal and the n type semiconductor, this is illustrated in figure 2. As can be seen from

the sketch, here too the electrons undergo a net energy jump of Ec - EF. Two similar conduction mechanisms are seen in the conduction of p type semiconductor ohmic contacts. Thermionic conduction is illustrated in figure 3 while tunneling is illustrated in figure 4. Note that the current direction has been changed with respect to figure 1 and 2. A net energy of EF - Ev is released per hole flowing through the contact. The four situations shown produce heating, but if we change the current direction, cooling occurs. This contact thermoelectric effect is a Peltier effect, the power exchange depends upon the current direction, the doping type and doping magnitude and also upon the type of metal at the junction. Besides the Peltier effect, a Joule effect will occur in the residual resistance of the ohmic contact, this latter is current square dependent and is therefore current direction independent.

For cosine wave excitation the power P(t) dissipated at an ohmic contact can be written:

for a n type:

$$P(t) = \left(\frac{Ec - EF}{q}\right) I_0 \cos \omega t + R I_0^2 \cos^2 \omega t \qquad (1)$$

$$P(t) = \frac{R I_0^2}{2} + \frac{R I_0^2}{2} \cos 2\omega t + \left(\frac{Ec - EF}{q}\right) I_0 \cos \omega t \qquad (2)$$

for a p type:

$$P(t) = \frac{R I_0^2}{2} + \frac{R I_0^2}{2} \cos 2\omega t + \left(\frac{Ev - EF}{q}\right) I_0 \cos \omega t \qquad (3)$$

If we only take the time varying part of the energy exchange, we observe a Joule contribution at 2ω (harmonic 2 of the excitation) and a Peltier contribution at ω (harmonic 1). By comparing relation 2 and 3 we see that the Peltier contributions of n and p type semiconductors have opposite phases. Energy exchange in the ohmic contact produces temperature changes which induce thermal expansion. Periodic displacements of the contact surface will be produced at ω and at 2ω.

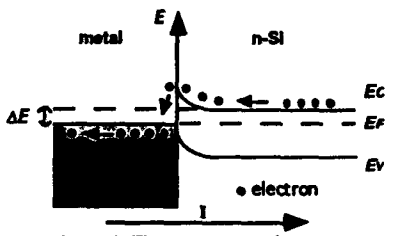


figure 1. Thermoionic conduction

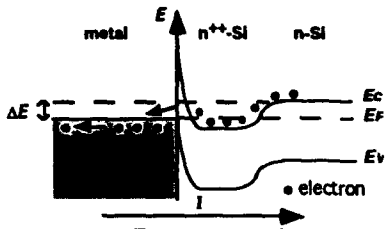


figure 2. Tunnel effect conduction

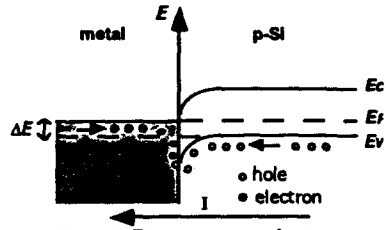


figure 3. Thermoionic conduction

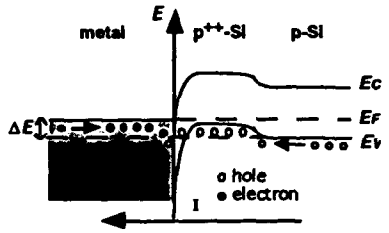


figure 4. Tunnel effect conduction

3. TEST STRUCTURE AND EXPERIMENTAL METHOD.

We have chosen to study the Peltier response of ohmic contacts upon a very simple structure in an integrated circuit, a Van der Pauw structure. A picture of it is shown in figure 5a. The four contacts in a Van der Pauw structure are used for conduction measurements of the different dopings in integrated circuit. We use this structure as a surface resistance and drive current between A and B, two opposite ohmic contacts in the structure. Figure 5b shows a sketch of the same structure, the dotted line shows the limits of the doped area.

For measuring thermal effects in integrated circuits, we use a high resolution interferometric laser probe we have developped (Refs. 1, 2). Absolute surface displacements are measured. The thermoelectric effects producing heat exchange in an ohmic contact, induce thermal expansion. The corresponding surface displacement of the ohmic contact is measured with the laser probe. The laser beam spot is 1 µm in diameter, which is much smaller than the contact surface. As an indication, the distance between contacts A and B in figure 5a is 112 µm.

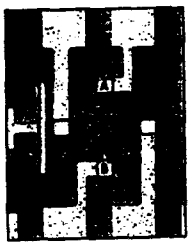


figure 5a.

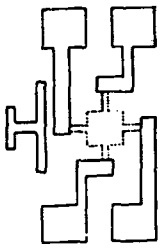


figure 5b.

Van der Pauw test structure

4. OPTICAL AMMETER.

Figure 6 shows the absolute expansion amplitude measured upon the ohmic contact A (figure 5) at harmonic 1 and 2 for different excitation current amplitudes (frequency 5 kHz). Pure linear first harmonic (Peltier effect) and pure quadratic second harmonic (Joule effect) responses are observed. This indicates that the Peltier effect can selectively be detected and that current amplitudes can linearly be derived from the displacement amplitudes. This shows the interferometric laser probe to be a non contact optical ammeter. Currents of a few microamperes at 5 kHz have been measured. It is interesting to see that for small current sensing Peltier heat detection is more sensitive upon ohmic contacts than the Joule one. This also shows the quality of the ohmic contact, as it is designed to have a negligible resistance. It is mportant to observe note the possibility to measure periodic surface displacements of a few femtometers with our interferometric probe.

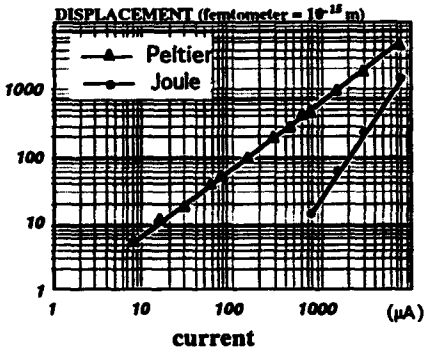


figure 6. Joule and Peltier expansion current dependence

Furthermore, if we measure the Peltier and Joule response upon ohmic contact B (figure 5), we observe the same amplitude responses as in figure 6, except that the phase of the Peltier effect has turned by an amount of 180°. This is easily understood as the Joule effects are identical in A and B while the Peltier effects are opposite. This phase change will be exploited in the next section when thermal waves will be observed.

5. THERMAL WAVES GENERATED AT OHMIC CONTACTS.

When current is driven through the semiconductor structure Peltier heat is generated in the ohmic contacts A and B of figure 5. They are the sources of two heat waves of opposite phase that propagate inside the semiconductor. We have explored with our interferometric laser probe the surface first harmonic thermal expansion (amplitude and phase) along the line joining A and B. Figure 7 shows the exponential decay in amplitude starting from the two sources 112 μm apart and interfering destructively at mid distance. Figure 8 shows the

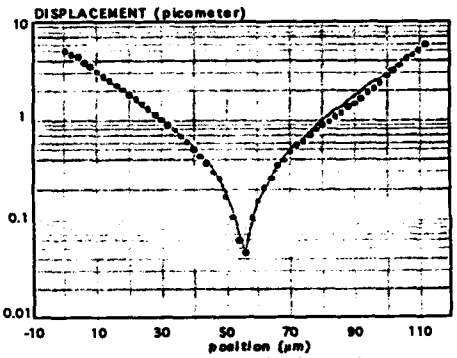


figure 7. Surface waves amplitude interferences

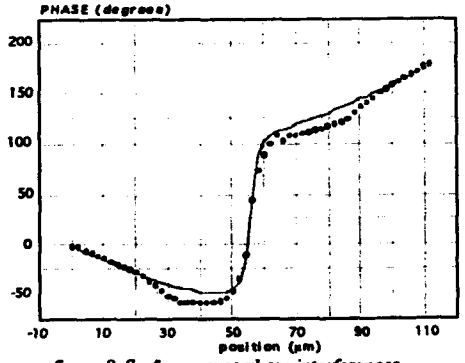


figure 8. Surface waves phase interferences

phase response indicating a linear dephasing starting from both sources opposite in phase and turning rapidly by 180° in the region of the interferences. Full curves in figures 7 and 8 correspond to the following simple expression of wave interferences :

$$A(x) = A(e^{-\alpha x} e^{-j\beta x} - e^{-\alpha(112-x)} e^{-j\beta(112-x)}) \quad (4)$$

with x in µm, α exponential decay coefficient and β linear dephasing coefficient.
Excellent agreement is observed between this simple model and the measurements. This indicates the thermal waves generated at ohmic contacts to be a good investigation means for semiconductor homogeneity testing. We show in the next section integrated circuit thermal wave imaging.

6. THERMAL WAVE INTEGRATED CIRCUIT IMAGING.

By flowing a sine wave current through opposite contacts in the Van der Pauw test structure presented in figure 5, we generate thermoelastic surface waves in the entire component. Figure 9 presents an image of the component obtained by mapping the amplitude of the thermoelastic first harmonic surface displacement.We observe a monotonous decrease in amplitude starting from both ohmic contacts. This gives information concerning the homogeneity of the semiconductor material around the contacts.

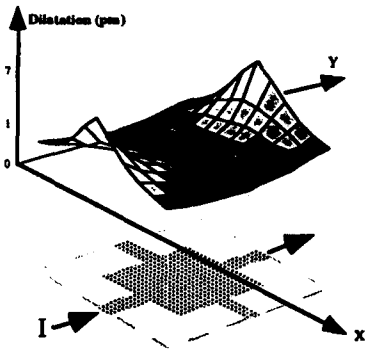


figure 9. Amplitude of Peltier displacement imaging of IC

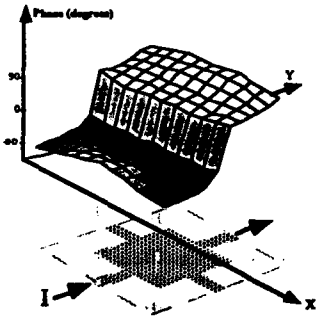


figure 10. Phase of Peltier displacement imaging of IC

Figure 10 shows the image obtained by mapping the phase of the first harmonic surface displacement. The borderline where the two opposite surface waves interfere is clearly outlined by the 180° phase jump. The fact that this borderline is straigth indicates the device to be homogeneous, the propagation of the waves has not been altered by accidents in the structure.

7. DOPING TYPE DETERMINATION.

Relations (2) and (3) show that the sign of the Peltier effect is dependent upon the sign of the current and upon the doping type. We analyse in table 1 below the Peltier result for different situations of non degenerated semiconductor ohmic contacts. First line of the table is easily understood by looking at the situation described in figure 1. We take the following convention : the current is positive when positive charges flow from the metal to the semiconductor in the ohmic contact.

Doping	Current sign	Power dissipated	Peltier effect
n	+ metal → SC	$\left(\frac{E_c - E_f}{q}\right) I \geq 0$	heating
n	−	$\left(\frac{E_c - E_f}{q}\right) I \leq 0$	cooling
p	+	$\left(\frac{E_v - E_f}{q}\right) I \leq 0$	cooling
p	−	$\left(\frac{E_v - E_f}{q}\right) I \geq 0$	heating

Table 1.

For pulsed or DC current we measure if the contact is expanding (heating) or contracting (cooling). We can thus obtain from table 1 the doping type of the semiconductor provided the direction of the current is known. On the contrary the direction of the current can be determined provided the type of the doping is known. The same holds for phase change in sinusoidal excitation. We have tested the statements of table 1 for different dopings upon an integrated circuit and found exactly the specifications of the design.

8. CONCLUSION.

We have shown that the measurement of the surface displacement of an ohmic contact with our high resolution interferometric laser probe provides a contactless reading of the current amplitude running through it. This reading is related to both the Peltier and Joule thermoelectric effects. Selective detection of these effects is possible as the first is linear with current intensity and the

second is quadratic. This detection method upon ohmic contacts is the basis of an optical ammeter. Applications in the field of functionnality, reliability and fault analysis is readily seen.

We have also shown the possibility to determine the doping type of the semiconductor of the ohmic contact. This leads to applications in the field of non-destructive external design inspection.

We have finally shown ohmic contacts to be the source of thermal waves propagating in the surrounding material. The detection and mapping of the generated surface waves, in amplitude and phase, provides a new imaging method for integrated circuits. These images are new tools in the fied of quality and reliability testing.

9. REFERENCES.

1. Claeys W., Dilhaire S., Sarger L. and Zardini Ch. "Joule displacement and thermoelastic optical test probe for electronic components. ", ESREF 91, 7-10 October Bordeaux, (1991) pp 819-827.

2. Claeys W., Dilhaire S. and Quintard V. "Laser beam probing of thermal behaviour of electronic components and its application in quality and reliability testing.", Microelectronic Engineering 24, (1994), pp 411-420.

Process Maturity Grids used as a Decision Support Tool

R.G.J. Arendsen[*Δ], A.C. Brombacher[Δ+*], B. Habraken[Δ], O.E. Herrmann[*]

[*]Twente University, Dept. Electrical Engineering, Laboratory for Network Theory, Bldg EL/TN 9234, P.O. Box 217, 7500 AE, Enschede, The Netherlands

[Δ] Philips Consumer Electronics, Dept. Quality Engineering, Bldg SK6.32 , P.O. Box 80002, 5600 JB, Eindhoven, The Netherlands

[+]Technical University Eindhoven, Dept. Mechanical Engineering, Group WOC, Bldg W-hoog 3.118, P.O. Box 513, 5600 AE, Eindhoven, The Netherlands

Summary

Failure rates of modern ASICs and ASSPs can not be demonstrated using reliability tests on product level. Reliability demonstration must be done by quantifying the quality of the process that produces these integrated circuits. Process capability however, is not fixed over time. As a process gets more mature process capability tends to improve. In this paper we introduce the concept of process maturity growth to quantify the quality of suppliers processes as a function of time. The method is illustrated for a submicron double metal CMOS process in which many mixed standard cell and full custom designs are processed.

Introduction

Traditionally qualification of application specific integrated circuits (ASICs) and application specific standard products (ASSPs) is done on product level. This means that after design, manufacturing and evaluation of the integrated circuits, product tests are carried out to ensure product quality and reliability. For several reasons this approach will become obsolete in the near future (Refs. 1-3). Most important reasons are:

- currently required low failure rate levels can no longer cost effectively be measured on product level because this would require too many test samples
- decreasing time to market of new ASIC designs does not allow a long qualification phase
- there is a trend towards more functional diversity and smaller manufacturing volumes, full product qualification does not pay off for these small manufacturing volumes

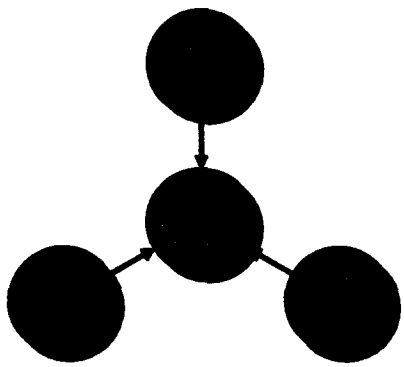


Fig. 1. *Changing prerequisites for IC qualification processes*

These developments are schematically given in Fig. 1 and they require a change in vendor/customer relationships with respect to qualification and reliability demonstration (Ref. 3). From consumer electronics industry we observe two changes that are going on at this moment. *First*, the number of suppliers delivering integrated circuits are minimised to a number of preferred suppliers.

Secondly, as a consequence, not only the products of these preferred suppliers are qualified but more important, the design and manufacturing processes of the supplier are qualified. For a semiconductor supplier these are the IC design process, the wafer fab process and the assembly process. This change from product qualification to process qualification is schematically given in figures 2 and 3.

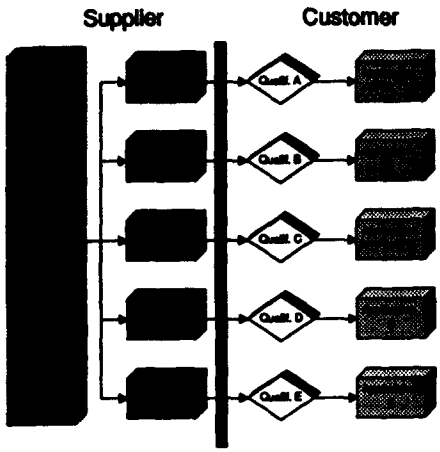


Fig. 2. *Traditional type related IC qualification processes*

The approach, given in Fig. 3 is conceptually very simple. It is important, however, to emphasise three major prerequisites, necessary for a successful completion of this process.

1. A supplier and a customer of an integrated circuit should agree to exchange information beyond the normal product specifications

2. A supplier and a customer should agree what parameters in (design and production) processes of a family of ICs are relevant for a customer

3. A supplier and a customer should agree how the results of such a process assessment are monitored; not only during a first qualification but also in time

This process based approach towards device qualification is not new in electronics industry. Today already many ASICs that are used in consumer electronics products are qualified by their design and production processes. Some examples are gate arrays and digital standard cell designs. These design technologies have in common that design and production processes can be highly standardised. The design freedom is limited and device qualification can be done on basis of design and production processes for two reasons:

1. There is a large structural similarity between the devices
2. Performance and reliability are regularly monitored respectively demonstrated using a representative test device. For standard cell this can be a special designed IC with part of the library and test structures on it.

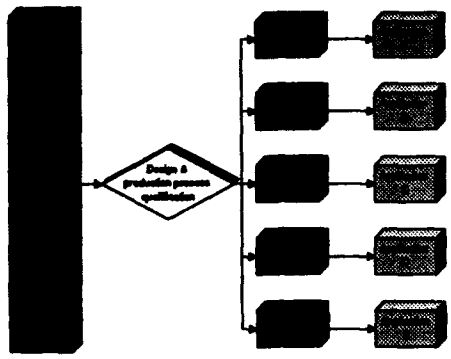


Fig. 3. *Qualification, based on design and production processes*

The opposite regarding design freedom are ASSPs that are full custom designed. This is given in Table 1. The freedom in this design style is such that electrical characterisation and reliability demonstration are typically done for each IC type. Many ASIC designs that we face in consumer electronics products are standard cell designs with some custom designed instances or mixed digital/analogue ASICs. For cost and/or performance reasons they may be realised in a new production technology. Given the developments from Fig. 1 it are these type of ASICs where it is most difficult for, to develop efficient standard process based qualification procedures.

IC type	type of qualification
FPGA/ gate array	process
standard cell	process
mixed std./custom mixed analogue/dig.	process/type
full custom ASSP	type

increasing design freedom

increasing qualification effort on type basis

Table 1: *Different types of ASICs in consumer electronics industry have different qualification strategies*

At this moment for most of this type of ASICs type qualification is performed. For reasons mentioned before our goal is to come to a process way of qualification also for these type of circuits. In this paper such an approach is described for IC production process. Design- and assembly are not included. It is a two step approach:

1. Assess the IC production process for the risks involved for the customer

2. Limit the type specific qualification tests to the items that pose a risk for the customer

This approach is not so straightforward as it may seem because in general process capability is a function of time. As more products are made in a specific process the maturity of the process will increase. To monitor this, we will introduce the concept of process maturity growth to be able to quantify the capability of the suppliers processes with regard to customer products as a function of time.

Process maturity growth

With a simple example we will illustrate what is meant by process maturity growth. For this example we will use "process defect induced" early failure rate (EFR) as a set of failure mechanisms that needs to be described over time. What typically is found when a new production process grows towards a mature process is given in Fig. 4. Typically once the process is able to make ICs that perform to function and reliability (end of life) requirements (this is called potential phase) the EFR is still much higher than the ultimate capability of the manufacturing line. Due to all sorts of improvement actions (e.g. on scratches, particles) the EFR of the new process improves rapidly. However due to special failure modes the performance is not continually decreasing but in the beginning of the lifetime of the new process some "humps" are seen in EFR rate due to some special failure modes. This period is called consistency phase. At the end of the consistency phase these special causes are eliminated and remaining defects are successfully contained the EFR of the new process will become stable. Containment of the defects can not only be done on process level by the control of particles but also by yield screening on wafer level (Ref. 4) or by I_{DDQ} testing on product level (Ref. 5). Most interesting for the customer of products from the new process is the point in time where the process defect induced EFR from the new process begins at the performance phase.

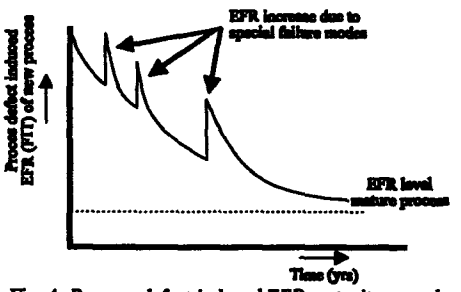


Fig. 4. *Process defect induced EFR maturity growth*

As an example we have sketched the development of process defect induced EFR as a function of time. The maturity growth is valid for many process aspects. Some of these can be PCM parameters, random yield, PPM levels, etc. Process parameters that are under SPC typically develop in the same way. This is given in Fig. 5.

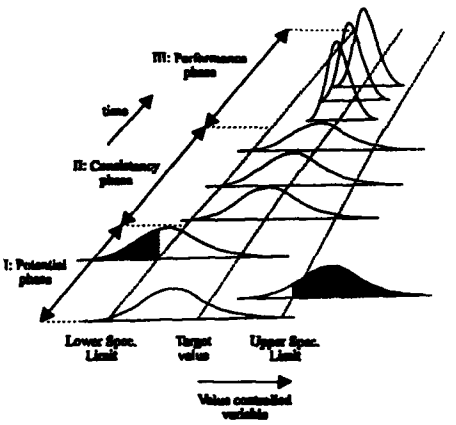


Fig. 5. *Different phases during process life of process variables under SPC*

In Fig. 5 the following three phases can be distinguished.

1. Potential	At the end of this phase the process is able to produce devices that conform to function and reliability (end of life) requirements.
2. Consistency	At the end of this phase all special causes of variation are eliminated so that the yield and early life failure rates are stable.
3. Performance	During this phase common cause (inherent) variation is reduced such that quality and early life failure rates are met.

In the previous examples we have shown how the SPC controlled variables as well as non SPC controlled variables typically develop as the production process grows to a mature process. This trend is valid for many process aspects and therefore a similar trend can also be found in the overall process capability. In Fig. 6 an example is given of how the overall process capability changes during a process lifetime.

In Fig. 6 two lines are drawn. One is indicating the process capability of an existing mature process. The other line describes the process capability of a new process which technology resembles the mature

process (e.g. only the minimum dimensions are smaller).

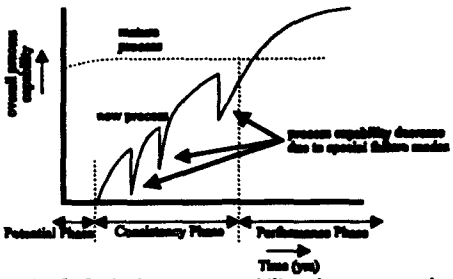


Fig. 6. *Typical process capability of a mature and new manufacturing process*

We see that the mature process has a stable process capability. The capability of the new process is at the start less than the capability of the mature process but when the new process becomes mature its capability will become better than the existing process. Most interesting for customers is the point in time where the capability of the new process becomes better than the capability of the existing process.

Because the process capability of a new production process depends so much on the maturity of the process we developed a simple method to monitor process maturity growth.

Monitoring process maturity growth

In order to monitor the maturity of the process on a regular basis we will use *maturity grids*. *Maturity* grids are two dimensional drawings in which the maturity of one process aspect can be indicated. The maturity grid can have any size but for the sake of simplicity we will use example grids that have four columns and four rows making a total of sixteen fields. This grid is shown in Fig. 7. The maturity grid is a risk assessment tool. This means that a customer can rate the risk that certain process aspects has for the products he buys from the supplier for his application. A risk for the customer can be defined as the product of the chance on failures and the severity of these failures for the customer. Thus:

Risk for customer =
P(failures) * Severity(failures)

Along the horizontal axis of a maturity grid the chance for failures is indicated. Along the vertical axis the severity of the failures for the customer are indicated. An example maturity grid is drawn in Fig. 7.

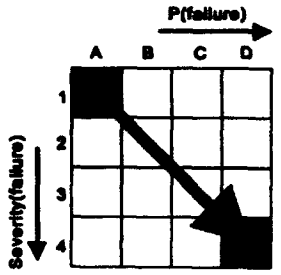


Fig. 7. *An example of a maturity grid*

Position A1 is the area with the highest risk. Position D4 is the area with the lowest risk. As the suppliers process gets more mature process items will shift from the upper left corner to the lower right corner in the maturity grid, typically somewhere along the diagonal line, as indicated by the arrow in Fig. 7.

To give a technical example of a maturity grid, in Table 2 we have given the values along the axis for the risk for the customer with respect to EFR. Although the values along the horizontal axis are purely illustrative the chance for failures is given by the EFR of similar products in the same production process. The consequence of a failing device due to process defects is likely to be that the application of the customer does not work according to specification. The severity of this problem for the customer largely depends on the ability of the supplier to decrease the EFR. There is a large difference in severity between the situation that the supplier faces a process limit or the situation that the supplier will on short term be able to improve the EFR. Therefore the vertical axis indicates the severity of the problem.

X-axis		Y-axis EFR (FIT)	
A	no improvement, equipment/ process capability limit	1	> 100
B	50% of all failures are of known cause	2	< 100
C	most failure modes are known, long term improvement	3	< 50
D	most failure modes are known, short term improvement	4	< 10

Table 2. *Values along the axis for the maturity grid regarding EFR*

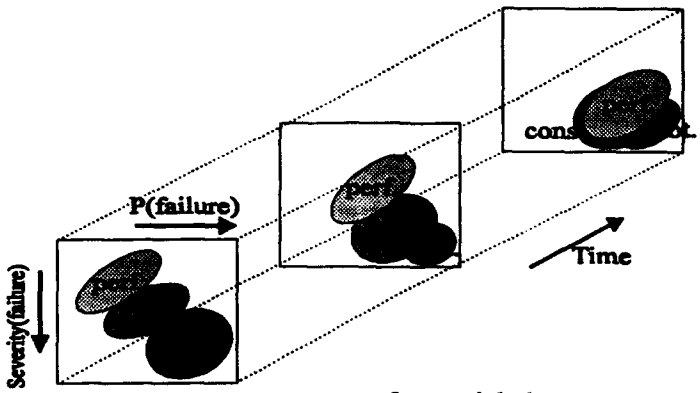


Fig. 8. *Expected process maturity growth*

In the approach presented in this paper such a grid is made for many process aspects that influence product quality and reliability. By stacking all the grids on top of each other it becomes clear which process items need most attention from the supplier and/or customer in order to improve product quality and reliability. In Fig. 8 a set of maturity grids is given that describes how the maturity of the process typically will develop as a function of time. In this figure the axes of the maturity grids have become continuous (Ref. 6). Given the maturity growth from Fig. 6 it is expected that process items from the potential phase will be mature before items from the consistency phase. After the consistency phase items from the performance phase will become mature.

An example

As was described before, during the maturity growth of a semiconductor manufacturing process for many process aspects three phases can be distinguished. These are the potential phase, the consistency phase and the performance phase. In Table 3 the customer concerns regarding these three phases are described.

process maturity phase	*customer concern*
potential phase	Is the process capable of delivering ICs that conform to functionality and reliability (end of life) requirements?
consistency phase	Are special causes identified and removed on time such that no delivery problems will occur?
performance phase	Are EFR and quality levels (PPM) met?

Table 3. *Customer concerns during process phases*

Process defect induced EFR has been the only failure mechanism that has been treated so far in this summary as an example failure mechanism. Of course in an actual process maturity assessment many more process items are reviewed. In this section we will shortly give an overview of the maturity grids that are used to assess an industrial submicron double metal CMOS process. Table 4 lists the most important items of which maturity grids are made. The items mentioned here are only top level items. In order to be able to assess process details while keeping an overall picture we have implemented a hierarchy in this process assessment. For example the EFR related maturity grid in the performance phase was given in Table 2. However if the maturity of this item is not D4 the supplier is asked to fill in more detailed maturity grids. These are based on a pareto of the most important failure modes (e.g. gate oxide, litho defects, inter-metal oxide particles). For each failure mode the maturity status must be indicated.

Potential	Technology maturity
	Library functionality
	Library reliability
	Reliability evaluation modules
Consistency	Special causes in-line failures
	Special causes ESORT failures
	Special causes PCM failures
	Special causes EFR failures
Performance	PPM risk factor
	Cpk PCM parameters
	Cpk in-line parameters
	EFR risk factor

Table 4. *Main groups of process items*

Within many of these groups detailed maturity grids are drawn. By stacking all top level maturity grids

we get a good overview which process aspects are the least mature for the product. At this moment we have implemented the approach outlined in this paper with one supplier of ASICs and ASSPs in a submicron CMOS technology. Since the method has been recently developed, at this moment only one maturity assessment has been done. Because the tool is still in an evaluation phase at this moment the method is used in addition to the traditional way of type qualification. In our experience so far, we have found that there are several advantages in using maturity grids for describing process maturity. Some of these are:

- the tool is easy to use, people responsible (e.g. process engineers, test engineers, reliability engineers) can fill in these maturity grids easily
- the maturity grids give a good overview which process items pose the highest risk for the customer of the ASICs
- by filling out the maturity grids on a regular basis it is possible to assess the process maturity as a function of time

Conclusions

Future reliability targets of ASICs and ASSPs can not be demonstrated by product life tests. These figures must be shown by the production process that produces these products. This requires a change in vendor/customer relationship.

The capability of the production process is not constant. As the process gets more mature the capability tends to improve. This is an important aspect for qualification of production processes.

Once the axes along the maturity grids are established, maturity grids are easy to use. By performing process maturity assessment on a regular basis the maturity can be found as a function of time.

References

1. Crook D.L., Evolution of VLSI Reliability Engineering, Proc. ESREF91, Bordeaux, France
2. Shafft H.A., Baglee D.A., Kennedy P.E., Building in Reliability: Making it Work, Proc. ESREF91, Bordeaux, France
3. Gerling W., Integration of Efforts for the Reliability of Microelectronic Devices, Proc. ESREF93, Bordeaux, France.
4. Huston H.H., Clarke, C.P., Reliability Defect Detection, Monitoring and Screening During Processing, Proc. ESREF92, Schwäbisch Gmünd, Germany.
5. McEuen S.D., Reliability Benefits of I_{DDQ}, JETTA, vol3, nr 4, December 1992.
6. Hugge P.B. A Quality Process Approach To Electronic Systems Reliability, Proc. of the Reliability and Maintainability Symposium, 1993.

RELIABILITY ASSESSMENT OF POTENTIOMETERS BY 1/f NOISE ANALYSIS

E.P. Vandamme and L.K.J. Vandamme

Eindhoven University of Technology
Department of Electrical Engineering
5600 MB Eindhoven, The Netherlands

ABSTRACT

The 1/f noise of the conductive track, end contacts and stationary wiper in non-wirewound potentiometers have been investigated. The resistance and the noise of the stationary wiper are constriction dominated. Physical models are proposed to explain experimentally observed trends. Potentiometer 1/f noise is due to contact noise between grains in the track. Coherence measurements serve as a fast diagnostic tool for reliability and quality assessment of the movable contact.

1. INTRODUCTION

Our aim is to provide a fast diagnostic tool to judge the quality and reliability of a potentiometer which is often the weakest component of a complex electronic system. The excess noise in the track, in a stationary wiper and in the end contacts as well as the coherence between them are used as a fast and non destructive technique for quality testing.

2. SIMPLE MODEL FOR RESISTANCE, NOISE AND COHERENCE

In our model two parameters are used, (i) the ratio ν of the noise of the movable contact to the noise of the total resistance, and (ii) the ratio σ of the total resistance of the potentiometer track to the resistance of the movable contact. Reliability problems can be expected when ν and σ depend on the position of the wiper contact on the track, because this points to a fluctuating pressure of the movable contact on the track.

The noise in conducting tracks for non-wirewound potentiometers is of the same variety as the noise observed in thick-film resistors, having power density spectra inversely proportional to the frequency f. A refined treatment can be found in (Ref. 1). The resistance of the conductive track is given by

$$R=LR_{\square}/W=L/q\mu nWt \qquad (1)$$

with R and $R_{\square}$ the track resistance and sheet resistance repectively; L, W and t the length, width and thickness of the track and q, μ and n the elementary charge, the free carrier mobility and concentration respectively. Hence the sheet resistance is given by

$$R_{\square}=1/q\mu nt \qquad (2)$$

The relative noise in the resistance of the track S_R/R^2 is inversely proportional to the track area WL with a proportionality coefficient C_{us} which has the dimension of mm² (Refs. 2-4)

$$S_R/R^2=C_{us}/WLf=\alpha q\mu R/L^2f \qquad (3)$$

hence

$$C_{us} = \alpha q\mu R_{\square} = \alpha/nt \qquad (4)$$

with α the 1/f noise parameter. This noise parameter is for homogeneous films of the order of 10^{-4} (Refs. 2,3).

Contact complications such as uniform films (Ref. 5) or punched films (Ref. 6) between wiper contact and track are ignored. Then the resistance of a stationary wiper contact on the track is given by

$$R_c\approx 1/2\pi aq\mu n\propto R_{\square} \qquad (5)$$

with a the contact radius. For the relative noise of the wiper contact holds (Ref. 7)

$$S_{R_c}/R_c^2=\alpha/10\pi a^3nf \qquad (6)$$

The ratios $\sigma=R/R_c$ and $\nu=S_{R_c}/S_R$ can be calculated using (1), (3), (5) and (6) and are given by eqs. (7) and (8)

$$\sigma=R/R_c\approx 2\pi aL/Wt \qquad (7)$$

and

$$\nu=S_{R_c}/S_R=(Wt)^3/40\pi^3a^5L \qquad (8)$$

The noise in a stationary wiper is very sensitive ($\nu\propto 1/a^5$) to changes in contact spot diameter along the track. This is the physical background why 1/f noise and coherence measurements are such sensitive diagnostic tools for the detection of wiper contact problems and reliability studies.

We restrict ourselves here to coherence measurements between voltage fluctuations. The potentiometer is biased with a noise-free constant current source I as in Fig. 1. We assume that V_1 and V_2 are amplified with a noise-free voltage amplifier. In fig. 1 the resistance and noise of the end contacts on the track are ignored.

The observed voltage fluctuations in V_1 and V_2 are due to conductance fluctuations in the track. The spectra S_{V_1} and S_{V_2} and the cross-correlation spectrum $S_{V_1V_2}$ are given by equations (9), (10) and (11)

$$S_{V_1}=I^2(S_{R_c}+S_{R_1}) \qquad (9)$$

$$S_{V_2} = I^2 S_{R_1} \tag{10}$$

and

$$S_{V_1V_2} = S_{V_2} \tag{11}$$

The coherence function defined as

$$\gamma^2_{V_1V_2} \triangleq \frac{S^2_{V_1V_2}}{S_{V_1} S_{V_2}} \tag{12}$$

now becomes

$$\gamma^2_{V_1V_2} = \frac{1}{1 + S_{R_c}/S_{R_1}} = \frac{p}{p(\nu+1)+\nu} \tag{13}$$

with $p = R_1/R_2$. Large values of $\gamma^2_{V_1V_2}$ indicate relative small contributions of the resistance and the noise of the wiper contact to the track resitance and the noise of the track, respectively. This points to a reliable and high quality potentiometer.

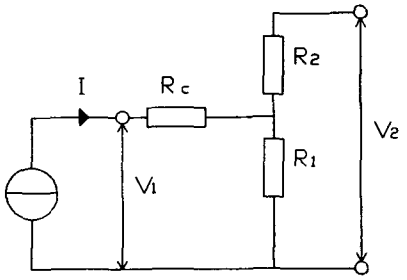


fig. 1: Measurement setup. The movable contact separates the track resistance R into R_1 and R_2. The resitance of the stationary wiper contact on the track is presented by R_c.

3. EXPERIMENTAL RESULTS

We have observed the relative noise of tracks in different potentiometers with different dimensions and sheet resistance. For potentiometers having the same sheet resistance the observed values of C are inversely proportional to the area as can be seen in (Ref. 8), where C is defined as fS_R/R^2.

Fig. 2 shows the C_{us} values of several carbon potentiometers with different sheet resistance. Only for constant values of $\alpha\mu$ we can expect $C_{us} \propto R_\square$. Fig. 2 shows that C_{us} is quadratically dependent on $R_\square$.

In fig. 3 the resistance of the wiper contact on the track is plotted versus the sheet resistance. Because R_c increases linearily with $R_\square$ we conclude that the contacts are constriction dominated, without film complications as can be seen from eq. (5). The same trend has been observed in (Ref. 9).

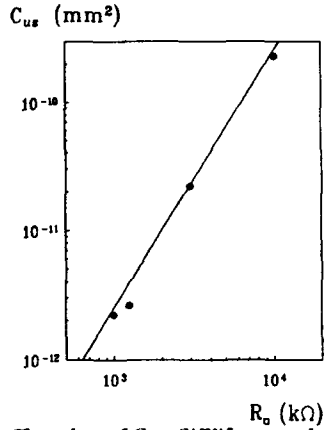


fig. 2: The values of $C_{us} = C*W*L$ versus the sheet resistance for four carbon tracks of different microstructure.

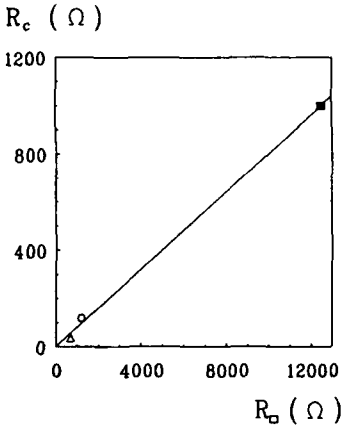


fig. 3: The resistance of the movable contact versus the sheet resistance for three different tracks.

Fig. 4 shows the spectra S_{V_1} (△), S_{V_2} (○) and the coherence $\gamma^2_{V_1V_2}$ (■) versus frequency f of a potentiometer of 10 kΩ. The large thermal noise contribution of R_2 shows up in S_{V_2} above 300 Hz. The values of $S_{V_1V_2}$ and $\gamma^2_{V_1V_2}$ were observed with an HP 35665A double-channel FFT spectrum analyzer. The value of $\gamma^2_{V_1V_2}$ below 300 Hz agrees with the calculated value from our equations using ν and p from an independent measurement. The observed noise due to conductance fluctuations in all investigated potentiometers showed a 1/f spectrum. The results are summa-rized in tables I and II.

Table I: Characteristics and noise results of potentiometers with a nominal value of 10kΩ

potentiometer no.	I_{max} (mA)	L/W (mm/mm)	$R_\square$ (Ω)	$C_R = fS_R/R^2$	$C_{R_c} = fS_{R_c}/R_c^2$	$C_{us} = WLC_R$	$\nu = S_{R_c}/S_R$	$\sigma = R/R_c$
1	4.5	24/1.6	667	10^{-13}	10^{-9}	$3.8*10^{-12}$	0.3	170
2	5.3	24/2.4	1000	$5*10^{-14}$	$2*10^{-10}$	$2.9*10^{-12}$	0.5	80
3	6.3	24.7/2.8	1134	$3*10^{-14}$	10^{-10}	$2.1*10^{-12}$	0.4	90
4		5.5/0.9	1636	10^{-13}	$2*10^{-10}$	$5.0*10^{-13}$	1.5	30
5	5.3	24/2.4	1000	$3*10^{-14}$	$5*10^{-10}$	$1.7*10^{-12}$	6.1	50
6	4.5	24/1.6	667	10^{-13}	$3*10^{-10}$	$3.8*10^{-12}$	0.1	180

Table II: Characteristics and noise results of potentiometers with a nominal value of 100kΩ

potentiometer no.	I_{max} (mA)	L/W (mm/mm)	$R_\square$ (kΩ)	$C_R = fS_R/R^2$	$C_{R_c} = fS_{R_c}/R_c^2$	$C_{us} = WLC_R$	$\nu = S_{R_c}/S_R$	$\sigma = R/R_c$
7	1	13.8/2.15	15.6	$5*10^{-12}$	$2*10^{-9}$	$1.5*10^{-10}$	0.1	60
8	2.2	15.3/1.8	11.8	$9*10^{-12}$	$3*10^{-9}$	$2.5*10^{-10}$	0.04	90
9	2.2	14.1/2	14.2	$2*10^{-14}$	$3*10^{-10}$	$5.6*10^{-13}$	2.2	100
10	7.1	36.3/3.9	10.7	$3*10^{-15}$	10^{-9}	$4.2*10^{-13}$	0.8	670

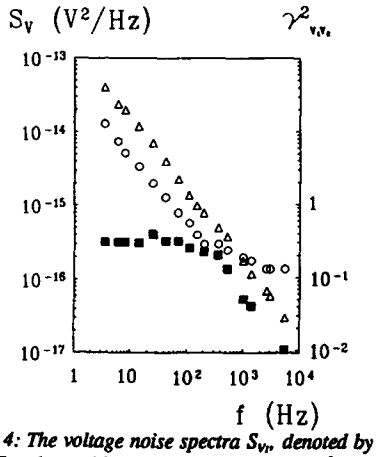


fig. 4: The voltage noise spectra S_{V_1}, denoted by △, S_{V_2}, denoted by ○, and the coherence $\gamma^2_{V_1V_2}$, denoted by ■, versus the frequency f, with $R_1/R=0.05$, $R=9.8k\Omega$ and $I=0.17mA$.

In figs. 5 and 6 the 1/f resistance noise, normalized for frequency, fS_{V_1}/I^2 (△) and fS_{V_2}/I^2 (○) observed in the circuit of fig. 1, are shown as a function of the position of the movable contact for two carbon potentiometers together with the plateau value of the coherence $\gamma^2_{V_1V_2}$ (■). A high value of $\gamma^2_{V_1V_2}$ for p=1 or $R_1/R=0.5$ indicates a high-quality potentiometer, considering its wiper contact noise. A large scattering in fS_{V_1}/I^2 points to a non-uniform resistance of the wiper contact along the track and is an indication for a poor quality potentiometer. The 1/f noise in the wiper and end contact resistance or in the end contact resistance only is obtained from the intersections of fS_{V_1}/I^2 and fS_{V_2}/I^2 in figs. 5 and 6 with the y-axis.

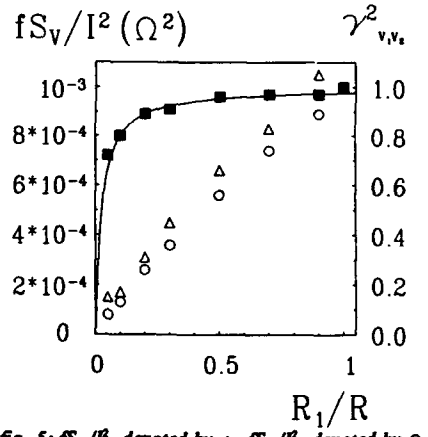


fig. 5: fS_{V_1}/I^2, denoted by △, fS_{V_2}/I^2, denoted by ○, and $\gamma^2_{V_1V_2}$, denoted by ■, versus R_1/R. The full line is the calculated value of $\gamma^2_{V_1V_2}$ using (12) with $\nu=2.5*10^{-2}$.

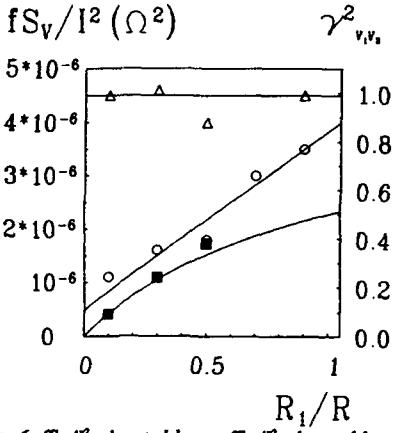


fig. 6: fS_V/I^2, denoted by △, fS_{V_n}/I^2, denoted by ○, and $\gamma^2_{V_1V_2}$, denoted by ■, versus R_1/R. The full line is the calculated value of $\gamma^2_{V_1V_2}$ using (12) with $\nu=1$.

Figs. 5 and 6 also show the experimentally obtained coherence $\gamma^2_{V_1V_2}$, (■) versus R_1/R, and the calculated $\gamma^2_{V_1V_2}$ (full line). Fig. 6 shows the typical results for a poor quality potentiometer with a low reliability and a relatively large S_{R_c} contribution which is not constant along the track.

4. CONCLUSIONS

Poor quality potentiometers have too small electrical contact area between wiper and track and large track conductance noise. Carbon black resin types are much noisier than metal-oxide glass cermet types. Reliability problems occur when a strong scattering in the ratio $\nu=S_{R_c}/S_R$ is observed for different positions along the track. Early aging after some position shifts of the wiper along the track is better seen from the noise ($\nu \propto 1/a^5$) than from $R_c \propto 1/a$. Coherence measurements serve as a fast and sensitive diagnostic tool for the quality and reliability of the movable contact.

REFERENCES

(1) E.P. Vandamme and L.K.J. Vandamme, "1/f noise and its coherence as a diagnostic tool for quality assessment of potentiometers", IEEE Trans. Comp. Hybr and Manuf. Tech. (1994), accepted for publication

(2) F.N. Hooge, T.G.M. Kleinpenning and L.K.J. Vandamme, "Experimental studies on 1/f noise", Rep. Prog. Phys., 44, pp 479, 1981

(3) L.K.J. Vandamme, "noise as a diagnostic tool", IEEE Trans. El. Devices, special Sept. issue on noise (1994)

(4) L.K.J. Vandamme, "Criteria of low-noise thick-film resistors", Electrocomp. Science and Technology, 4, pp 171, (1977)

(5) L.K.J. Vandamme, "1/f Noise of point contacts affected by uniform films", J. Appl. Phys., 45, pp 4563 (1974)

(6) L.H.F. Ortmans and L.K.J. Vandamme, "Characterization of impulse-fritting procedures of contacts by measuring 1/f noise", Appl. Phys., 9, pp 147 (1976)

(7) L.K.J. Vandamme, "On the calculation of 1/f noise of contacts", Appl. Phys., 11, pp 86 (1976)

(8) B.S. Chiou, J.Y. Sheu, and W.F. Wu, "Size effect on the electrical conduction and noise of RuO_2-based thick-film resistors", J. Electr. Mat., 21, pp 1105 (1992)

(9) A. Banovec and A. Zalar, "Investigation of sliding contact resistance of Ni-Cr/Au and Ni-Cr/Au-SiO_2 thin resistive films", Thin Solid Films, 164, pp 129 (1988)

RELIABILITY TEST RESULTS ON COATING AND UNCOATING PCB's

Adell, J.A., E. Nogueira, J.E. Vila
Reliability & Materials Division
TELEFONICA INVESTIGACION Y DESARROLLO
C/ Emilio Vargas, 6
28043 Madrid
Spain
Phone:34-1-337.43.88/34-1-337.44.89
Fax: 34-1-337.42.12

ABSTRACT

This paper present a comparative reliability analysis of coated and uncoated PCB's. The influence of the cleanness process and the different kinds of conformal coatings were evaluated.

1.INTRODUCTION

Several failures were detected in equipments installed in high humidity environments or in areas nearest to the sea-side. In these equipments the PCB's were uncoated and the main important *failure mode detected was shortcircuit*, caused by dendritic growth, between conductors or solder joints.

In these tests the failures produced in the field were studied and were compared with the failures obtained in the climatic test in the laboratory.

These tests have been carried out in order to reproduce the failure mechanism and to stablish the mean time between failures (MTBF) on coated and uncoated PCB's. Furthermore, the influence of the cleanness process, prior to coate the PCB's, on the reliability was determined.

2.TEST METHODS

For comparison purposes circuits without protection, with conformal coating without cleanness process prior to coating and with cleanness process prior to coating were tested. Insulation resistance and dielectric strength were measured on all test circuits prior and after the climatic test. The insulation resistance reading was taken after application of 500 V DC for one minute. In the dielectric strength measure test voltage was applied by increasing from 0 to 1250 V DC at a rate of 500 V/s, after which leakage current was measured.

Test samples were subjected to a 1000 hours compressed reliability test, the cycle is showed in Figure 1. Test samples were biased with 48 V and the drop voltage in a limitation resistance was measured every ten minutes. Limitation resistance of 1 Mohm value were introduced in the control PCB, outside of the climatic chamber, and connected with the samples in order to limite the current when the dentritic growth take be place and avoid the volatilization of the dentrites by the power suddenly dissipated through them.

The aim of this test was to reproduce the drops of condensation during the changes between night and day, wich could be the main principal mechanism of failure in the field. Each cycle with a duration of 3 hours simulate one day, then is possible obtain an accelerated factor of 8.

Furthermore, high temperature and high humidity conditions were used in order to accelerate the failure mechanism, but the acceleration rate of these factors can't be determined because several tests in diferent conditions were necessary and it was difficult to reproduce the drops of condensation in different conditions.

3.DISCUSSION

Test results were plotted on Weibull paper. From this representation is possible to obtain shape (β), scale (characteristic life, η) and the mean time between failures (MTBF).

Uncoated PCB's showed a poor performance in all the manufacturers tested. The MTBF's were

between 10 and 30 hours, the shape (ß) was arround 1 as shown in Figure 2. The main important failure mechanism detected was dendritic growth between solder joints and conductors. The insulation resistance and dielectric strength of this circuits have an important drop.

Conformal coated PCB's without cleanness process showed a better perfomance than uncoated PCB's. The MTBF's were between 60 and 80 hours, the shape were between 3.5 and 4.2 as shown in Figure 3. The increase in the shape value can be explained by the lateness in the access of the humidity to the surface of the PCB, however the failures apparition rate increase faster at a consequence of the impossibilty to eliminate the humidity of the surface of PCB due to the presence of the coating. In these samples the dendritic growth was present in areas where solder fluxes residues were present. These findings can be explained by the presence of ionic contamination on the PCB's, wich in presence of moisture, retained in the manufacturing process of PCB's and in the climatic chamber, can increase the conductivity, furthermore, in these samples blistering and worst adherence of coating to PCB'S were detected, at a consequence corrosion in the conductors track was detected. Conformal coated PCB's increased the MTBF at double than uncoated PCB's.

Conformal coated PCB's with cleanness process prior to the application of conformal coating products, showed the best performance and adhesion of coating. The MTBF of these samples were arround 300 hours and the shape was arround 3 as shown in Figure 4. Dendritic growth wasn't present in these lots and the failure mechanism were associated with the failure on uncoated components mounted on the circuit. The insulation resistance and dielectric strength remained unalterable.

Conformal coating products with different chemical composition and from different manufacturers showed no significant differences in the results of the coated PCB's in the climatic tests.

4.FAILURES ANALYSIS

The main important failure mechanism detected in the samples failed in the outside plant was dendritic growth between the solder joints with different polarity or between conductor tracks of the PCB as shown in Figures 5 and 6. The failure occurs as a consequence of the presence of humidity inside the equipment which forms a layer of a polar liquid on the surface to form an electrolyte. This layer, together with the presence of contaminants and bias, gives rise to chemical action at the anode to form positive metals ions (Sn) which then migrate to the cathode. Dentritic growth of the metal occurs from the cathode to the anode , resulting in excessive current leakage and eventually a resistive short is created, with a resulting equipment failure.

In the test carried out in the laboratory the same failures detected in the samples from the outside plant were reproduced , as shown in figures 7 and 8.

Another factor contributing to the apparition of this failure mechanism is the presence of solder flux residues on the board surface. The presence of a halogen species such as chloride ions in the solder flux residues promotes the formation of stable complexes with the metals there by helping to dissolve and transport the formation of stable complexes with the metals there by helping to dissolve and transport the metal ions. This effect was obeserved in the boards without conformal coating (as shown in Figure 7) and in the boards with conformal coating and with inadequate cleaning procedures. Under test conditions and due to the presence of solder flux residues, blistering of the conformal coating takes place which allows the access of humidity to the board surface and the formation of dendrites.

The boards with conformal coating and a previous cleaning process did not show this kind of failures which proves that the presence of humidity and contaminants on the board surface are determining factors in the apparition of this failure mechanism.

5.CONCLUSIONS

1) These tests confirmed that conformal coating products protect the PCB's in high humidity environments.

2) The cleanness process prior to the application of conformal coating is a very important factor in order to obtain good adhesion and increase their reliability in high humidity environments.

3) The MTBF of cleanness conformal coated PCB's is 15 times better than uncoated PCB's.

4) The time compressed reliability tests are a good method to obtain a quick answer that allows to evaluate differents manufacturing process and afecting factors.

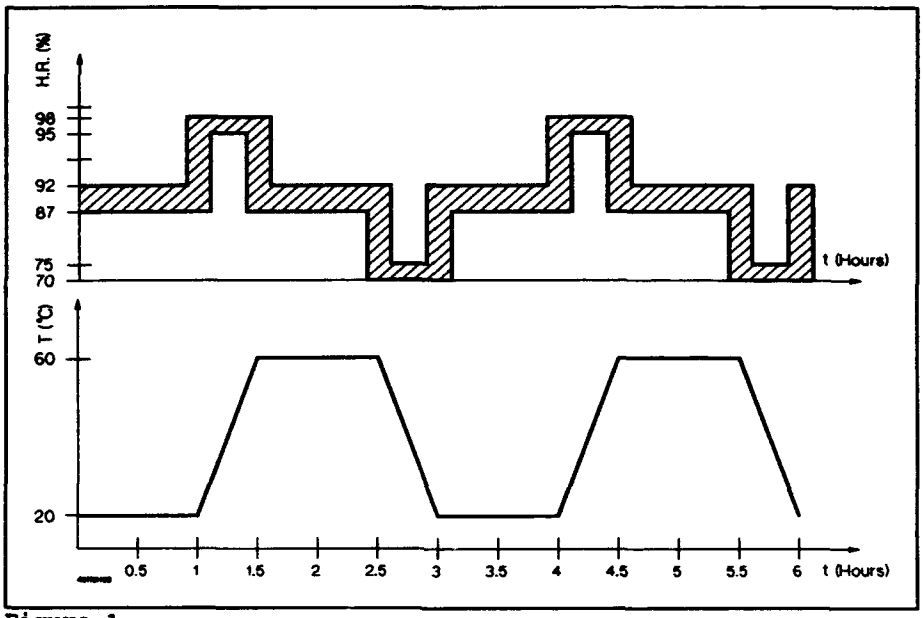


Figure 1.

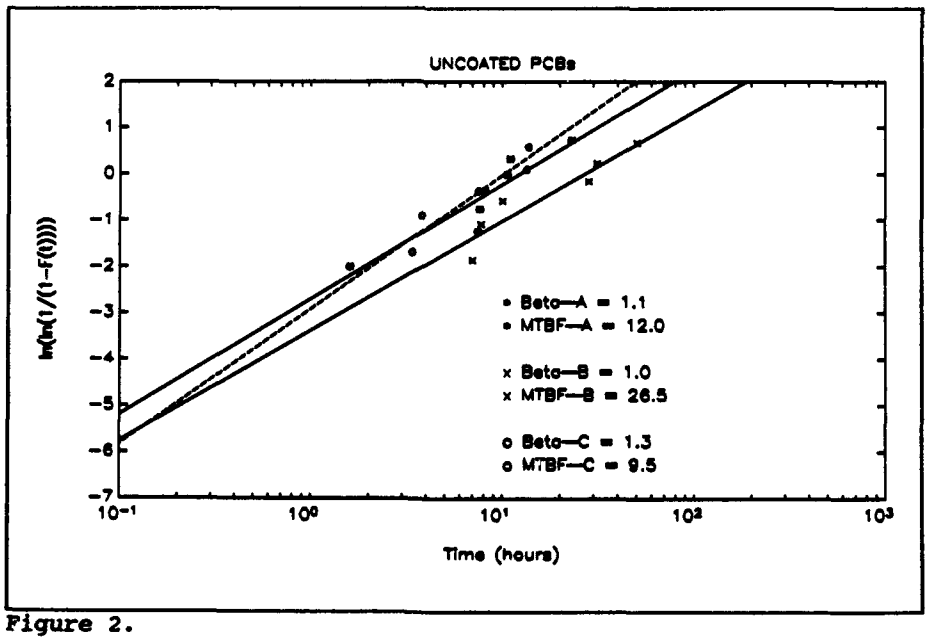


Figure 2.

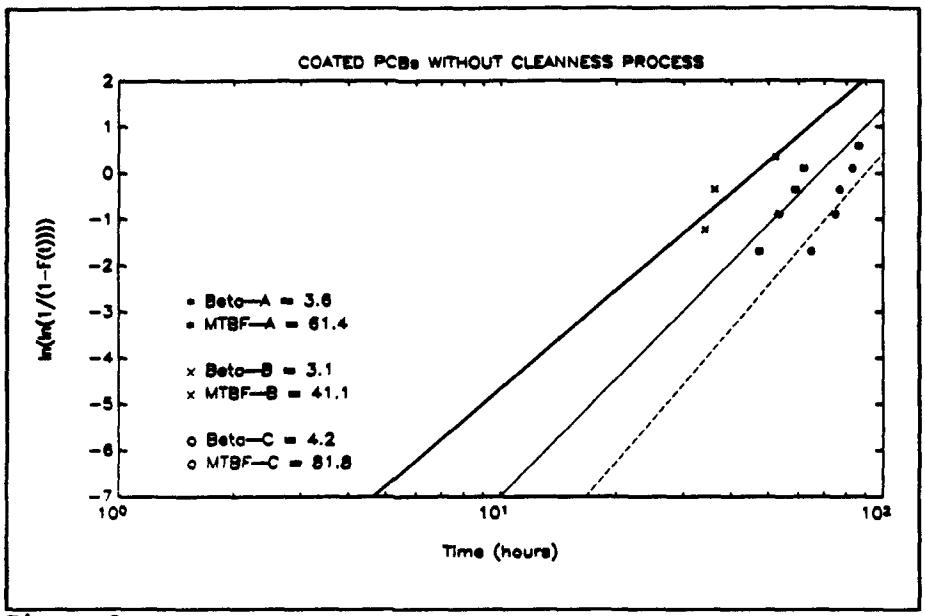


Figure 3.

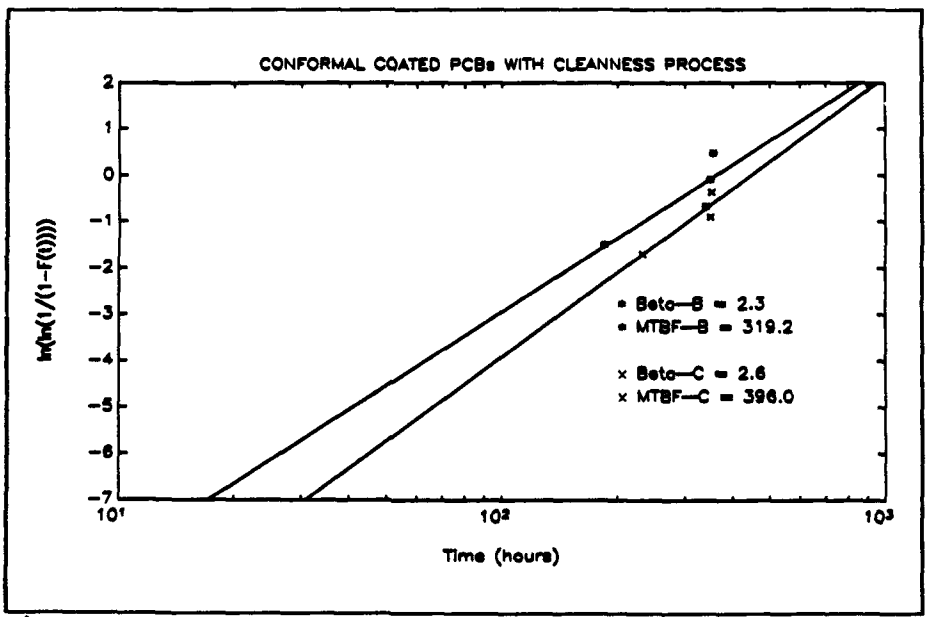


Figure 4.

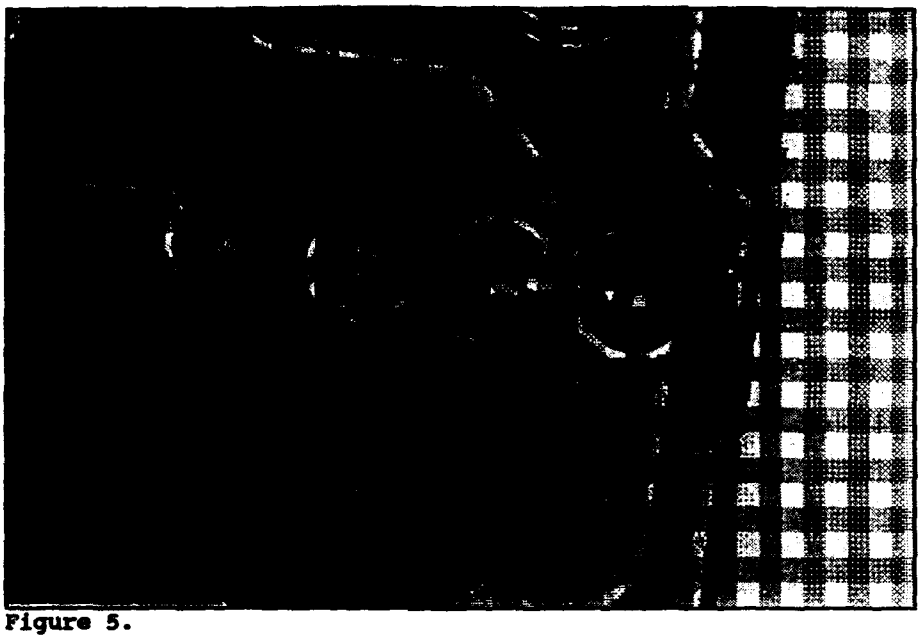


Figure 5.

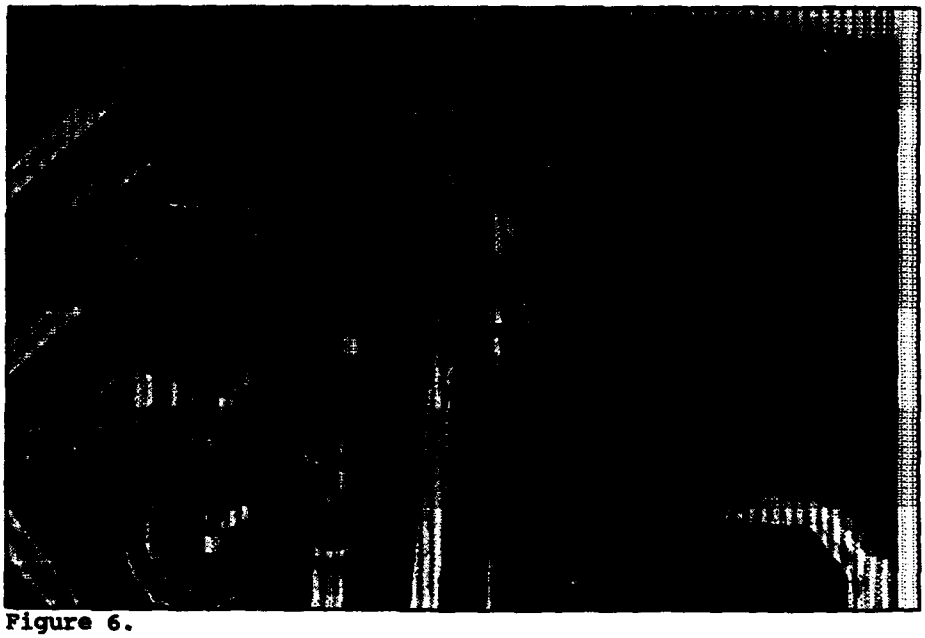


Figure 6.

Figure 7.

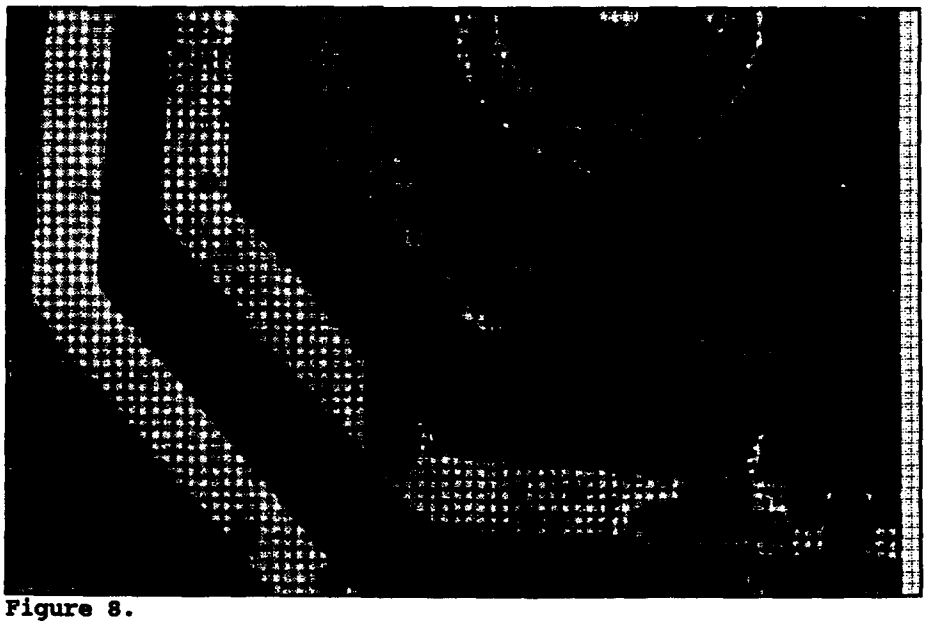


Figure 8.

OBSERVATION OF SIMULTANEOUS EXTRUSIONS AND VOIDS, IN TUNGSTEN-FILLED SUB-MICRON VIAS, INDUCED DURING HIGH-TEMPERATURE PROCESSING.

J.A.Walls MIEEE MIEE CEng
Digital Equipment Corporation, South Queensferry, Scotland.
email: jwalls@theraj.enet.dec.com

1. ABSTRACT:

Recently, Shibata et al [1] demonstrated the sensitivity of sub-micron via integrity to high-temperature processing after via etch. Initially, thermal expansion of Al into the etched via takes place; this results in deformation of the tungsten plug. Our work indicates that subsequent contraction of the Al during a period of relaxation can give rise to void formation in the via during electromigration stress. This mechanism represents a potential reliability hazard for a tungsten-plugged interconnect system.

2. INTRODUCTION:

The recent work of Shibata et al [1] demonstrated the sensitivity of sub-micron via integrity and tungsten plug formation to high-temperature processing steps after via etch. Through a length-dependency study, two effects were demonstrated. Firstly, deformation of the tungsten plug through a thermal expansion of Al into the etched via, and secondly, void formation in the inter-connecting lines due to subsequent contraction of the Al during a period of relaxation.

Recently we have observed a similar effect during experimentation, with the principle difference being the location of the voids. An anneal step was introduced immediately after the contact etch, on a sample basis, to investigate the sensitivity of our process in the light of the work by Shibata et al. In our work, voids formed in the immediate vicinty of the tungsten vias, at the M1-W interface (see figure 1), and not in the metal line as reported previously. Initially, analysis of the fail sites suggested electromigration of Al away from the M1-W interface, but closer inspection revealed that vias immediately adjacent to the fail site were similarly affected. Their location thus did not relate to electron current direction, and hence to the expected direction of mass transport.

This paper describes failure analysis techniques by which this phenomena was observed, the characteristics of this unique failure mechanism, and a model to explain how the effect of thermally-induced stress significantly reduces fail time.

3. EXPERIMENTAL PROCEDURE:

Test samples were produced using a modified production process flow; with the omission of front-end (transistor) processing and with an anneal step added after contact clean, immediately prior to plug-fill. The anneal was performed at 450 deg C in H_2 for 60 mins. The test structures employed consisted of a chain of twenty identical W-plugged M1 to M2 vias, with approximately 200 um long M1 or M2 lines linking each via. 200 um is greater that the Blech length for the metallisation scheme employed, and ensures electromigration may take place. This is longer than the threshold length of 10 um or saturation length of 100 um observed by Shibata et al for extrusion formation. M1 and M2 were composed of Al1%Si0.5%Cu, the M1 linewidth was 1.3um, and the nominal via diameter was 0.68 um.

Electromigration testing was performed at 200 C on batches of 16 packaged parts from the test sample lots. A current of 4.6 mA was applied to the test structure which was monitored in a Kelvin cofiguration for changes in resistance. At this low stress current, no significant Joule-heating was observed. A threshold of 10% change in the entire structure resistance was set as a means of detecting the onset of fatal void formation. Smaller increases in resistance could invariably be detected, but since only gross voids such as those reported previously were of interest, this higher threshold sufficed. The higher threshold also made the task of locating the voids somewhat easier during subsequent failure analysis.

Control samples were placed under similar conditions of temperature, but with zero current applied, to assess the effect of temperature alone as an acceleration factor.

4. EXPERIMENTAL RESULTS:

For the annealed samples tested, the 10% failure criterion was reached in a considerably shorter time than would have been normally expected from this structure under these test conditions. The MTTF of all samples have been normalised to an equivalent T50 to take account of differences in applied currents and test temperatures. This also allows for meaningful comparisons to be made.

The results from a single split lot are shown in table 1:

sample	normalised T50	gamma
Annealed	173 ±8 hrs	0.17 ±0.03
No anneal	555 ±61 hrs	0.36 ±0.1

Table 1: Results from split lot.

The distribution of the failures in the annealed wafer appears lognormal as can be seen from figure 2.

Additional lots were also annealed, but included various process splits known to affect T50. In all cases, however, the normalised T50's were remarkably similar.

Clearly the effect of the anneal is to activate a dominant early failure mode. The smaller gamma of the annealed samples is also suggestive of a new failure mechanism.

No failures were observed from the zero-current samples, not even after over 2000 hours of thermal stress.

5. FAILURE ANALYSIS:

The failing areas were identified as single or multiple contacts within the chain using the thermal liquid crystal technique. Failing contacts were sectioned using focussed ion beam (FIB), and on occasions some apparent via deformation was visible (see figure 3. This was initially mistaken for an artefact (oblique cut) of the FIB process due to the very occasional nature of the defect. Furthermore, the visibility of any deformation was a rapid function of FIB slice depth, and early image quality was poor.

A specialised tripod-mounted polishing jig, normally employed for TEM sample preparation, was subsequently employed. This allowed for all vias in the chain to be polished and viewed simultaneously. Also, better 'slice-depth' resolution could be achieved than was attainable by FIB. Figures 4 clearly show the deformation of each via in the chain due to the thermal expansion of aluminium into the via during the anneal step. This 'mushroom' style extrusion in the via is identical in appearance to that observed by Shibata et al. However, no immediately obvious cause for a resistance increase (EM fail) was visible.

Refininements to the FIB technique and the secondary ion (SIM) imaging enabled the extrusions to be located with greater ease. This was important since there was significant wafer-wafer and die-die variability observed. In short, the extrusions which we would have expected to find were often difficult to locate and image.

Closer inspection of some failing vias indicated some small voids located at the M1/W interface as shown in figure 5. Furthermore, voids were occuring at M1/W interfaces only. Upon comparing the current polarity to the failure site locations, however, it was revealed that the location of the voids did not always relate to electron current direction, and hence to the expected direction of mass transport. A FIB/SIM image of a failure site, and the immediately adjacent vias is shown in figure 6. This indicated that voiding was taking place independent of current direction.

FIB cross-sections of the zero-current samples were not obtained.

6. MODELLING THE FAILURE MECHANISM:

Stress-enhanced electromigration is thought to be the mechanism of failure in this case for two reasons. Firstly, the location of the voids found appears to not always relate to current direction; indeed voids were found at the 'upwind' and 'downwind' ends of the M1 lines. Secondly, the interconnect system used in this study is virtually symmetrical about the W-plug, and yet no evidence of voiding has ever been detected in the M2 immediately above the via.

Surprisingly, however, no evidence of voiding in M1 was found in the electrically un-stressed (zero curent) samples. Although this has yet to be confirmed by FIB cross-sectional analysis, no measurable resistance increases were observed. This is supportive of a two-step failure process, where the first step is taking place without an applied stress current.

Such a two-step process has been postulated previously [2] whereby time-to-failure (tf) is the summation of a nucleation time (tN) and a period of growth, or void movement (tG):

$$tf = tN + tG \quad (1)$$

The nucleation and growth phases are also considered to have different current density (j) dependencies [3]:

Nucleation: $1/j^2$ dependence

Growth: $1/j$ dependence

Normally, the fail time is dominated by a long incubation period before growth can take place, hence tf is $1/j^2$ dependent (or thereabouts) for most observations, and tf can be modelled by the Shatzkes & Lloyd relation [4]:

$$tf = A\,(T/j)^2 \exp(Ea/kT) \quad (2)$$

Our observed ~4x reduction in lifetime for this structure can be explained if the rate-limiting step is the void growth process with the the nucleation, enhanced by a pre-existing stress gradient, taking place almost spontaneously. It is proposed that the stress gradient is a result of the aluminium extrusions induced by the anneal step, and that this is sufficient to considerably shorten the first phase *in the vicinity* of the via hole.

It has been argued that mechanical stress gradient and vacancy concentration are essentially the same thing in that they can be modelled similarly in terms of their effect upon time-to-failure [5]. Our samples could be considered as having a close to critical stress gradient in the vicinity of the via hole as a result of the anneal step. This stress was insufficient to result in immediate actual void nucleation (as evidenced from cross-sections of unstressed samples), but that soon after the stress current was applied, voids formed preferentially in the regions of highest stress, namely in the M1 under the vias. Thereafter there followed a phase of void growth resulting in the failure of the via. This could, of course, occur at either the 'upwind' or the 'downwind' via as we have seen.

7. CONCLUSIONS:

Our initial analysis of the failure sites had suggested electromigration of aluminium away from the M1/W interface, but closer inspection revealed that vias immediately adjacent to the fail site were similarly affected. This suggested a new failure mechanism was at work. The anneal step after contact etch was sufficient to enhance early electromigration failure through the enhancement of the void nucleation process. Our results suggest an n=1 current dependency which can be explained by a pre-existing critical stress. Further work to evaluate the effect of current density would, of course, be required to substantiate this.

Such anneal steps are commonplace to permit the early in-line measurement of transistor parametrics, and anneal out plasma-related oxide charges. We conclude that their exact location in the processing sequence is critical to the reliability of a tungsten-plugged interconnect system.

8. ACKNOWLEDGEMENTS:

This work was conducted in conjunction with the Reliability and Failure Analysis groups of South Queensferry (Scotland), Ayr (Scotland), and Hudson (Massachusetts) wafer fabs, and with the cooperation of the South Queensferry Materials Analysis group. Thanks are also due to G.McKee and S.Pizzanello for performing the electromigration tests and assembly of the samples, and to FEI Europe Ltd for the FIB/SIM images. Thanks are also due to Terry Spooner, Jim Lloyd, and Roger St Amand for many fruitful discussions.

9. REFERENCES:

1 "Via-hole related simultaneous Stress-Induced Extrusion and Void Formation in Al Interconnects", H.Shibata, T.Matsuno, K.Hashimoto, Proceedings International Reliability Physics Symposium, pp. 340-344, 1993.

2. "Electromigration Failure", J.R.Lloyd, J.Appl.Phys., 69 (11), pp. 7601-7604, 1991.

3. Ibid

4. M.Shatzkes, J.R.Lloyd, J.Appl.Phys., 59, p 3890, 1986.

5. "Mechanical Stress and Electromigration Failure" J.R.Lloyd, MRS Symposium Proceedings, Vol 225, 1991, pp 47-52, Material Reliability Issues in Microelectronics, Eds J.R.Lloyd, F.G.Yost, P.S.Ho.

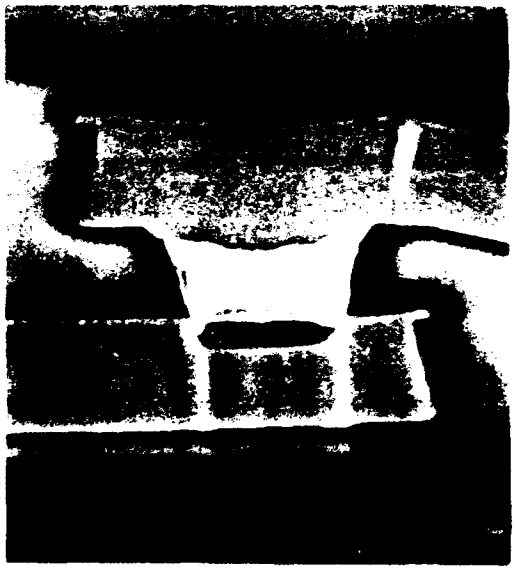


Figure 1: A typical void formed at an extrusion site.

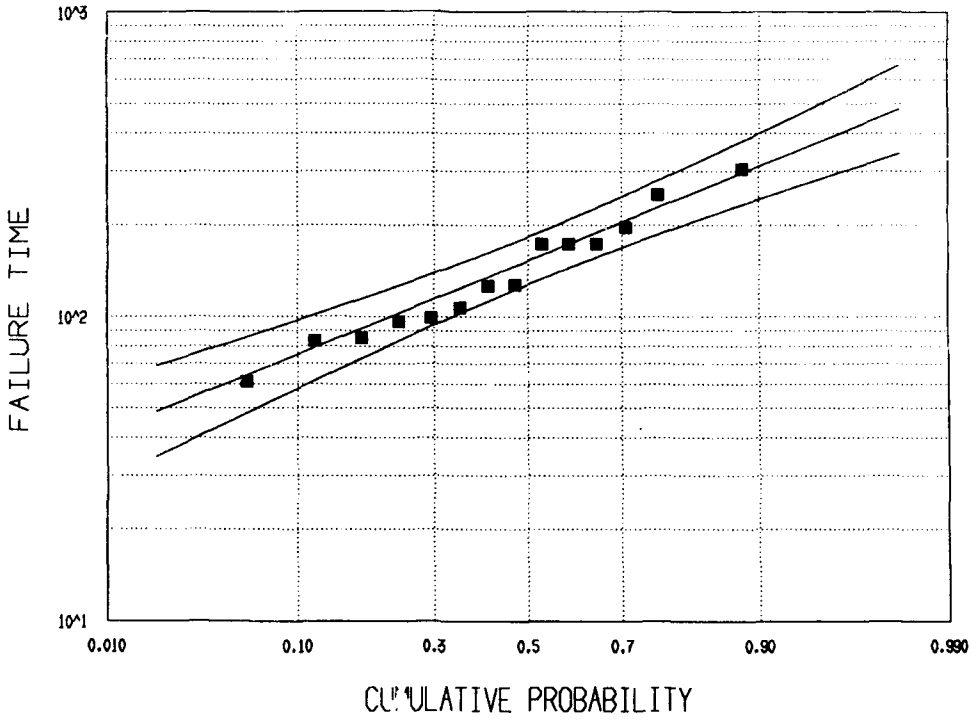


Figure 2: Cumulative fail plot for annealed sample.

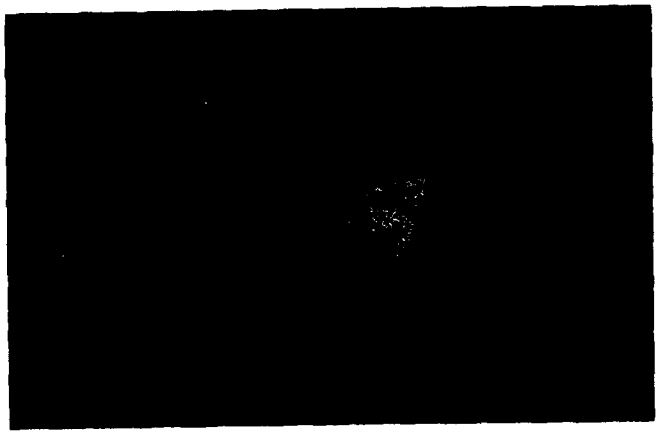


Figure 3: Tungsten via deformation due to aluminium extrusion

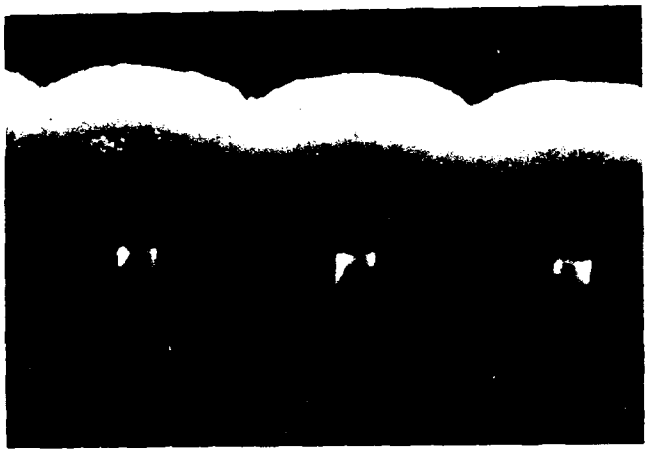


Figure 4a: Cross-sectional SEM from tripod-polished sample which had been subjected to the anneal after contact etch.

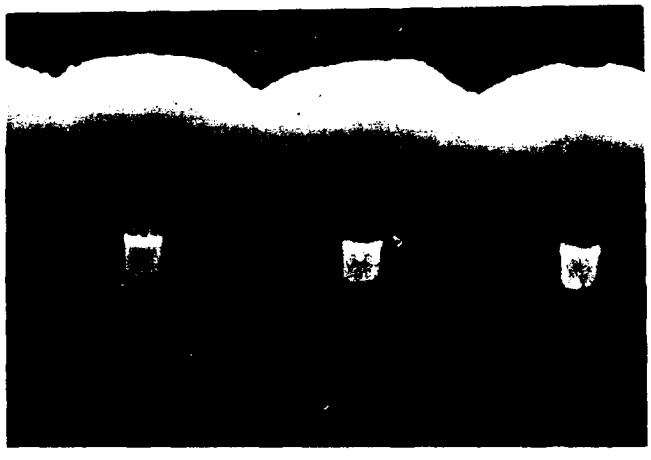


Figure 4b: Cross-sectional SEM from tripod-polished sample without anneal step (for comparison with 4a).

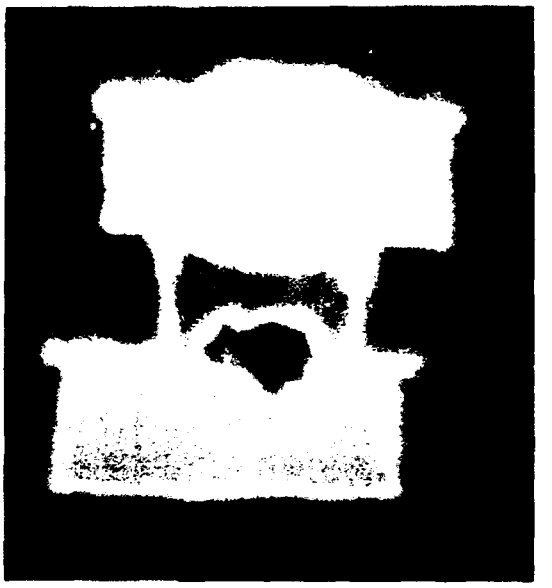


Figure 5: SIM/FIB image of failing via showing simultaneous extrusion and voiding of the aluminium interconnect.

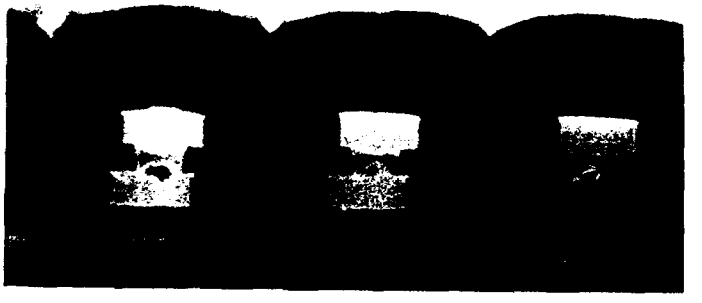


Figure 6: SIM/FIB image showing presence of multiple voids in vias adjacent to the failure site shown in figure 5.

COMPUTER ANALYSIS OF MICROSCOPIC ELECTROMIGRATION FAILURE MODELS

J T Trattles, A G O'Neill and B C Mecrow
Department of Electrical and Electronic Engineering
Merz Court
University of Newcastle upon Tyne
NEWCASTLE UPON TYNE
NE1 7RU
UK

1. ABSTRACT

Time-to-failure and resistometric electromigration reliability models used in the analysis of experimental data to determine interconnect reliability are compared using a new computer model for simulating failure in polycrystalline thin film conductors. The inclusion of backflux in the model and the use of ambient or conductor temperature on the extracted failure parameters is illustrated and the need for including backflux in future electromigration studies is demonstrated. The limitations and unsuitability of time-to-failure and resistometric models for extracting certain material parameters is highlighted and a correlation between the time-to-failure and the early stages of the rate of relative resistance changes is shown.

2. INTRODUCTION

Electromigration damage of metal conductors is a major cause of failure in integrated circuits. Over the past 30 years or so a number of analytical and empirical models have been used in an attempt to understand the processes involved and provide design rules that increase the reliability of the conductor. Unfortunately, as a result of the numerous processes involved in the conductor deterioration, it has been difficult to determine what, if any, physical parameters can be extracted from the experimental data. A new computer model for simulating electromigration damage in thin film polycrystalline conductors (Ref. 1) is used to analyse and compare three time to failure models (based on different assumptions) (Refs. 2-4) and the classic resistometric failure model (Ref. 5). The failure models analysed are described in section 3 and the new computer model is briefly outlined in section 4. Results from the simulations are discussed in section 5.

3. ELECTROMIGRATION FAILURE MODELS

The lifetime test of metal conductors is the most common method used for determining the resistance to electromigration damage. Due to the statistical nature of the results they are usually presented in terms of the time-to-failure (TTF) of a group of conductors. This is usually expressed as (Ref. 2)

$$TTF = Aj^{-n} \exp(\frac{E_a}{k_B T}) \qquad (1)$$

where A is a constant, T is the conductor temperature, k_B is the Boltzmann constant, j is the current density, n the current density exponent and E_a is the activation energy of the process. Equation (1) is usually called Black's equation and is used to provide design rules once n and E_a are determined.

The uncertainty in the true value of the current density exponent n (which is usually taken to be either = 1 or 2) has led to this model has being modified by several authors. Shatzkes & Lloyd (Ref. 3) considered both diffusion and electromigration to produce a modified version of Black's equation with a current density exponent of 2 and an additional temperature term. This is expressed as

$$TTF = BT^2 j^{-n} \exp(\frac{E_a}{k_B T}) \qquad (2)$$

where B is a constant. However, the assumptions used in deriving equation (2) neglect the void growth process and assume that the period required for the void to incubate is the dominant failure process: if the void growth dominates then $TTF \propto T/j$ (Ref. 6).

Another version of Black's equation which assumes void growth is the dominating failure mechanism was produced by Walter (Ref. 4). In this model the current density exponent was taken to be 1 while the temperature coefficients of electrical resistivity and effective charge (α and β respectively) were introduced. This is expressed as

$$TTF = \frac{Cj^{-n}}{(1+\alpha T)(1+\beta T)} \exp(\frac{E_a}{k_B T}) \qquad (3)$$

where C a constant.

The resistometric method was first used by Rosenberg and Berenbaum (Ref. 5) who related changes in the resistance in the early stages of conductor deterioration with damage primarily as a result of electromigration. This is a method commonly used for determining activation energies, although there have been some doubts to the validity of this approach (Ref. 7). Many authors have related the relative change in resistance due to electromigration to a term similar to that used in time-to-failure analysis, i.e.

$$R_{rc} = \frac{1}{R}\frac{\partial R_{em}}{\partial t} = Dj^{n}\exp(-\frac{E_a}{k_B T}) \quad (4)$$

where D is a constant.

4. COMPUTER MODEL

The new computer model uses a coupled finite element model to produce a self-consistent solution to the electrical and thermal conduction equations in two dimensions. These are used to provide realistic current density and temperature distributions required for accurately simulating electromigration damage. Unlike previous models (Refs. 4,6,8,9) diffusion and stress driven backfluxes are included for the first time. Backfluxes are calculated explicitly in each of the grain boundaries using concentration and stress gradients resulting from the initial electromigration flux. The stress dependent diffusivity term is also included directly in the formulation. It is assumed that the main cause of the flux divergence is the grain structure of the conductor and that these divergences occur at the triple point junctions of the grain boundaries.

The model solves the transport equation for atomic migration along a grain boundary in the presence of an electric field, stress gradient and concentration gradient:

$$J_{gb} = D_{gb}\left(\frac{N_{gb}Z^{*}_{gb}e\rho j_o \cos\psi}{k_B T} - \nabla N_{gb} - \frac{N_{gb}\Omega}{k_B T}\nabla\sigma_n\right) \quad (5)$$

where the diffusion constant D_{gb} is given by

$$D_{gb} = D^{o}_{gb}\exp\left(-\frac{E_a + \Omega_a\sigma_n}{k_B T}\right)\sin\left(\frac{\theta}{2}\right) \quad (6)$$

and the other terms in (5) and (6) are:
D^{o}_{gb} - constant of diffusion
N_{gb} - atomic density
Z^{*}_{gb} - effective charge
ρ - electrical resistivity
j_o - applied current density
T - temperature
E_a - activation energy for diffusion
Ω - atomic volume
$\nabla\sigma_n$ - stress gradient
∇N_{gb} - concentration gradient
Ω_a - activation volume
σ_n - hydrostatic stress
θ - angle of mismatch between adjacent grain lattices
ψ - grain boundary/current flow angle of orientation
gb - specifies grain boundary values

The angle of mismatch θ between adjacent grain lattices is a randomly generated between 0 and 60 while the angle of orientation between the grain boundary and current flow is determined by means of a scalar product.

The model uses two iterative processes, one for calculating the grain boundary fluxes and another for the finite element calculations (Ref. 10). This is due to the computational expense of the finite element calculations and the stability of the backflux calculations which require short time steps. Conductors are assumed to fail when the temperature reaches the melting value when the conductor will break resulting in an open circuit. While the choice of failure criterion is rather arbitrary it is simple to implement.

5. RESULTS

A set of five conductors with different grain structures have been used in a simulation of conductor deterioration and failure. The conductors were subjected to stress conditions typical of those used in experimental test conditions. The results from these simulations have been analysed using the reliability models presented in equations (1) - (4). These are commonly used to provide reliability design rules for integrated circuit interconnections.

The simulated deterioration and failure of polycrystalline structures is similar to that observed experimentally (Refs. 5,11,12). This is illustrated in figure 1 which shows the final three finite element time steps of a conductor undergoing failure with the stress condition of 5×10^{10}A/m^2 at 175°C (the darker regions indicate where the void is growing). This highlights the rapid rate of deterioration immediately prior to failure and the unpredictability in the location of final failure. Even when the conductor is close to failure (upper image) the expected failure location is near the left of the conductor. However, this is not the case and there is a rapid deterioration on the right of the conductor where the failure finally occurs.

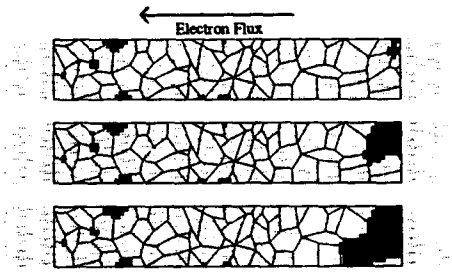


Figure 1. Simulated failure in a polycrystalline conductor

Figure 2 shows the simulated resistance change as a function of time in the period terminating immediately before failure. This illustrates that there are periods where the conductor degradation is halted and resistance decreases. There are also cases where there is a sudden increase in conductor resistance which just as suddenly decreases.

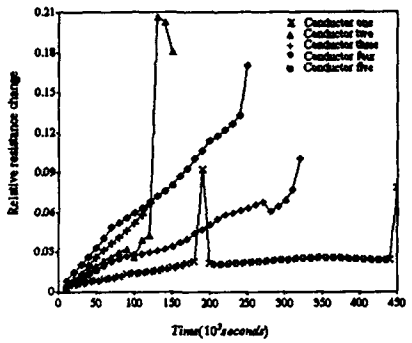


Figure 2. Simulated relative resistance change as a function of time for five conductors with the same temperature/current density conditions.

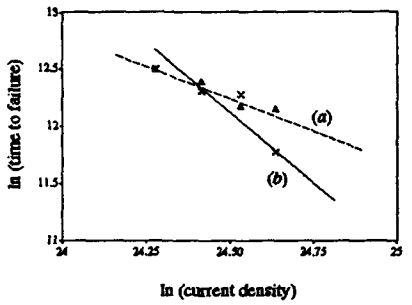


Figure 3. ln(TTF) as a function of ln(j) for simulations without backfluxes (line (a)) and with backfluxes (line(b)).

The impact of the inclusion of backflux calculations in the simulation model is demonstrated by the estimate of the current density exponent. This is illustrated in figure 3 which shows a typical plot of ln(*TTF*) as a function of current density ln(*j*). When backflux calculations are not included a value of approximately 1.04 (figure 3 line (a)) is predicted for the current density exponent in all time-to-failure models. This is in agreement with other simulations (Refs. 4,6,8,9) but does not agree with values found in experimental studies of approximately 2 (Refs. 13,14). However, when the backflux calculations are included in the simulations a value of approximately 2.3 is obtained (figure 3 line (b)).

The analytic model of Shatzkes & Lloyd (Ref. 3) obtained a value of 2 (see equation 2) by assuming the void incubation period is the dominant factor in conductor deterioration. However the simulated conductor deterioration is dominated by the void growth period, in fact there is no void incubation present in these simulations. Under these conditions a value of 1 would be expected (Refs. 6,8), however, a value of 2.3 is obtained. This suggests that the assumption of a void incubation period is not necessary to obtain agreement between theory and experiment: the key to the interpretation of the data is the inclusion of backflux in the analysis.

Comparison of equations (2) and (4) suggest a relationship between the time-to-failure (TTF) and the early rate of relative resistance changes (R_{rc}). Maiz and Segura (Ref. 11) suggested the use of a distribution of relative resistance changes when comparing time-to-failure with resistance changes. Using this method they obtained a relationship of the form $TTF=CR_{rc}^{-m}$ where *m* and *C* are contants. When all simulated time-to-failure (TTF) and rate of relative resistance changes (R_{rc}) are considered together a correlation between them is found . This is illustrated in figure 4 which shows a plot of $\ln(R_{rc})$ as a function of ln(TTF) for all the data obtained in the simulations. The form of this relationship is $TTF=0.223R_{rc}^{-1.11}$ and indicates that resistance measurements can be used in producing lifetime parameters for use in providing reliability rules for conductor design.

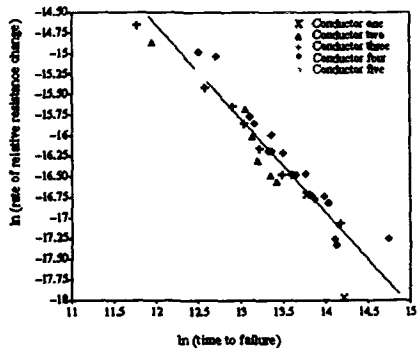


Figure 4. Relationship between the early rate of relative resistance change and the time-to-failure.

Table 1 shows the average activation energy predicted using the time-to-failure models (Refs. 2-4) and the classic resistometrc method (Ref. 5) with both ambient and conductor temperatures. Comparing the estimated activation energies with the default value of 0.51eV used in the simulations clearly demonstrates the dangers in relating the time-to-failure and resistometric activation energies with that associated with diffusion. Time-to-failure analysis is less accurate than the classic resistometric method (although the model of Shatzkes and Lloyd (Ref. 3) is only slightly less accurate than the resistance model). However, even the resistometric model produces an estimate (0.45eV) which is still considerably lower than the default value (0.51eV). The use of the conductor temperature in estimating the activation energy exponent merely increases the value in all failure models (Refs. 2-5). In general, this is closer to the activation energy for diffusion however this is not always the case: Conductor 3 analysed using the model of Shatzkes and Lloyd (Ref. 3) produced estimated activation energies of 0.50eV and 0.57eV using ambient and conductor temperatures respectively.

Table 1. Activation energy E_a estimated using time-to-failure (TTF) and Classic Resistometric analyses

Temperature Used	Model Type	Activation Energy (eV)	reference
Ambient	TTF	0.29	[2]
Conductor	TTF	0.33	[2]
Ambient	TTF	0.37	[3]
Conductor	TTF	0.43	[3]
Ambient	TTF	0.30	[4]
Conductor	TTF	0.35	[4]
Ambient	Resistometric	0.39	[5]
Conductor	Resistometric	0.45	[5]

6. SUMMARY AND CONCLUSIONS

A new computer model for simulating failure in polycrystalline thin film conductors has been used to study three time-to-failure (TTF) and the classic resistometric electromigration reliability models. Unlike previous computer models (Refs. 4,6,8,9) diffusion and stress driven backfluxes have been included for the first time. The results demonstrate the limitations and unsuitability of the time-to-failure and resistance models in extracting certain material parameters (e.g. thermal activation energy) and highlighted the need for including backflux in the simulation model.

A direct correlation between the time-to-failure and rate of relative resistance change was found when all conductors are considered together. Also, the variance in the results obtained using the resistometric method (0.04 for E_a and 0.17 for n) are smaller than those obtained using the time-to-failure models (these are, for E_a and n respectively, 0.09 & 1.03 using equations 1, 2 and 3). This suggests that resistance measurements can be used in producing lifetime parameters for use in providing reliability rules for conductor design.

The use of ambient or conductor temperature on the extracted failure parameters has also been studied. While, in general, the use of the conductor temperature give more accurate estimates of the activation energy for diffusion this is not always the case.

7. ACKNOWLEDGEMENTS

Dr J T Trattles would like to thank the Science and Engineering Research Council (SERC) for providing a research studentship to finance this study. Thanks also go to Dr C M Johnson (University of Newcastle) for allowing time for the preparation and presentation of this paper.

8. REFERENCES

1) Trattles J T, O'Neill A G & Mecrow B C, "Computer Simulation of Electromigration in Thin Film Metal Conductors", *J. Appl. Phys.*, to be published (1994)

2) Black J R, "Electromigration-A Brief Survey and Some Recent Results", *IEEE Trans. ED*, **16**, pp 388-347, (1969)

3) Shatzkes M & Lloyd J R, "A Model for Conductor Failure Considering Diffusion Concurrently with Electromigration Resulting in a Current Exponent of 2", *J. Appl. Phys.*, **59**, pp 3890-3893, (1986)

4) Walter M J, "Implications of a Two Dimensional Description of Electromigration in Aluminum", *Proc. IEEE V-MIC86*, pp 173-180, (1986)

5) Rosenberg R & Berenbaum L, "Resistance Monitoring and Effects of Nonadhesion during Electromigration in Aluminum Films", *Appl. Phys. Lett.*, **12**, pp 203-204, (1968)

6) Kirchheim R & Kaeber U, "Atomistic and Computer Modeling of Metallization Failure of Integrated Circuits by Electromigration", *J. Appl. Phys.*, **70**, pp 172-181, (1991)

7) Schreiber H-U & Grabe B, "Electromigration Measuring Techniques for Grain Boundary Diffusion Activation Energy in Aluminum", *Solid State Electronics*, **24**, pp 1135-1146, (1981)

8) Marcoux P J, Merchant P P, Narodisky V & Rehder W D, "A New 2D Simulation Model of Electromigration", *Hewlett-Packard Journal*, pp 79-84, June 1989

9) Smy T J, Winterton S S & Brett M J, "A Monte Carlo Computer Simulation of Electromigration", *J. Appl. Phys.*, **73**, pp 2821-2825, (1993)

10) Trattles J T, "Finite Element Simulation of VLSI Interconnections with Application to Reliability Design Optimisation and Electromigration Modelling", PhD Thesis, University of Newcastle upon Tyne (UK), (1993)

11) Maiz J A & Segura I, "A Resistance Change Methodology for the Study of Electromigration in Al-Si 'nterconnects", *Proc. IRPS'88*, pp 209-215, (1988)
12) Greer A L & Shih W C, "Microstructure and the Development of Electromigration in Narrow Interconnects", *Proc MRS Symposium*, **265**, pp 83-94,(1992)
13) Schafft H A, Grant T C, Saxena A N & Kao C-Y, "Electromigration and the Current Density Dependence", *Proc. IRPS'85*, pp 93-99, (1985)
14) Shih W C, Denton T C & Greer A L, "A Microscopical and Statistical Study of Electromigration Damage and Failure in Al-4wt.%Cu Tracks", *Proc. MRS Symposium*, **265**, pp 101-106, (1992)

SHAPE CHANGES OF VOIDS IN BAMBOO LINES: A NEW ELECTROMIGRATION FAILURE MECHANISM

O. Kraft, U.E. Möckl and E. Arzt

Max-Planck-Institut für Metallforschung and Institut für Metallkunde der Universität Stuttgart

Seestr. 71, 70174 Stuttgart, Germany

1. ABSTRACT

The behaviour of electromigration-induced voids in narrow, unpassivated aluminum interconnects is examined. Failure voids are categorized in two different types: extended wedge-shaped voids and narrow slit-like voids. The occurrence of slits is found to increase with decreasing current density and line width. It is observed that shape changes of voids lead to the formation of the slit-like voids. These shape changes may consume a major part of the lifetime of a conductor line. A model to describe the evolution of the void shape is presented.

2. INTRODUCTION

Failure of metallic interconnects is considered to be a key concern in the reliability of microelectronic devices, which is further accentuated by the continuing trend toward miniaturization. The phenomenon of electromigration is one possible reason for such failures: transport of matter driven by high electric currents produces damage in the lines (e.g. Ref. 1). Failure then results from voids growing over the entire line. In the last years, microscopy studies have revealed that these voids are able to move along the conductor line (Refs. 2-6) and more recently, it was reported that in bamboo interconnects voids grow transgranularly with a slit-like morphology (Refs. 7-10). Further it was observed that voids which produced an open circuit show a specific asymmetric void shape. In contrast to the contour of the void at the anode end, the cathode boundary is flat and more or less perpendicular to the line (Ref. 11). Interrupted electromigration tests revealed that in addition to void nucleation, growth, and motion, void shape changes can consume a major part of the lifetime of a conductor line and may be the decisive mechanism in the formation of a slit-like transgranular void. It is the aim of this article to elucidate the conditions under which slit voids occur and to identify the driving forces behind the shape change.

3. EXPERIMENTAL

Two different sets of samples, pure Al and Al-0.5%Cu-1.0%Si, were provided from different manufacturers. The films had been sputter deposited onto thermally oxidized silicon wafers. Film thicknesses were 0.5 µm (pure Al) and 0.85 µm (AlSiCu). Parallel line arrays of 20 lines with a length of 1 mm and line widths in the range of 1-5 µm were patterned using standard lithography and etching techniques. Prior to patterning, the Al film was annealed in forming gas at 400°C for 45 minutes, resulting in a grain size of 1.4 µm. The AlSiCu samples were annealed in forming gas after patterning at 480°C for 20 minutes resulting in a grain size in the continous film of 1.9 µm. All samples were tested without passivation.

Electromigration testing was performed on wafer level using two different types of experimental set-up: probe station including a hot chuck with post-

Fig. 1: SEM micrograph exemplifying an extended wedge-shaped void in a pure Al line. Testing conditions: line width 1.8 µm, current density 1.2 MA/cm^2, and temperature 227°C. Electron flow from right to left.

Fig. 2: SEM micrograph showing a slit-like void in an AlSiCu line. Testing conditions: line width 1.9 µm, current density 1.2 MA/cm^2, and temperature 227°C. Electron flow from right to left.

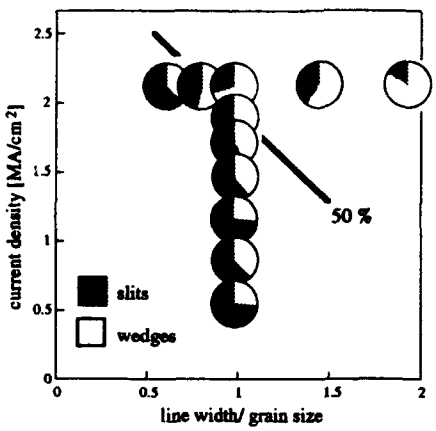


Fig. 3: Experimental electromigration damage map for AlSiCu: the fraction of slit failures is shown as a function of current density and the ratio of line width to grain size.

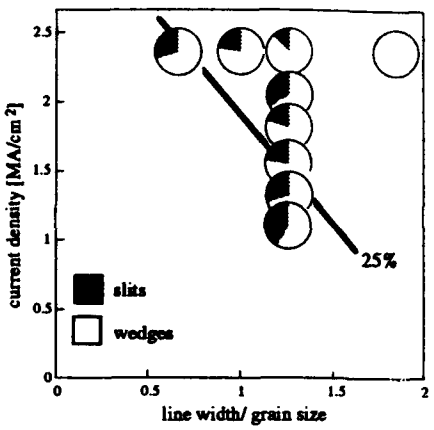


Fig. 4: Experimental electromigration damage map for Al.

testing failure analysis and in-situ testing in an SEM with a specially designed prober module. Tests were performed under constant voltage conditions with current densities in the range of 0.7 to 2.5 MA/cm^2, at a temperature of 227°C. Further experimental details and results can be found in Ref. 12.

4. ANALYSIS OF VOID DAMAGE

The detailed investigation of several hundred failure sites suggested a differentiation of fatal voids into at least two different categories, which are exemplified in Figs. 1 and 2: a more or less extended wedge shaped void (Fig. 1), or a narrow slit-like void (Fig 2). The voids were found to have a characteristic asymmetry (Ref. 11), where the boundary of the void at the cathode end is perpendicular to the line. Wedge voids are usually found to be intergranular, while slits are apparently transgranular. Systematic trends, consistent with other reports (Refs. 8 and 13), for the occurrence of slit failures were observed. These observations are summarized in experimental damage maps in Figs. 3 and 4 where the relative fractions of failures due to slit- and wedge-shape voids are shown as a function of current density and the ratio linewidth to grain size. It is shown that the importance of slit failures increases with continuing miniaturization (decreasing line width) and decreasing current densities. The latter corresponds to the transition from accelerated testing conditions to use conditions and might severely affect the

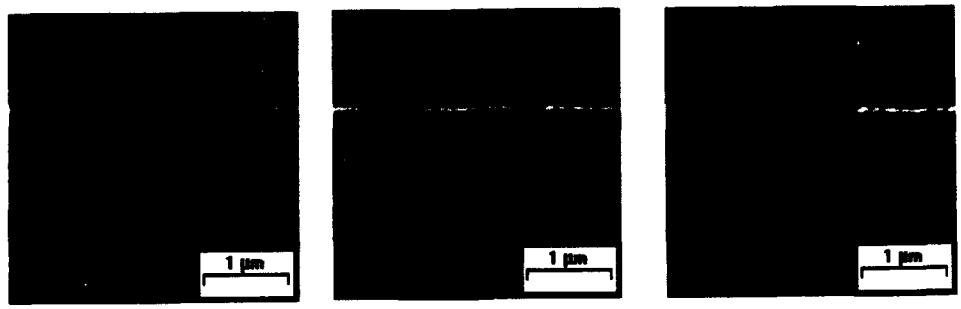


Fig. 5: SEM micrographs showing a time sequence of void nucleation (left), growth (middle), and ultimate failure due to shape change (right). Testing conditions: line width 1.2 µm, current density 1.75 MA/cm^2, and temperature 225°C. Electron flow was from right to left.

prediction of time-to-failure under use conditions from accelerated tests.

Fig. 5 shows an example of the formation of a slit-like failure. Typically nucleation occurs at the edge of the conductor line as shown in Fig. 5a. The subsequent growth is observed to be linear with time, but the ultimate failure is very often related to a void shape change. The void in Fig. 5b has a more or less rounded void tip and a measured size of 0.24 μm^2 (void area on the micrograph). The void then extends across the line by becoming narrower without significant further void growth (Fig. 5c, void area 0.26 μm^2). The final shape of the void exhibits the asymmetry described above. In addition the void moves slowly over a distance of 0.6 μm (Fig. 5a to b) and 0.1 μm (Fig. 5b to c) against the electron flow direction.

Many of the slit-like voids were found to start from small voids at the edge of a line, as shown in Figs. 2 and 5. It is striking that the growth direction of such a slit is either perpendicular to the electron flow or more often tilted against the electron flow. Slits which grew in the direction of the electron flow were very rarely observed.

5. MODELING OF SHAPE CHANGES

Our results suggest that surface diffusion on the void surfaces is responsible for the void shape, which can lead to narrow slit failures as shown in Fig. 5. An analytical treatment of void motion and shape changes by diffusion on the void surface is presented in Ref. 14. It has been shown that a semi-circular void in an infinite half-space does not change its shape under the action of an electron wind, but moves along the edge of the half-space against the electron flow. In contrast, a void with an semi-elliptical shape transverse to the electron flow direction suffers a shape distortion which tends to produce a flatter face at the cathode end and a slight lateral growth of the void. This result is in good qualitative agreement with our experimental observations. The analytical treatment is however limited to simple initial shapes and cannot predict a full temporal evolution of the void shape. In addition it is not possible to treat a void in a narrow line, because "edge" effects cannot be included in the analytical treatment.

Therefore a numerical model was developed which describes the motion and shape evolution of a "two-dimensional void" in an isotropic medium when subjected to a high current density. Current density, temperature distribution and resulting thermal stress in the vicinity of a void are calculated by application of a commercial finite element code (ANSYS, Swanson Analysis Systems, Inc.). Subsequently the void shape changes by surface diffusion are examined using a finite difference scheme, which includes the influence of gradients in curvature along the void surface. The drift velocity v_n of a point on the void surface (which was used as a node in the previous finite element calculation) can be calculated using Eqs. 1 and 2:

$$v_n = \frac{\delta D_s}{kT} \frac{\Delta}{\Delta s}\left[eZ^* \rho j_s + \frac{\Delta \mu}{\Delta s} \right] \qquad (1)$$

and

$$\Delta\mu = -\Omega (\gamma \Delta\kappa + \Delta\sigma) \qquad (2)$$

where D_s is the surface diffusivity, δ the thickness of the surface phase, eZ^* the effective charge, ρ the resistivity, j_s the surface component of the current density, μ the local chemical potential, and Δs the arc length along the surface. Ω is the atomic volume, γ the surface energy, κ the curvature of the surface,

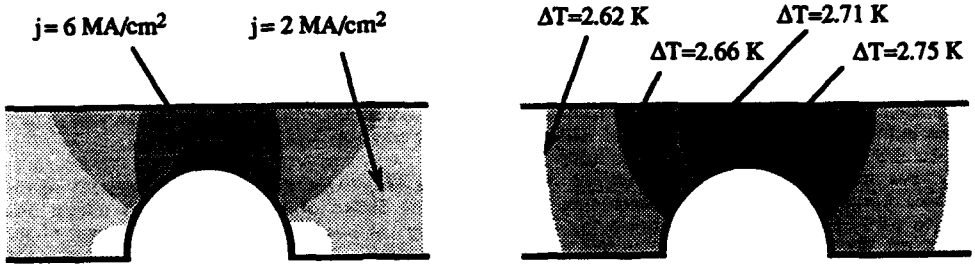


Fig. 6: Results of the finite element calculations: current density distribution in the vicinity of a semi-circular void with a radius of 0.6 μm in a 1 μm wide Al line and a applied current density of 2 MA/cm² (left), and resulting temperature distribution (right).

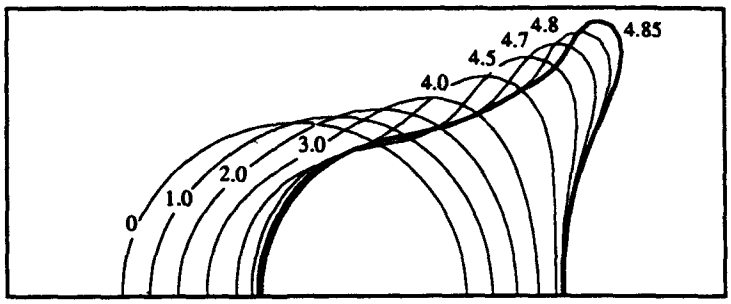


Fig. 7a: Time sequence showing a void shape change as calculated by the numerical simulation. The time is given in arbitrary units. The height of the frame is equal to the line width used in the simulation.

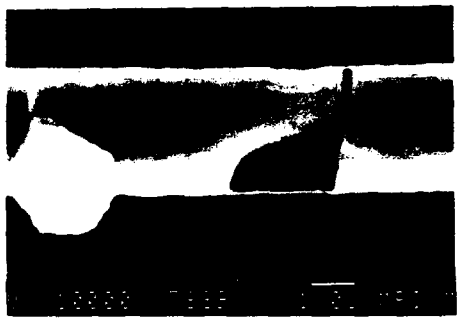


Fig. 7b: SEM micrograph of a fatal void in an AlSiCu line, showing the predicted asymmetric void shape.

surface, and σ the local hydrostatic stress.

As an example of this investigation, Fig. 6 shows the behaviour of an initially semi-circular void, with a radius of 0.6 μm in a 1.0 μm wide line Al stressed with a current density of 2 MA/cm^2. Fig. 6a shows the distribution of the current density and Fig. 6b of the temperature in the vicinity of the void. As expected, the highest values for current and temperature were obtained at the void tip. Consequently the highest compressive stresses were calculated in this region. It turned out that a compressive stress is much more critical for failure by shape change than a tensile stress, because the chemical potential is raised and this favors diffusion away from the tip. The full temporal evolution is shown in Fig. 7a and compared with a fatal void in a AlSiCu line (Fig. 7b). As a result of the numerical calculation two important features are obtained: first, the void moves in the opposite direction from the electron wind (as experimentally observed, e.g. Ref. 5). Second, the void exhibits a shape change which produces the characteristic asymmetry. The ultimate failure of the line is due to the rapid growth of a slit at the cathode end of the void. The void volume does not change in this simulation.

6. DISCUSSION

The extended failure analysis shows that the occurrence of slit voids depends strongly on the microstructure of the film, line width and testing conditions. These dependencies can be rationalized as shown in Ref. 15, where a theoretical damage map (similiar to Fig. 8) was developed. Comparing our experimental maps with the theoretical map, it is encouraging that the simplified approach which was used to obtain the theoretical map leads to qualitative agreement with the experimental observations. It is striking that the occurrence of slit-like voids in the pure Al is shifted to lower current densities and line width/grain size ratios in comparison to the alloyed metallization. It is conceivable that the addition of Cu slows down the formation and/or growth of wedge-type voids and enhances the lifetime under testing conditions. But the Cu and Si additions seem to favor the formation of slit-like voids, which will be the dominant failure mechanism under use conditions.

Obviously the initial circular shape of the void is stabilized by the surface tension. In the calculations the surface energy was chosen as $\gamma = 1$ J/m^2. Using a smaller surface energy would lead to a more rapid failure and a narrower slit in the final void shape. In order to avoid the formation of slit voids, a high surface energy would be favourable. It can be argued that, in fact, the additions of Cu and Si might lower the surface energy by segregation to the void surface and promote the formation of a slit.

These considerations have several important aspects for the design of new metallization alloys. The strategy of raising the mechanical strength of the conductor line in order to increase the electromigration resistance (Refs. 16-18) might be limited. Certainly, strengthening the alloy suppresses voiding under testing conditions, but could increase the importance of slit formation under use conditions. It appears that optimizing an alloy for accelerated test conditions does not guarantee an optimum alloy under use condditions. Following the results of the

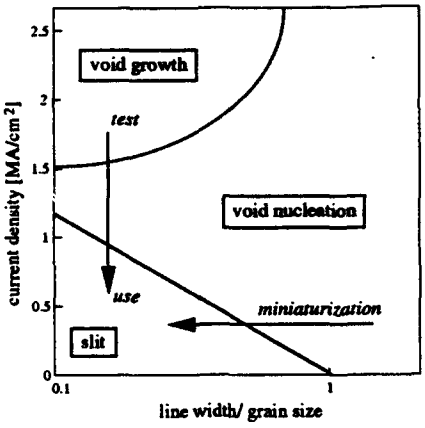


Fig. 8: Schematic of a theoretical electromigration damage map for near-bamboo lines (after Ref. 15): The upper curve delineates the area for which void growth or void nucleation control the time-to-failure. The lower curve marks the transition to failure by slit formation.

present simulations, the decisive material property would be a high surface energy to stabilize a round void shape and to avoid slit formation.

7. SUMMARY

1. An experimental damage map has been presented, which shows a trend towards suppression of wedge-type voids with decreasing current density and line width/grain size ratio. Under these conditions slit-like voids are more likely to form. The transition from one void type to the other is shifted for pure Al to lower current densities and narrower lines.

2. A model which assumes that diffusion on the surfaces of the void is the decisive mechanism for shape changes was able to successfully reproduce the typical asymmetric shape of fatal voids.

3. Reliable extrapolations of lifetime data from accelerated laboratory conditions to use conditions have to consider the change of the failure mechanism. In the same way, a uniform strategy for new metallization alloys will have to include these effects.

8. REFERENCES

1. Hummel R.E. "Electromigration and Related Failure Mechanisms in Integrated Circuit Interconnects", International Materials Reviews, in press

2. Thomas R.W. and Calabrese D.W. "Phenomenological Observations on Electromigration," 21st Annual Proc. of IEEE IRPS 1983, pp. 1-9.

3. Shingubara S. and Nakasaki Y. "Electromigration in a Single Crystalline Submicron Width Aluminum Interconnection", Appl. Phys. Lett. 58, 42-44 (1991)

4. Castano E., Maiz J., Flinn P.A. and Madden M.C. "*In Situ* Observations of DC and AC Electromigration in Passivated Al Lines" Appl. Phys. Lett. 59, 129-131 (1991)

5. Besser P.R., Madden M.C. and Flinn P.A. "*In Situ* Scanning Electron Microscopy Observation of the Dynamic Behavior of Electromigration Voids in Passivated Aluminum Lines" J. Appl. Phys., 72 (8), 3792-3797 (1992)

6. Madden M.C., Abratowski E.V., Marieb T. and Flinn P.A. "High Resolution Observation of Void Motion in Passivated Metal Lines under Electromigration Stress", MRS Symp. Proc., Vol 265, 33-38 (1992)

7. Rose J.H. "Fatal Electromigration Voids in Narrow Aluminum-Copper Interconnect," Appl. Phys. Lett., 61 (18), 2170-2172 (1992)

8. Rose J.H. and Spooner T. "The Microstructural Nature of Electromigration and Mechanical Stress Voids in Integrated Circuit Interconnect", MRS Symp. Proc., Vol 309, 409-416 (1993)

9. Sanchez J.E. Jr. and Arzt E. "Microstructural Aspects of Interconnect Failure", MRS Symp. Proc., Vol 265, 131-142 (1992)

10. Sanchez J.E. Jr., Kraft O. and Arzt E "Electromigration Induced Transgranular Slit Failures in Near Bamboo Al and Al-2% Cu Thin-Film Interconnects", Appl. Phys. Lett., 61 (26), 3121-3123 (1992)

11. Kraft O., Bader S., Sanchez J.E. Jr. and Arzt E. "Observation and Modelling of Electromigration-induced Void Growth in Al-based Interconnects", MRS Symp. Proc., Vol 309, 199-204 (1993)

12. Möckl U.E., Bauer M., Kraft O., Sanchez J.E. Jr. and Arzt E. "Detailed Study of Electromigration Induced Damage in Al and AlSiCu Interconnects", MRS Proc. Vol 338, (1994) in press

13. Atakov E.M., Clement J.J. and Miner B. "Characterization of Two Electromigration Failure Modes in Submicron VLSI", MRS Proc. 309, 133 (1993)

14. Arzt E., Kraft O., Nix W.D. and J.E. Sanchez J.E. Jr. "Electromigration Failure by Shape Change of Voids in Bamboo Lines", accepted by J. Appl. Phys.

15. Arzt E., Kraft O. and Möckl U.E. "Electromigration Damage in Conductor Lines: Recent Progress in Microscopic Observation and Mechanistic Modelling", MRS Proc. Vol. 338 (1994) in press

16. Arzt E. and Nix W.D. "A Model for the Effect of Line Width and Mechanical Strength on Electromigration Failure of Interconnects with "Near-Bamboo" Grain Structures", J. Mater. Res. 6 (4), 731-736 (1991)

17. Arzt E., Kraft O., Sanchez J.E. Jr., Bader S. and Nix W.D. "Electromigration Resistance and Mechanical Strength. New Perspectives for Interconnect Materials?", MRS Proc. Vol. 239, 677-682 (1992)

18. Dirks A.G. and Augur R.A. "Reliability and Microstructure of Al-Si-V-Pd alloy Films for Use in Ultralarge Scale Integration", Appl. Phys. Lett. 64, 704-706 (1994)

ACTIVATION ENERGY OF THE EARLY STAGES OF ELECTROMIGRATION IN Al-1%Si/TiN/Ti BAMBOO LINES

I. De Munari°, A. Scorzoni*, F. Tamarri*, F. Fantini°

° Dip. di Ing. dell'Informazione, Univ. di Parma, Viale delle Scienze, 43100 Parma (Italy)
* CNR - Istituto LAMEL, via P. Gobetti 101, 40129 Bologna (Italy)

Abstract *The activation energy of electromigration in "bamboo" type, passivated Al-1%Si/TiN/Ti lines was determined by means of high-resolution resistance measurements at wafer level. Both very early stages, with a non linear behaviour of the resistance, and following stages, characterized by an approximately constant rate of resistance change, were analyzed with an existing model that correlates electromigration and mechanical stress evolution. An activation energy could be extracted only in the phase of linear resistance increase, where a value of 0.95 eV was found. Diffusion at the interface between Al-Si and the barrier metal or the passivation could be responsible for this value of the activation energy.*

1. INTRODUCTION

Electromigration (EM) is electronic-current-induced atomic diffusion, essentially due to momentum transfer from electrons to atoms in the conductor (the so-called "electron wind") [1]. If one-dimensional atomic motion is concerned, the force applied by the electron wind is $F_e = Z^* qE \approx Z^* q\rho j$, where q is the electronic charge, ρ the metal resistivity, E the electric field, j the current density and Z^* is an "effective charge number", usually negative since atoms are pushed in the electronic flow directions.

It is well known that thin metallic films deposited on oxidized silicon substrates are usually in a highly stressed state. The stress in the crystalline matrix is usually induced by the thermal treatments the devices undergo during technological steps, but it can also be originated by the metal deposition itself (intrinsic stress).

If we restrict our study to Al and Al-alloys, which are the most widely used in VLSI, it is commonly found that at room temperature the thermally-induced stress is of tensile type. This can be explained if one recalls that the thermal expansion coefficient of Al is much greater than that of Si and deposited oxides. As a consequence, following a relatively high temperature process (typically 400°C during passivation deposition), the metallization tends to shrink relative to the substrate, and a tensile stress forms in the Al line.

Tezaki et al. published experimental results of metal stress in passivated 2 μm wide Al-Si lines during temperature cycles [2]. They demonstrated that at room temperature huge thermal stresses, of the order of 200-400 MPa, arise in the metallizations. Anyway, these values are much lower than what can be simulated starting from a zero-stress state at the temperature of passivation deposition [3]. This suggests that stress relaxation through void growth occurs during cooldown and room temperature aging [3].

Thermally induced stress is not the only source of mechanical stress in the metal lines of integrated circuits. high current tests, accumulation and depletion of atoms due to EM lead to local changes in the mechanical stress σ of the line [4]. This effect is enhanced by the presence of a passivation. A stress gradient arises between accumulated (more compressed) and depleted (in a more tensile state) parts of the metal line, thus creating a back flow of atoms. Then, the local atomic flux assumes the following form:

$$J_a = (DN_a/kT)[\Omega(\partial\sigma/\partial x) + F_e] \qquad (1)$$

where N_a is the atomic concentration, Ω the atomic volume $1/N_a$, kT the thermal energy

per atom and D is the applicable self-diffusion coefficient, which depends on the diffusional mechanisms acting inside the polycrystalline material (grain boundary, bulk, dislocation assisted diffusion, etc.).

Even if eq. (1) and the theory behind it are generally accepted by the research community, different models related to the basic physical mechanisms leading to EM have been proposed [4 - 8], but the overall comprehension of the phenomenon is still unsatisfactory.

The aim of this paper is to demonstrate the capability of a high resolution resistometric method for electromigration [9] to test the validity of existing EM theories, taking into account the effect of the stress.

Since resistance fluctuations can be induced both by EM and by thermal instabilities of the ambient in which an EM measurement is performed, a good high resolution EM technique should be insensitive to these temperature-induced fluctuations. In this work the compensation of temperature fluctuations is achieved by means of a ratio of resistances [10].

An introductory section will be devoted to a brief description of a new stress-diffusion EM model [8], which tries to describe with relatively simple mathematics the complex interaction between mechanical stress and EM, and to the assumptions that have to be made to correlate the results of this model with resistance changes during EM.

Experimental results obtained with the high resolution resistometric technique will be presented and the stress-diffusion EM model will be adopted to describe quantitatively the measured resistance behaviours when a high stressing current is introduced in the narrow metal lines.

A short discussion will follow, with the aim to highlight some limitations still present in this theory.

2. EM MODEL AND RESISTANCE VARIATIONS

Recently, several research groups developed analytical or numerical models of early stages of EM [4 - 8]. Some of these model also concern resistance variations [5, 7]. Moreover, simplified

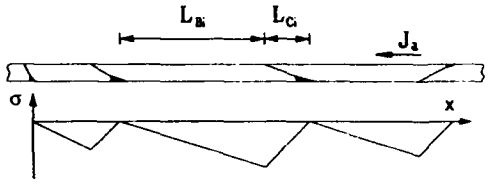


Figure 1: *Bamboo structure of a narrow metal line. L_{Ci} and L_{Bi} are the lengths of the cluster and bamboo sections, respectively. The steady-state stress distribution is also shown.*

treatments on reciprocal influence of EM and mechanical stress are available [4, 6, 8], but only in the case of a hydrostatic stress σ in a rigidly passivated line.

These models are usually applied to narrow, passivated, Al-based metal lines, in the special, but technologically very relevant, case of alternate "bamboo" sections and so-called "cluster" sections with connecting, longitudinal grain boundaries. A cluster section is always bounded by two transverse "blocking boundaries". This structure can be described as "near-bamboo".

It must be underlined that a perfect "bamboo" structure (see Fig. 1) can be considered as a particular case of a "near-bamboo", with very short "cluster" regions. In fact the boundaries between blocking grains are likely to have a nonzero component of the electric field along their direction since they usually form a non orthogonal angle with the line edges.

Regarding the effect of a passivation, it is worth stressing that the resistance behaviour of non-passivated samples usually has the same characteristics of passivated ones. This implies that the basic mechanisms do not change, but the rate of resistance change usually increases in the case of absence of a passivation, since hillocks are more prone to grow.

Starting from the flux and continuity equations, Korhonen et al. [11] and Kirchheim [4] showed that

$$\frac{\partial \sigma}{\partial t} = \frac{\partial}{\partial x}\left[\kappa\left(\frac{\partial \sigma}{\partial x} + G\right)\right] \qquad (2)$$

where $\kappa = DB\Omega/kT$, $G = Z^*q\rho j/\Omega$, B is an appropriate bulk modulus and all the other symbols were previously defined. Eq. (2) looks like the equation of heat conduction in solids.

Supposing constant ρ and D, analytical solutions of eq. (2) during transients in near-bamboo lines were derived [8], so that correlations among stress evolution and void [illegible] could be attempted.

During EM, in the cluster sections compressive stress accumulation on one side should induce the growth of a void on the other side. The final volume increment of the void is $V_0 = wd(\sigma^T L_C/B + GL_C^2/2B)$, where σ^T is the initial constant thermal stress and L_C is the cluster length. V_0 is accumulated in the characteristic time $t_o = L_C^2/\kappa_C$, where κ_C is referred to the grain boundary diffusion coefficient in the cluster. $t_o \simeq 300$ s at 250°C [8], while at 200°C $t_o \simeq 1000$ s ($E_a = 0.6$ eV was assumed for this calculation).

The detection of time t_o depends on the time-resolution of the method used to detect EM damage. Clearly, the greater the resolution, the greater is the capability to discriminate different effects in the early stages of EM.

After t_o the void volume should increase distributing atoms further away to neighbouring bamboo sections. If $t' = t - t_o$, the resulting increase in the void volume V_1 is given by:

$$V_1(t') \simeq wd[(2\sigma^T + GL_C)/B]\sqrt{\kappa_B t'} \qquad (3)$$

If m cluster sections are present in the line, the total void volume accumulated is a simple sum of the individual volumes $V_{0i} + V_{1i}$, $i = 1, m$. If $t' \gg t_o$ a typical square root volume increase as a function of time should be observed. If we suppose that relative resistance variations are proportional to relative volume variations, a similar behaviour should be detected during high resolution resistance measurements.

In general, electromigration can induce both resistivity variations and void motion, growth and shrinkage. Due to high thermal stresses generated after cooling from passivation deposition, voids might be present at the beginning of an EM test (stress-induced voiding) [8].

In the following sections, resistance variations will be mainly attributed to void growth/shrinkage. Regarding the intriguing effect of resistivity variations see the discussion in [10].

It is easy to demonstrate that in the case of (a) slit-like voids having constant depth and variable length (the void might span the whole thickness of the Al line when refractory metal films are present as under and overlayers) and (b) voids with constant width along a longitudinal grain boundary, the relative resistance variation of the metal line is of the same order of the volume variation per unit volume of the line itself. In ref. [5] it was demonstrated that this conclusion is also true for cylindrical voids. In the following we will use this simple proportionality in order to correlate resistance variations and change of void volume.

3. STUDY OF THE RESISTANCE CHANGES

In this section we will analyze EM experiments performed on glass passivated, 0.9 μm wide, 1000 μm long, Al-1%Si(900 nm)/TiN(60 nm) /Ti(50 nm) lines. True test temperatures were 190, 200, 220, 230 and 240°C.

At least 10 samples were simultaneously measured at each temperature at the wafer level. A constant current density of 2 MA/cm^2 was used.

Previous microanalytical investigations revealed an almost perfect "bamboo" grain structure of these lines [14].

Let us now concentrate our attention on resistance changes measured at different stages of the EM process. As stated before, we will suppose that resistance changes during EM are only induced by geometry changes as void growth/shrinkage and not by resistivity variations.

It is well known that resistance changes may also occur during and after heating the sample to the aging temperature [12]. Obviously, these changes are not caused by electrical current, but by structural relaxation or reversible and irreversible solid-state reactions of the aluminum with the additives in the crystalline matrix [13]. In this case resistance changes can be induced both by resistivity and geometrical changes. In order to get rid of these effects, not directly related to EM, EM tests were always

preceded by the typical measurement sequence described in [10]. Failure to do this could affect the detection of EM-induced very early resistance changes, as one can realize comparing data presented in this paper and in ref. [9].

Fig. 2 shows typical experimental behaviours obtained at 220 and 240°C.

In Fig. 2(a) two different measurements at a true test temperature of 240°C are reported. An initial region is clearly observable, in which the resistance increase can seemingly be described by a function $\sqrt{(\kappa t)}$. This initial behaviour is typical of all high resolution measurements.

After about 10 hours of test, the resistances continue to grow with an approximately constant rate.

Measurements at all the other temperatures highlight again a "square root" followed by a linear resistance increase. A further example is reported in Fig. 2(b), showing the behaviour of 10 lines measured at 220°C.

The initial non-linear resistance increase was analyzed by means of the model of eq. (3), provided that relative resistance variations are proportional to relative volume variations. For a metal line with m cluster sections and total length L_{tot} the relative resistance change could be calculated as:

$$\frac{R(t) - R_o}{R_o} \propto \frac{\sqrt{\kappa_B t}}{L_{\mathrm{tot}}} \sum_{i=1}^{m} \frac{2\sigma^T + GL_{Ci}}{B} \qquad (4)$$

thus providing the expected $\sqrt{t}$ trend. Clearly, the relative resistance change is strongly influenced by the initial thermal stress σ^T.

Fig. 2(c) shows the relative resistance change vs. $\sqrt{t}$ for the 10 lines of the previous figure, measured at 220°C. The linear shape of the curves is evident in the early stages of EM. Linear fittings of good quality were also found at the other test temperatures.

Note that eq. (4) assumes that only one diffusion mechanism operates during EM. Since at this time the grain boundary diffusion stage should be concluded, possible mechanisms are diffusion in bulk Al-Si or at the interface between Al-Si and the passivation or the underlying Ti-based barrier.

We have seen that, after a certain elapsed

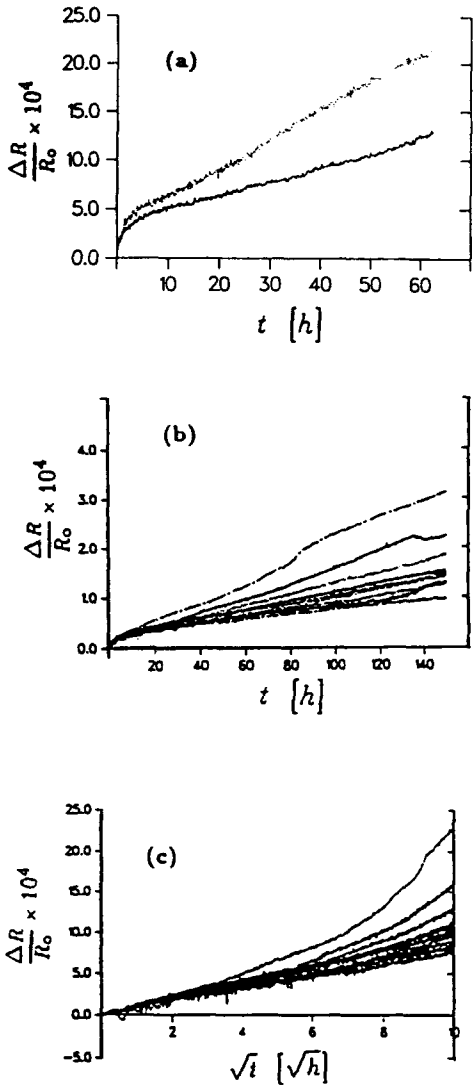


Figure 2: *Resistance changes during and after EM experiments on passivated, Al-1%Si/TiN/Ti submicron lines subjected to a current density of 2 MA/cm^2. (a) two different behaviours at a true test temperature of* 240°C *in a common time scale. (b) Resistance change of 10 lines at* 220°C. *(c) Relative resistance change vs. $\sqrt{t}$ for the 10 lines of (b).*

time, depending on temperature, the resistance growth becomes approximately linear. It is argued that in this phase the internal stress of the lines has reached the steady-state oscillating (or saw-toothed) pattern described in ref. [8] (see Fig. 1). Now couples formed by cluster and bamboo sections (the so-called "failure units") start interacting by stress-assisted bulk diffusion through the bamboo sections. At this stage some voids shrink, while some grow at the expense of others. In such a steady state for stress it is likely that the rate of void growth/shrinkage is simply proportional to the atomic flux.

4. DEPENDENCE ON TEMPERATURE

In the previous section we have seen that the stress-diffusion model of EM is able to describe resistance changes in two different stages of EM. In principle, an activation energy can be extracted in both cases.

In Fig. 3(a) the coefficients of $\sqrt{t}$ in eq. (4) extracted by linear fitting applied to the data measured at 190, 200, 220, 230 and 240°C (see e.g. Fig. 2(c)) are reported. From Fig. 3(a) it can be seen that data are highly dispersed, and a non-uniform increase of the average values as a function of temperature is evident. Reasons for dispersion can be found in the different microstructure, or in non negligible and non uniform change of resistivity in the various samples.

According to the model of eq. (4), a possible reason for the unusual thermal activation of the data is that, notwithstanding all resistance transients due to change of temperature were concluded before starting the tests, the initial stress σ^T was different at the various temperatures. For these reasons the extraction of the activation energy was not attempted in this phase.

In the linear, steady-state phase the activation energy E_a was extracted supposing that the relative resistance variations are proportional to the atomic drift velocity, $\Delta R/R \propto v_B = J_a/N_a$. Even if a high dispersion of the data was found also in this case (see Fig. 3(b)), a smooth and monotonic increase of the average values of the relative rates of resistance change was detected and a value $E_a = 0.95\,\text{eV}$, with an error $\Delta E_a = 0.1\,\text{eV}$ (99% confidence interval), was extracted. This activation energy is lower than commonly accepted values for bulk Al diffusion ($E_a \approx 1.4\,\text{eV}$), but higher than values for grain boundary diffusion ($E_a \approx 0.6\,\text{eV}$) [1]. Diffusion at the interface between Al-Si and the barrier metal or the passivation could be responsible for this intermediate value of activation energy. This explanation is consistent with previous failure analysis done on similar samples [15].

5. DISCUSSION AND CONCLUSIONS

In this paper it was shown that an existing stress-diffusion model on void growth is in substantial agreement with resistance changes in near-bamboo metal lines during EM. An activation energy of $0.95 \pm 0.1\,\text{eV}$ was extracted in zones of linear resistance increase.

Only geometrical variations have been supposed to influence the resistance changes. This model has the important merit to be very simple, since only analytical solutions of the stress-diffusion equation are needed to approximate the experimental data of resistance changes collected during EM. Moreover, the model itself partly explains the high dispersion of the measured data.

The analytical solutions are based on specific assumptions. The problem of a stress-dependent diffusion coefficient has already been addressed in ref. [11], where it was shown that the analytic solutions may in several cases be sufficient for practical estimations of stress buildup during EM.

More important, regarding resistance measurements, is the assumption of a constant resistivity. The electrical resistivity depends on defects present or induced in the polycrystalline metal lines. In particular, vacancy supersaturations [7] and an increase of dislocation density during EM could cause measurable resistivity variations. Moreover, the presence of large hydrostatic mechanical stresses in the metal polycrystalline matrix should induce additional resistivity variations. The high dispersion of rates of resistance change can also be explained in terms of these resistivity variations.

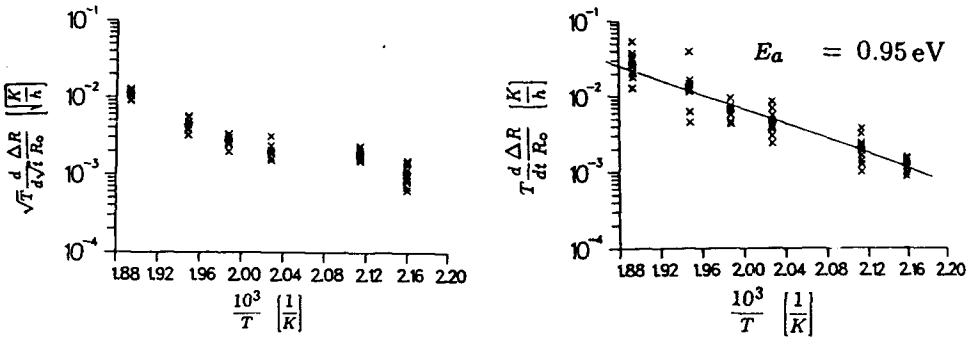


Figure 3: *(a) Arrhenius plot of the coefficients of $\sqrt{t}$ obtained during the initial, non-linear phase of EM at 190, 200, 220, 230 and 240° C. (b) Arrhenius plot of the slopes obtained during the linear phase of EM from measurements at the same temperatures of (a). Values are corrected to take into account the T^{-1} dependence of the parameter κ and of the drift velocity v_B. × = measured data. • = average value.*

To be complete, a model for the description of resistance changes during EM in near-bamboo lines should address also these important issues.

ACKNOWLEDGMENTS

Thanks are due to Professors D. Nobili and A. Desalvo for helpful discussions and to Dr. C. Caprile (ST Microelectronics) for supplying structures. This work was partially supported by the ESPRIT III-ADEQUAT project.

References

[1] A. Scorzoni, B. Neri, C. Caprile, F. Fantini, Material Science Reports, **7** (4-5), 143 (1991).

[2] A. Tezaki, T. Mineta, H. Egawa, T. Noguchi, Proc. 28th IEEE Int. Rel. Phys. Symp., 221 (1990).

[3] M. A. Korhonen, W. R. LaFontaine, P. Børgesen, Che-Yu Li, J. Appl. Phys., **70** (11), 6774 (1991).

[4] R. Kirchheim, Mat. Res. Symp. Proc., **309**, 101 (1993).

[5] Z. Li, C. L. Bauer, S. Mahajan, A. G. Milnes, J. Appl. Phys., **72** (5), 1821 (1992).

[6] W. D. Nix, E. Arzt, Metallurgical Trans. A, **23 A**, 2007 (1992).

[7] J. Niehof, H. C. de Graaf, J. F. Verwey, Proc. of the 4th Europ. Symp. on Reliab. of Electron Dev., Failure Phys. and Analysis (ESREF '93), Bordeaux (France), 135 (1993).

[8] M. A. Korhonen, P. Børgesen, D. D. Brown, Che-Yu Li, J. Appl. Phys., **74** (8), 4995 (1993).

[9] G. L Baldini, A. Scorzoni, IEEE Trans. on Electron Devices, **38** (3), 469 (1991).

[10] A. Scorzoni, I. De Munari and H. Stulens, MRS 1994 Spring Meeting Symposium Proceedings, S. Francisco, April 4-8 1994 (in print).

[11] M. A. Korhonen, P. Børgesen, K. N. Tu, Che-Yu Li, J. Appl. Phys., **73** (8), 3790 (1993).

[12] H. Stulens, G. Knuyt, W. De Ceunick. L. De Schepper, L. Stals, in press in J. Appl. Phys., **75** (4) (1994).

[13] D. S. Gardner, P. A. Flinn, J. Appl. Phys., **67** (4), 1831 (1990).

[14] G. L. Baldini, I. De Munari, A. Scorzoni, F. Tamarri, C. Caprile, F. Fantini, ESREF '93, Bordeaux (France), p.147 (1993).

[15] C. Caprile, G. Specchiulli, D. Sala, F. Fantini, Quality and Reliab. Eng. Int., **7** (4), 275 (1991).

ELECTROMIGRATION BEHAVIOUR OF ON-CHIP INTERCONNECTS: INFLUENCE OF STRESS PARAMETERS STUDIED BY EARLY RESISTANCE CHANGES

V. D'Haeger[1], W. De Ceuninck[1,2], H. Stulens[1], L. De Schepper[1,2] and L.M. Stals[1]

[1] Limburgs Universitair Centrum - Institute for Materials Research - Materials Physics Division - Wetenschapspark 1 - B-3590 Diepenbeek (Belgium)
Tel.++ 32 - 11 - 26 88 26 - Fax ++ 32 - 11 - 26 88 99

[2] DESTIN N.V. - Wetenschapspark - Universitaire Campus - B-3590 Diepenbeek (Belgium)
Tel.++ 32 - 11 - 26 88 11 - Fax ++ 32 - 11 - 26 88 99

ABSTRACT

Several models have been developed in order to relate electromigration (EM) induced resistance changes to time and stress parameters. Most of these models state a Black-type equation. In this paper, a critical evaluation of the influence of these parameters is presented. First, the influence of time has been studied in order to verify the validity of models which state a linear relationship between EM induced resistance changes and time. Also, the dependence on temperature and current density is examined. Our results show that these models are much too simplified and cannot be used at all stress conditions.

1. INTRODUCTION

A frequently used technique to analyse the early stages of electromigration, is to measure the resistance changes of interconnects during isothermal experiments at high stress conditions, such as temperature and current density. In 1969, Black published a simple theory in which he related the MTTF (median time to failure) results with temperature and current density (Ref.1). Since then, several models (Refs. 2-6) have been published, relating resistance measurements data to the MTTF data and the stress parameters. Here, a critical evaluation is performed in order to examine the validity of these Black-type models.

2. EXPERIMENTAL DETAILS

Experiments are performed on passivated metallizations sputtered on a Si-oxide layer with underlying Si substrate. The stripes have a width of 10μm and a meander structure with a total length of 6000μm. Six types of metallizations are studied and are listed in table 1.
Metallizations of type A, B, C and D are 0.8μm thick and those of type E and F have a thickness of 1.0μm.

Type	Metallization	Deposition temperature (°C)
A	Al 1%Si 0.5%Cu	220
B	Al 1%Si 0.5%Cu	400
C	Al 1%Si	250
D	Al 1%Si	400
E	Al 1%Si 0.5%Cu	180
F	Al 1%Si	180

Table 1 : The metallizations used in the experiments.

3. RESULTS AND DISCUSSION

In order to evaluate models which predict a Black-like behaviour, high resolution electrical resistance measurements are performed. High current densities are applied to the metal lines, and the resistance changes are measured as a function of time during high temperature isothermal experiments. A critical evaluation of the time dependence and the influence of the stress parameters on resistance changes is presented.
Models which relate resistance changes to stress parameters, introduce the following equation for the rate of resistance change (RRC) of the metal line, assuming a linear dependence on time (Refs.5, 6) :

$$RRC = \frac{1}{R_0}\frac{dR}{dt} = A\, j^n \exp(-E/kT) \qquad (1)$$

where all the parameters have their usual meaning.

3.1. Time dependence

The linear time dependence, as given by eqn.(1), is indeed observed in many experimental cases, i.e. our type D samples show such a dependency for standard stress conditions (Ref.8). However, if these samples are subjected to a very low current density,

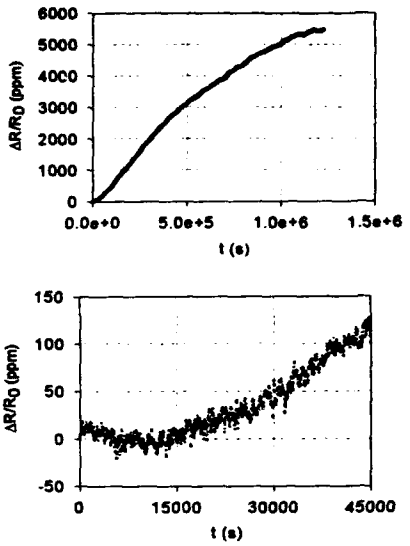


Figure 1 : EM for a pre-annealed type D sample at T=269°C and j=0.13 MA/cm² : total measurement (upper) and initial period (lower).

the time behaviour seems to be more complicated. For this type of measurements, all samples are pre-annealed during several hours at T=269°C, until no net resistance variation is observed. Then, the metallization is subjected to a low current density. The result of such an EM experiment is shown in figure 1. An initial incubation period is followed by a linear behaviour. However, at a certain moment, a deviation from this linear behaviour occurs. This type of measurements indicates that the underlying atomic processes are more complicated than is the case in the Black-like EM-models. In figure 2 the results are shown for different low current density experiments performed at the same sample temperature on type D metallizations. Since the

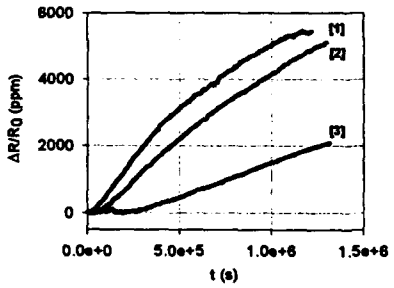


Figure 2 : Low current density experiments on pre-annealed samples of type D at T=269°C and j=0.13 MA/cm² [1], j=0.09 MA/cm² [2] and j=0.06 MA/cm² [3].

applied current densities are 10 to 20 times smaller than in conventional EM-experiments, the temperature rise due to Joule heating is negligible small. From our tests it is clear that the lower the current density, the longer the incubation time. In our experiments at high current densities (>> 10^5 A/cm²) no incubation time is observed. Note also that this incubation time can become very long, e.g. several days for j=0.06 MA/cm².

3.2. Current density dependence

The current density dependence is expressed in the models by the current density exponent, n. In theoretical models, derived from pure diffusion mechanisms, n should be equal to 1 (Refs.5,6). Other models (Ref.7) predict a current density exponent equal to 2. From experiments, a wide spread of values for n is reported from 1 up to 5 and higher.

We performed different isothermal experiments at current densities ranging from j=0.3 MA/cm² to j=2.5 MA/cm² at T=269°C. At these stress conditions, a linear increase of the resistance with time is observed during 48 hours and the RRC can be determined. Figure 3 shows the RRC's as a

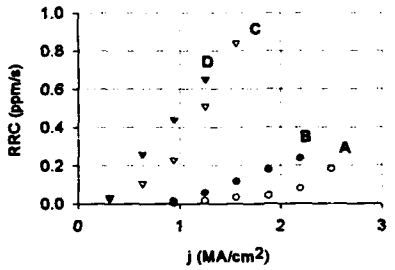


Figure 3 : RRC as a function of current density at T=269°C for samples of type A, B, C and D.

function of the current density for four types of samples. It is obvious that we can distinguish between the different metallizations. Cu-addition retards the EM process and a difference in deposition temperature is reflected in a more or less pronounced current density dependence.

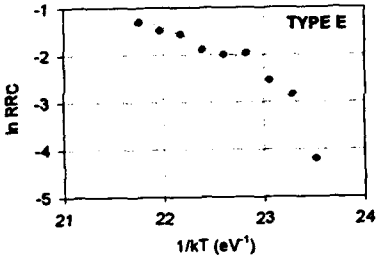
Applying eqn.(1) to our results, the current density exponent n can be determined. From figure 3 it is clear that n=1 requires the existence of a current density, below which no electromigration should occur. This assumption is verified by performing experiments at current densities much lower than the expected "threshold" value. For type D samples, we performed such low current density measurements. If n=1, no EM should occur at j<0.3 MA/cm², as deduced from figure 3. As shown in figure 2, an increase in the resistance versus time is observed, even at j=0.06 MA/cm². These

measurements indicate that EM is still active; even at current densities below the expected "threshold" value. A unique determination of n is difficult and seems to have no signification, because the time dependence at low and high current densities is different. Note that this "threshold" current density is not to be confused with the Blech threshold current density. The determination of this Blech threshold current density for our samples is virtually impossible. The expected Blech critical current density for our samples is certainly less than $j=5kA/cm^2$. The detection of this threshold current density would require unreasonable long measuring times, since the incubation time increases sharply with decreasing current density.

3.3. Temperature dependence

3.3.1. High current density measurements

The Boltzmann factor in eqn.(1) provides a method to determine the activation energy E. EM experiments are performed at $j=1.9$ MA/cm^2 and at



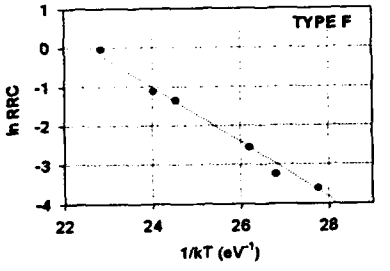


Figure 4 : Arrhenius plots to determine the activation energy for EM (type E and F samples).

different temperatures. Figure 4 shows the (ln RRC) vs. (1/kT) plots for type E and F samples. It is clear that for simple AlSi metallizations (type F) a good linear fit is obtained, yielding an activation energy of ~0.74eV. This corresponds approximately to the activation energy of grain boundary diffusion. For the more complex AlSiCu metallizations (type E) however, a deviation from the linear dependence is observed. The activation energy seems to be a decreasing function of temperature. This can be explained by the solubility of Cu in Al. At high temperatures no $CuAl_2$ precipitates are present and grain boundary diffusion is again the dominating process. From the fit at the high temperature range, an activation of ~0.67eV is obtained. At lower temperatures, Cu precipitates at the grain boundaries, impeding diffusion along these paths and bulk diffusion takes place, which gives rise to a much higher activation energy.

3.3.2. Low current density measurements

The temperature dependence is also examined for type D samples subjected to a low current density. Again, this samples are pre-annealed for several hours at T=269°C. Then, the temperature is

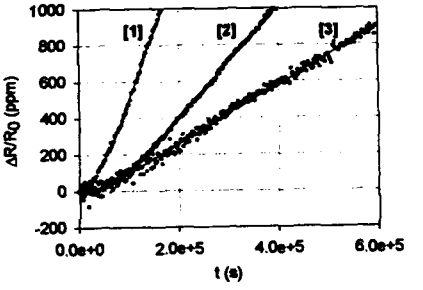


Figure 5 : Early resistance changes versus time during low current density measurements with $j=0.13$ MA/cm^2 at T=269°C [1], T=236°C [2] and T=225°C [3].

decreased to the sample EM test temperature. The metal line is again subjected to this condition until no net resistance change is observed, in order to eliminate possible reversible processes. Then, the metal line is subjected to a high current, in order to perform the EM experiment.

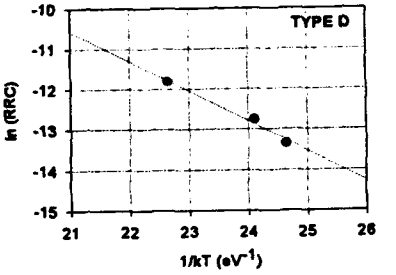


Figure 6 : Arrhenius plots to determine the activation energy for EM at $j=0.13$ MA/cm^2 (type D samples).

As discussed in section 1, at low stress conditions, the linear dependence of the resistance on time is only seen during the early stages of EM and after an incubation period. In order to determine an activation energy, the RRC's are calculated from the plots in figure 5. These RRC's are typical 100 times smaller than those measured at the high stress conditions. Figure 6 shows the (ln RRC) versus (1/kT) plots. An activation energy of ~0.74eV is obtained, which corresponds very good with the high current density results for AlSi metallizations.

Note that our high j measurements are performed on type F samples (T_{dep}=180°C) and the low j measurements on type D samples (T_{dep}=400°C). SEM investigations reveal a polygranular structure for both types of metallizations, with comparable grain sizes (1-2μm). As a consequence, the comparison between high and low current density results is justified. We can state that the activation energy is not dependent on the amount of applied current stress. The driving mechanism is the same in both cases and can be identified as grain boundary diffusion.

4. CONCLUSIONS

Early resistance measurements are useful to study electromigration. However, attention must be paid when using a model to analyse the results. Black-type equations cannot be applied in all cases and at all stress conditions.

First, time dependence is not always as simple as supposed in most of the models. A linear dependence is observed only at certain stress conditions.

Furthermore, experiments show that even at very small current densities EM is active and the current density exponent n≠1. The determination of n for the complete current density range is not meaningful because of the complex time behaviour.

The Arrhenius type dependence on temperature seems to be correct, since a reasonable activation energy is obtained. However, attention has to be paid for metallizations with Cu-addition. Here, the physical properties of the metal line change as a function of temperature, giving rise to different diffusion processes and different activation energies.

Finally, we showed that the temperature factor and current density factor are independent. For polygranular AlSi metallizations the EM activation energy corresponds to grain boundary diffusion at all stress conditions.

REFERENCES

1. Black J.R., Proc. IEEE 57 (1969) 1587.
2. Bobbio A. and Saracco O., Microelectron. Reliab. 14 (1975) 431.
3. Maiz J.A. and Segura I., Proc. of the 28th IEEE Int. Rel. Phys. Symp. (1990) 25.
4. Hummel R.E., DeHoff R.T. and Geier H.J., J. Phys. Chem. Solids 37 (1976) 73
5. Pasco R.W. and Schwarz J.A., Solid State Electron. 26 (1983) 445.
6. Li Z., Bauer L., Mahajan S. and Milnes A.G., J. Appl. Phys. 72 (1992) 1821
7. Shatzkes M. and Lloyd J.R., J. Appl. Phys. 59 (1986) 3890.
8. D'Haeger V., Stulens H., De Ceuninck W., De Schepper L., Gallopyn G., De Pauw P. and Stals L.M. , ESREF93, Bordeaux (1993) 141.

FAILURE LOCATIONS IN DIFFERENT VIA STRUCTURES DUE TO ELECTROMIGRATION

K.Weide *, W.Hasse **

* Laboratorium für Informationstechnologie, Universität Hannover,
Schneiderberg 32, 30167 Hannover, Germany

** Institut für Halbleitertechnologie und Werkstoffe der Elektrotechnik,
Universität Hannover, Appelstr. 11a, 30167 Hannover, Germany

1. ABSTRACT

For reliability prediction of via structures in multilevel metallizations with numerical methodes (like finite element methode -FEM-) it is possible to identify the weakest link in the structure caused by local heating and current crowding. In this paper the influence of current crowding, local heating and mass flux divergences on different via structures as a conventional via with small step coverage, a tungsten plug and an aluminum plug via will be discussed. Also the dependance of a horizontal scaling on the current crowding, local heating and mass flux divergence will be shown.

2. SUMMARY

A 3-dimensional parametrized FEM-model of a double level metallization via structure was developed and the current density- and temperature distributions depending on the geometry and material properties were determined (Ref.1). The model was verificated by a comparision between simulation and measurements (Ref.2). For the calculation of the local mass flux and mass flux divergences the commercial FEM-program ANSYS was extended by a Fortran user routine. With this user routine it is possible to determine the influence of electromigration, thermomigration and triple points on the mass flux divergences in the metallization of the via structure as well as the determination of the failure location in the structure by the location of the maximum mass flux divergence. The comparison between the failure locations out of the simulation with the failure locations out of the measurement shows a very good agreement.

3. THEORY

The mass flux divergence $\operatorname{div} \vec{J}_A$ caused by electromigration (temperature gradients) is given by equation [1].

$$\operatorname{div} \vec{J}_A = \frac{e Z^* N D_0}{k_B} \left[\vec{j} \operatorname{grad} \frac{\rho}{T} \left(\exp\left(\frac{-E_A}{k_B T}\right) \right) \right] \qquad [1]$$

D_0 is the diffusion coefficent of aluminum, N the concentration of the atoms, eZ^* the effective charge of ions, k_B the Boltzmann constant, $\vec{j}$ the current density, T the temperature, ρ the resistivity and E_A the activation energy for an easy diffusion path like grain boundaries or interface diffusion.

By the occurrence of temperature gradients the thermomigration is not neglectable for high accelerated tests. The thermomigration mass flux divergence $\operatorname{div} \vec{J}_{TH}$ due to the temperature gradients is given by equation [2], with the heat of transport Q.

$$\operatorname{div} \vec{J}_{TH} = \operatorname{grad} T \operatorname{grad} \left(\frac{D_0 Q N}{k_B T^2} \exp\left(\frac{-E_A}{k_B T} \right) \right) + \frac{D_0 Q N}{k_B T^2} \exp\left(\frac{-E_A}{k_B T} \right) \operatorname{div} \operatorname{grad} T \qquad [2]$$

The mass flux divergence caused by triple points $\operatorname{div} \vec{J}_{TP}$ is proportional to the sum of $\vec{J}_A$ and $\vec{J}_{TH}$ [3].

$$\text{div } \vec{J}_{TP} = G \left(\vec{J}_A + \vec{J}_{TH} \right) \quad [3]$$

The total mass flux divergence div $\vec{J}_G$ is than given by adding equation [1], [2] and [3]. Positive mass flux divergences lead to voids and negativ mass flux divergences lead to material accumulation (continuity equation). The proportional factor G depends on the geometry of the triple point.

It was assumed that the complete structure is in equilibrium, so that the concentration gradients and mechanical stress can be neglected.

Fig.1: SEM-Picture of a via structure after stress test (T=150°C and 100mA). Failure locations (voids) marked by the arrows.

4. EXPERIMENTAL

The failure location after the accelerated stress test for a double via structure is shown in Fig.1.
The current flow direction was from the right to the left (M2-->M1-->M2). Voids were found at the corner of the via into the second metallization.
The calculated mass flux divergence in the model is shown in Fig.2. The voids expected at the maximum mass flux divergence agree very good with the failure location determined after electromigration stress test.

The influence of $\vec{J}_A$, $\vec{J}_{TH}$ and $\vec{J}_{TG}$ vs the applied current density is described elsewhere (Ref.3).

5. HORIZONTAL SCALING

The aspect ratio is defined by the heigth of the via divided by the width of the via. The investigations were carried out with a current density in the via of 4,7MA/cm^2 and a bulk temperature of 473K. The metallization thickness was 1µm. The overlap of M1 was 0,6µm and the overlap of M2 was 1,1µm. With increase of the aspect ratio the current crowding for the conventional via (10% step coverage) and the tungsten-plug via shows a strong increase (Fig.3).
The maximum current density found for the conventional via was nearly homogenous distributed in the step coverage. In the tungsten-plug via the current crowding has two maximum at the corner of M1 to the via and at the via to M2.

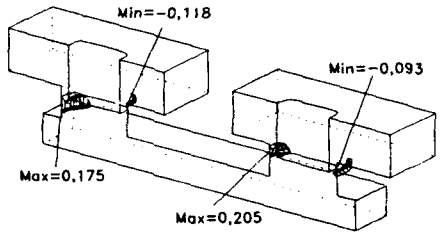


Fig.2: Calculated minimum and maximum mass flux divergences in a conventional via structure in a.u. corresponding to the failure locations shown in Fig.1.

For the local heating a strong increase in the maximum temperature was found for an increasing aspect ratio (Fig.4). The maximum temperature was calculated for a conventional via with 10% and 20%

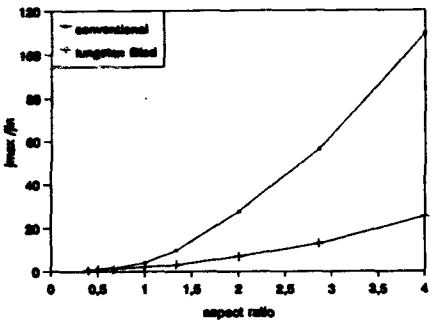


Fig.3: Current crowding (j_{max}/j_{in}) vs. aspect ratio.

step coverage and a tungsten-plug via. To avoid a local heating in the via the aspect ratio should be as small as possible.

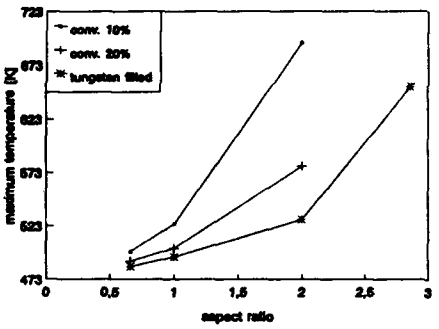


Fig.4: Maximum temperature vs. aspect ratio.

At least the maximum mass flux divergence in the aluminum depending on the aspect ratio was determined for the conventional via with 10% and 20% step coverage and the tungsten-plug via. In Fig.5 the maximum mass flux divergence due to temperature gradients and thermomigration (div $\vec{J}_A$ + div $\vec{J}_{TH}$)max is shown. The maximum mass flux divergence including the divergence of the triple points div $\vec{J}_{TP}$ is shown in Fig.6. For the conventional via with 10% and 20% step coverage no significant difference between both graphs were found. Only for the pluged vias a difference was found for small aspect ratios.

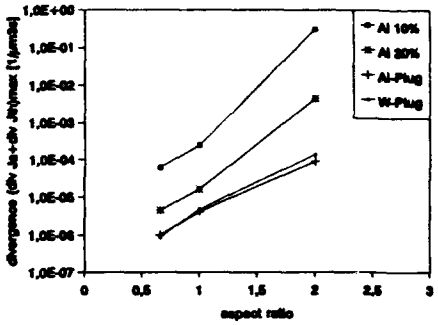


Fig.5: Maximum mass flux divergence out of thermomigration and temperature gradients vs. aspect ratio of the via structure.

For half micron pluged vias (aspect ratio 2) the maximum mass flux divergence is one order of magnitude higher than for one micron pluged vias.

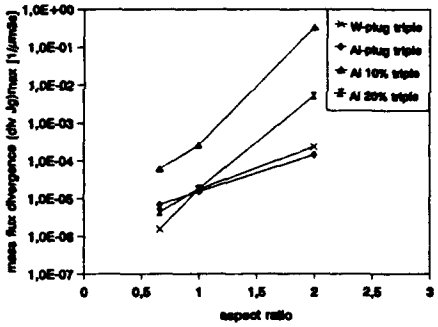


Fig.6: Maximum mass flux divergence vs. aspect ratio including the influence of the triple points.

The values for aluminum-plugs are nearly the same as for tungsten-plugs. So both plug structures seem to be comparable. But the tungsten plug in the via is a diffusion barrier for the aluminum, so in this case the behaviour of this barrier must be added to the mass flux divergence. Conventional half micron vias with a small step coverage have a nearly three orders higher maximum mass flux divergence than the one micron vias. This shows on one hand that the aspect ratio should be as small as possible and on the other hand that a completed via filling is necesarry to avoid high mass flux divergences.

6. COMPARISON OF DIFFERENT VIA TECHNOLOGIES

For a conventional via with 10% step coverage, a tungsten plug and an aluminium plug via the maximum temperatures T_{max}, current crowding j_{max}/j_{in}, maximum temperature gradients $(\text{grad } T)_{max}$, the mass flux $\vec{J}_G$, and the minimum and maximum mass flux divergence $(\text{div } \vec{J}_G)_{min/max}$ were determined. The investigations were carried out for one via out of a chain of parallel via chains. A distance of 5µm and 20µm between two of the via chains were choosen. The interplug distance was 7µm. For the vias a radius of 0,5µm and a heigth of 1µm was choosen. The metallization thickness was 1µm. The overlap of M1 was 0,6µm and the overlap of M2 was 1,1µm. In the simulations a bulk temperature of 150°C and a current density in the via of 4,7 MA/cm^2 with a current flow direction from M1 to M2 was choosen.

In Tab. 1 the values for a chain distance of 20µm are listed.

	10 %	Al-plug	W-plug
$\vec{J}_G[m^2s]^{-1} * e^{-4}$	3,47	0,592	0,595
$(\text{div}\vec{J}_G)_{max}$ $[m^3s]^{-1} * e^{-4}$	0,822	0,0242	0,0244
j_{max}/j_{in}	6,06	1,95	1,79
T_{max} [K]	502	483	485
$(\text{gradT})_{max}$ [K/µm]	7,5	0,7	1,4

Tab. 1: Comparison between a conventional via with 10% step coverage, an Al-plug and a W-plug via structure with a current flow direction M1-->M2 and a distance between the chains of 20µm.

The conventional via with 10% step coverage shows a one order of magnitude higher value for $(\text{div } \vec{J}_G)_{max}$ compared to both plug vias. Comparing the W-plug via with the Al-plug via the current crowding is higher and the temperature gradients are lower in Al-plug via.

A change in the current flow direction from M2 -->M1 leeds to a change in the values of $(\text{div } \vec{J}_G)_{max}$ (Tab. 2).

	Al-plug	W-plug
$\vec{J}_G[m^2s]^{-1} * e^{-4}$	0,604	0,598
$(\text{div}\vec{J}_G)_{max}[m^3s]^{-1} * e^{-4}$	0.0053	0,0065

Tab. 2: Comparison between an Al-plug and a W-plug via structure with a current flow direction from M2 to M1 and a distance between the chains of 20µm.

This behaviour is caused by the effect of the thermomigration. In this case $(\text{div } \vec{J}_G)_{max}$ for both structures are nearly one order of magnitude smaller than for a current flow from M1 to M2. This behaviour depending of the current flow direction compares to observations described in the literature (Ref.4, Ref.5).

A smaller distance between the via chains can lead to higher temperatures as well as higher mass flux divergences in the different structures. Also the locations of the mass flux divergences will change.

In Tab. 3 the calculated values for via chains with a distance of 5µm are shown. The maximum temperature as well as the maximum temperature gradient is higher than in the structure with a distance of 20µm (see Tab.1). For the Al-plug and W-plug via the values of $(\text{div } \vec{J}_G)_{max}$ are a factor of two higher for a smaller distance between the chains. The conventional via with 10% step coverage has compared to a distance of 20µm an about three times higher value. This behaviour shows that a change in the failure mechanism due to the self heating effect in the different via structures depending on the chain distance is possible.

The maximum mass flux divergence for the via with 10% step coverage and a chain distance of 5µm between the chains was located in the step coverage at the bottom of the via. For a distance of 20µm the maximum was found in the via at the corner via/M2. For the aluminum-plug structure the maximum was found independant of the distance between the structures in the via at the corner of M1/via. The tungsten plug via shows its maximum for both distances and a current flow direction M1-->M2 in M1 near the corner M1/via.

	10 %	Al-plug	W-plug
$\vec{J}_G[m^2s]^{-1} *e^{-4}$	0,82	0,756	0,863
$(div\vec{J}_G)_{max}$ $[m^3s]^{-1}*e^{-4}$	2,49	0,042	0,047
j_{max}/j_{in}	6,96	1,95	1,79
T_{max} [K]	525	491	497
$(gradT)_{max}$ [K/µm]	11,1	1,1	2,3

Tab. 3: Comparison between a conventional via with 10% step coverage, an Al-plug and a W-plug via structure with a distance between the chains of 5µm.

7. CONCLUSION

It was found that out of the calculation of the current crowding and joule heating (temperature gradients) the main influences can be determined.

In all investigated cases a very good correspondance of the simulations to measurement and literature was found. The calculated determination of the failure locations and the comparison for different types of vias show that finite element analysis is a good tool for electromigration failure prediction.

LITERATURE

1. Weide, K.; Hasse: "3-dimensional Simulations of Temperature and Current Density Distribution in a Via Structure", Proc.Conf.IEEE/IRPS, 1992, pp.361-365.

2. Weide, K.; Hasse, W.: "3-Dimensional FEM-Simulations and measurement of Via Structures", Proc. Conf., ESREF, Bordeaux, 1993, pp.313-317.

3. Weide,K; Hasse, W.: "Prediction of Electromigration Faailure Locations in Multi-level Metallizations Due to Triple Points, Current Crowding and Temperature Gradients", Proc. VMIC Conf., June 1994, pp.536-538.

4. Tao, J.; Young, K.K.; Cheung, N.W.; Hu, C.; "Eletromigration Reliability of Tungsten and Aluminum Vias and Improvements Under AC Current Stress", IEEE Trans. Elec. Dev., 1993(8), No.8, pp. 1398-1405.

5. Estabil, J.J.; Rathore, H.S.; Dorleans, F.: "The Effect of Metal Thickness on Electromigration Induced Extrusion Shorts in Submicron Technology", Proc. 29th IEEE/-IRPS, 1991, pp. 57-63.

THE EVOLUTION OF THE MICROSCOPIC DAMAGE IN ELECTROMIGRATION STUDIED BY MULTIPLE ELECTRICAL MEASUREMENTS

B.K. Jones, Y.Z. Xu and P. Zobbi*

School of Physics and Materials,
Lancaster University,
Lancaster. LA1 4YB
United Kingdom

Tel. No. + 44 524 593657
Fax No. + 44 524 844037
email B.K.JONES@LANCASTER.AC.UK.

* Present address:
Department of Information Engineering, University of Parma, Italy

1. ABSTRACT

Continuous measurements of the changes in resistance and self-heating effects during electromigration are analysed to reveal details of the void formation and healing effects in the final stages of failure.

2. INTRODUCTION

Electromigration is an important mechanism of failure in integrated circuit interconnects. Considerable progress has been made in the improvement of the electromigration lifetime by control of the photolithography, metal deposition conditions, crystal size and alloy composition. New structures and metals are being evaluated and developed. However the assessment of the failure time and the understanding of the detailed failure process is not easy. Failure time methods are either very slow or suspect because of the high acceleration factors used in wafer-level tests. Electron microscope studies are also very lengthy. Multiple electrical measurements taken continuously during the sample lifetime under stress can give considerable information on the microscopic processes leading to failure.

We report the results of the analysis of measurements in an AC bridge of the sample resistance, second harmonic amplitude and phase-shift. These are measures of the resistance, temperature coefficient of resistance and thermal time constant. With only simple assumptions these can reveal the cross-sectional area and local resistivity changes along the track.

3. APPARATUS

The AC bridge is a resistive bridge with a low-noise amplifier as the detector. The bridge is kept balanced by four servo loops to cancel the fundamental (1.25 kHz) and second harmonic signals with both the in-phase and out-of-phase components. The size of the feedback signals gives a measure of the resistance change, the self-heating and the effective thermal time constant. The bridge has been described briefly earlier (Ref. 1) and a more detailed description of its performance will be given in a later publication. The sample, which forms one arm of the bridge, can also be subjected to a D.C. electromigration stress.

The value of the change in the fundamental signal gives a measure of the resistance change. The second harmonic signal derives from the direct current passing through the sample which is changing its resistance at the second harmonic due to the Joule heating and consequent temperature change caused by the A.C. bridge signal. The change in the absolute magnitude of the second harmonic signal gives a measure of the Joule heating of the sample and the change in the phase-shift gives a measure of the thermal time constant of the heated part of the sample (Ref. 2). The thermal time constant is usually very short compared with the A.C. signal period. The results

are taken continuously with an interval of about two seconds.

A variety of specimens has been studied including Al 4% Cu passivated and unpassivated and with a TiN capping layer.

4. RESULTS

The apparatus has been used for several experiments including spot measurements taken at intervals during highly accelerated electromigration stess (Refs. 1-3) as well as the continuous measurements reported here. These results will be reported elsewhere. Here we will present briefly some results and early analysis. The data presented are the resistance change $\Delta R/R$ % and the change in the value of $\Delta\alpha\beta/\alpha\beta$ % derived from the magnitude of the second harmonic change. We will present, but not use, the phase-shift or delay data.

The quantity $\alpha\beta$ is the produce of the temperature coefficient of resistance, α, and the thermal resistance, β, given by the equations

$$\rho = \rho_0 (1 + \alpha T) \tag{1}$$

where ρ and ρ_0 are the total and temperature-independant resistivities.

$$\delta T = \beta I^2 R \tag{2}$$

is the temperature rise due to a current I flowing through a resistance R

$$V(2\omega) = \frac{3}{2} I_{DC} I_0^2 R^2 \alpha\beta \tag{3}$$

is the second harmonic voltage due to a direct current I_{DC} and an AC current $I_0 \cos \omega t$.

5. ANALYSIS

The evolution during the life of the specimen shows three main regions. Initially there is a very short period corresponding to the anneal of the sample strains. In the main period (phase II) there is a slow linear rise in resistance. In the later stages of the life (phase III) there are violent increases in resistivity which heal again (Ref. 3).

The analysis of these effects is based on the assumption of Mattheissen's Rule and also that during phase II the changes are happening uniformly throughout the length of the sample and during phase III each event is happening in a very short localised length of the sample.

Mattheissen's Rule states that for bulk metals the resistivity is due to carrier scattering processes which are independent of each other so that the total resistivity

$$\rho = \rho_0 + \rho(T) \tag{4}$$

where ρ_0 is the residual, temperature independant, resistivity due to all the crystal defects, impurities etc and $\rho(T)$ is the lattice scattering component.

6. PHASE II ANALYSIS

With the assumptions mentioned earlier, the effective average cross-sectional area and effective average resistivity changes in the main part of the degradation lifetime can be computed and are shown for a typical specimen in Figure 1. The linear decrease in cross-section with very small change in resistivity agrees with the scanning electron microscope (SEM) results which indicate a *linear increase* in small but macroscopic defects, small voids, alloy inclusions, hillocks, gooves etc. (Ref. 3).

7. PHASE III ANALYSIS

In the final stage of the lifetime violent increases in resistance are observed usually followed by a decrease. This is well known from SEM studies to be associated with the formation and healing of very large voids. With the assumptions mentioned earlier the effective cross-sectional area, resistivity change and temperature rise in the localised area of the void can be calculated. The temperature rise can then be used to calculate the consequent resistivity rise. On the resistivity plot the difference between the measured resistivity change and that due to the temperature rise indicates the extra resistivity due to scattering. Typical sequences are shown in Figure 2 for an event which is completely healed and Figure 3 for an event in which there is only partial healing.

8. CONCLUSIONS

These electrical measurements can be interpreted together with SEM studies to show the general trend in electromigration degradation. There is an early and short period where there is an an-

nealing of strains caused by temperature changes after fabrication or due to the elevated temperature of the accelerated stress. In the main electromigration damage lifetime there is a linear rise in damage due to small defects. At some point these reach a stage where large voids can form. These voids are sufficient to constrict the cross-section by a factor of 4 to 10 and cause extremely large local heating. These voids heal and can either completely disappear or leave some permanent damage when they reform elsewhere. These voids grow larger as more form and at some stage the local temperature reaches the metal melting point and the metal goes open circuit.

9. REFERENCES

1. Jones B.K., Xu Y.Z. and Emonts P.G.A. "Characterisation of Electromigration Damage by Multiple Electrical Measurements", Proc. 3rd European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF 92) Oct 92, pps. 363-6.

2. Jones B.K., Xu Y.Z. and Denton T.C. "Electrical Measurements as Performance Indicators of Electromigration" Proc. 4th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF'93) Bordeaux Oct 1993, pps. 83-7.

3. Jones B.K., Shih W.C., Greer A.L. and Xu Y.Z. "Microstructure, damage and resistance during life-testing of Al-Cu interconnects", Materials Research Society Symposium Proceedings **309** (1993) pps. 369-376.

10. ACKNOWLEDGEMENTS

This work was part of Project IED 2/1/2140 and financed by SERC grant GR/F 99144 and we would like to thank our colleagues at the University of Cambridge, GPS Roborough and BNR Europe Limited. Financial support was also given under the ERASMUS programme and the work of the students on that scheme is also acknowledged.

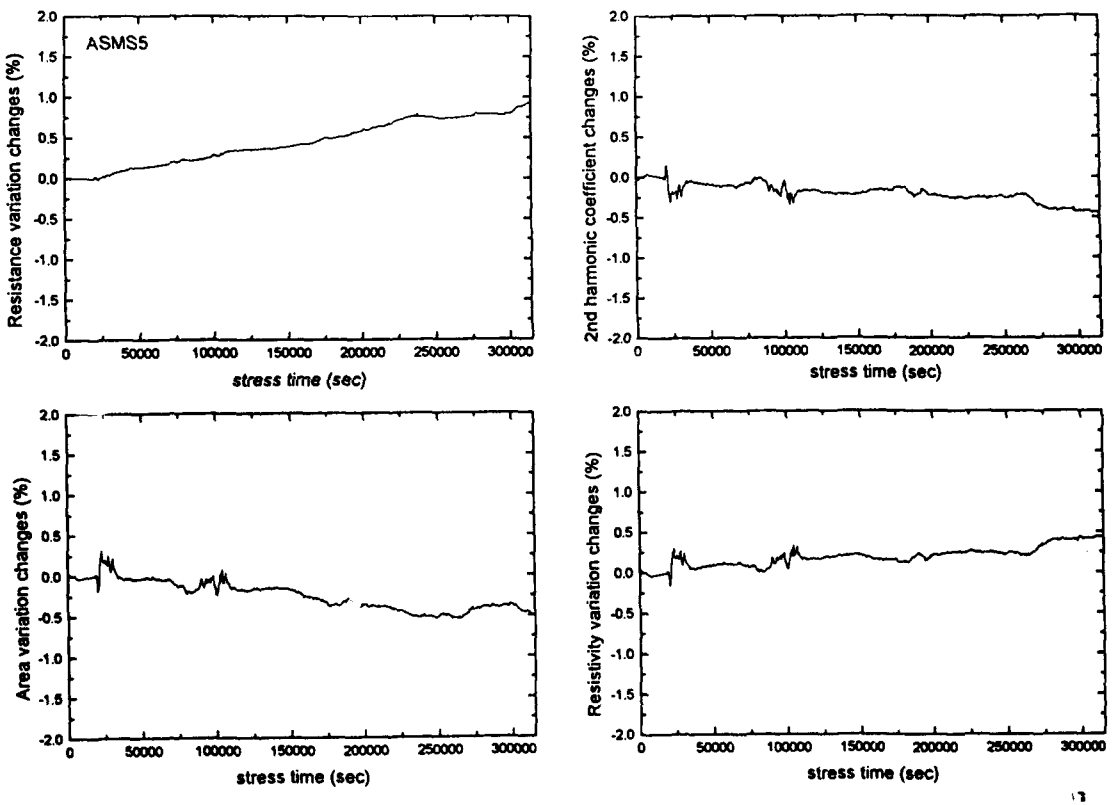


Figure 1. A typical analysis of specimen ASMS. during phase II. Shown are the resistance change $\Delta R/R\%$, the Joule heating change $\Delta \alpha\beta/\alpha\beta\%$ and the derived average cross-sectional area change $\Delta A/A\%$ and resistivity change $\Delta\rho/\rho\%$

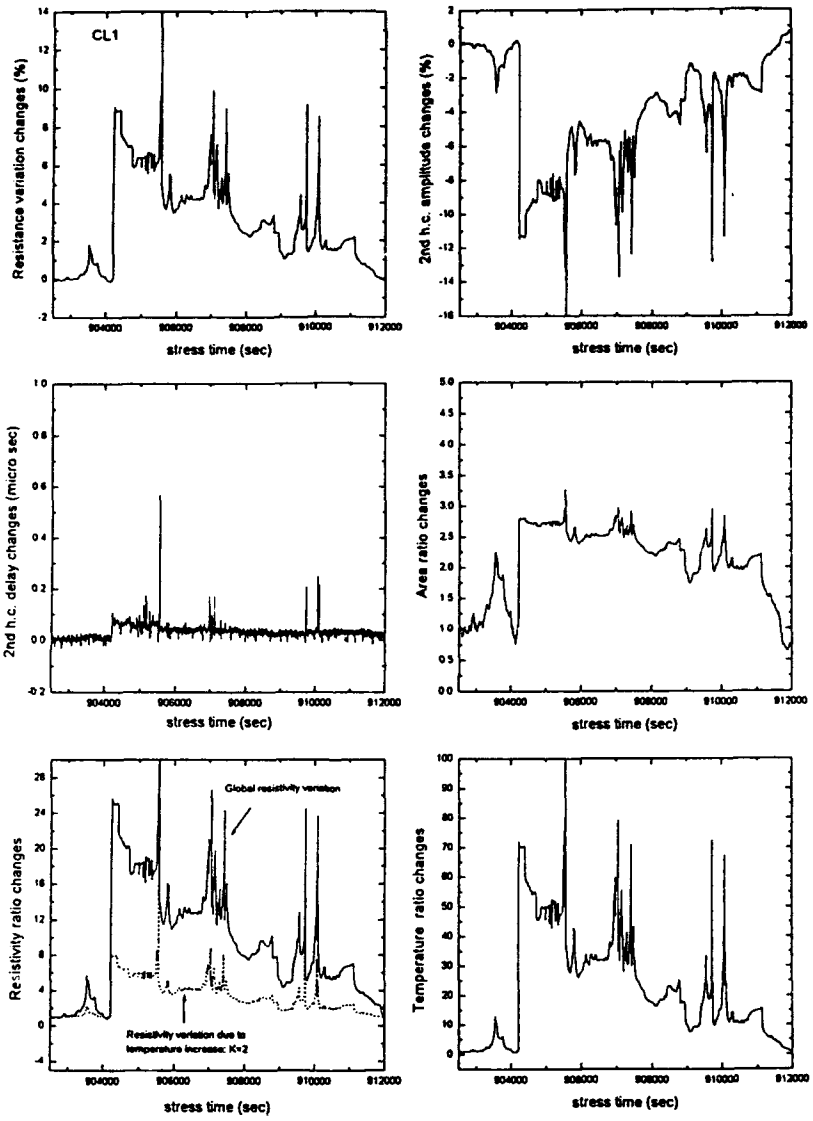


Figure 2. A typical analysis of specimen CL1 for a fully self-healing void formation event in phase III. Shown are the measured resistance change $\Delta R/R\%$, the Joule heating change $\Delta\alpha\beta/\alpha\beta\%$, the derived local cross-sectional area change A_1/A_2 where A_1 is the initial and A_2 the final area, the temperature ratio change $\delta T/\delta T_0$ where δT_0 is the initial Joule heating (about 10°C) and the resistivity changes ρ_2/ρ_1 where ρ_1 is the initial value and ρ_2 the final value for both the measured total resistivity and the derived part due to the temperature increase.

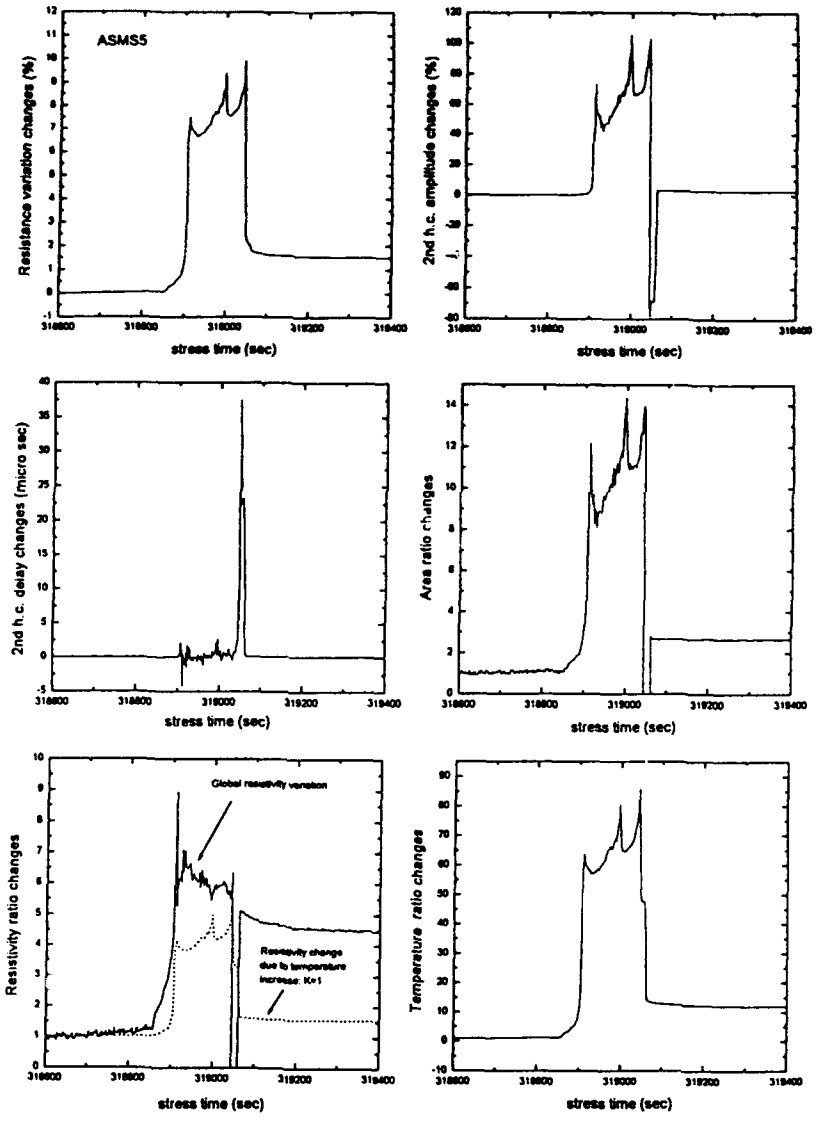


Figure 3. A typical analysis of specimen ASMS5 for a partially self-healing void formation event in phase III. Shown are the measured resistance change $\Delta R/R\%$, the Joule heating change $\Delta\alpha\beta/\alpha\beta\%$, the derived local cross-sectional area change A_1/A_2 where A_1 is the initial and A_2 the final area, the temperature ratio change $\delta T_0/\delta T_0$ where δT_0 is the initial Joule heating (about 10°C) and the resistivity changes ρ_2/ρ_1 where ρ_1 is the initial value and ρ_2 the final value for both the measured total resistivity and the derived part due to the temperature increase.

ELECTROMIGRATION IN AL BASED STRIPES: LOW FREQUENCY NOISE MEASUREMENTS AND MTF TEST

P.E.Bagnoli, C.Ciofi, B.Neri, G. Pennelli.

**Dipartimento di Ingegneria dell'Informazione:
Elettronica, Informatica, Telecomunicazioni
via Diotisalvi 2, I-56126, Pisa.**

ABSTRACT.

Four types of aluminum based samples have been analyzed by using both traditional (Median time to Failure) and not traditional (Noise measurements) techniques. The comparison bertween these two techniques has demonstrated that, if different stress conditions are used in the two cases, the results can be very different in terms of reliability index. This means that some typical noise parameters, evaluated for a given type of samples at relatively weack stress conditions, cannot be used to foresee the behavior of the same group of samples at strongly accelerated stress conditions. In particular it has been observed that unpassivated samples, containing 0.5% of Cu and 1% of Si, change deeply their noise characteristics when the test temperature overcomes 200 °C. Therefore it seems not possible, at the present state of the knowledge of the phenomenon, a straigthforward extension of the results obtained under weack stress conditions to accelerated stress conditions and vice versa.

1. INTRODUCTION.

In the last years a great effort has been devoted to the tuning up of techniques capable of characterizing the Electromigration (EM) in metal lines, in a short time and without using accelerated stress conditions. The result has been partially reached in that at least two different characterization procedures have been proposed, capable to provide information about the phenomenon in a quite short time (a few hours) and by using stress conditions not too different from the operating ones. The first one is based on the measurement of the resistance changes in the early stages of the phenomenon [1], by using high sensitivity instrumentation capable of displaying percentage resistance changes of a few hundreds of ppm. The second one is based on the measurement of the power spectral density of the resistance microfluctuations induced by the vacancy flux at the ends of a sample subjected to EM [2-4]. This second technique, named SARF (Spectral Analysis of Resistance Fluctuations), has been successfully employed to evaluate the activation energy of the phenomenon in different types of Aluminum based samples [3]. It has been also demonstrated that SARF technique is very sensitive to the different microstructural characteristics (grain dimensions) of samples obtained with sligthly different technological processes [5].

In order to use these non destruptive techniques in the field of the reliability, it is necessary to investigate the correlations, if any, between the information that they are capable to provide and some typical reliability parameter such as, for instance, the MTF (Mean Time to Failure). In this paper the preliminary results of a comparative study performed on different types of samples are presented.

In particular, the results of the MTF tests at 221 °C are compared with those of SARF tecnique in the temperature range betwee 70 and 225 °C.

2. EXPERIMENTAL RESULTS.

Four types of samples have been used: Al - Si 0.8% unpassivated (S) and passivated (SP), and Al - Si 1% - Cu 0.5% unpassivated (SC) and passivated (SCP). The test pattern had the following geometry: length = 1mm, width = 2 mm, the thickness was 0.6 μm for SC and SCP samples and 0.7 μm for S and SP samples. MTF tests, SARF characterization and SEM observations have been carried out on the four groups of devices.

2.1 MTF Tests.

The apparatus for the traditional MTF test has been purposely designed in order to drive, with a constant stabilized current, up to ten different samples simultaneously. The samples were placed in an electrically insulating oil bath whose temperature was controlled with an accuracy of 1 °C. The temperature stability was ± 2 °C.

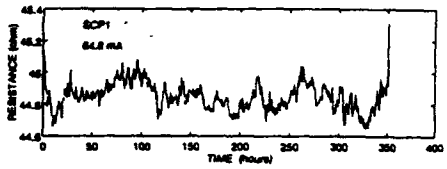
Each current driving unit consists of a voltage to current converter which employs narrow-band, low-offset operational amplifiers and power transistors. The current supplied by each channel can be separately and continuously set in the range 0.1 - 150 mA with an accuracy of about 0.1%. Using a computer controlled scanner and a voltage data aquisition system, the voltage drop across the sample voltmetric terminals, the supplied current and the oil bath temperature were measured for each resistance every five minutes. Therefore the system was able to monitor, during the entire life test, all the electrical parameters of the samples: voltage, current, resistance, power dissipation, environment and sample temperature.

For the MTF tests the temperature of the oil was set at 154 °C, and the nominal value of the current density was 4.6 10^6A/cm^{2}for all the samples. The temperature of the samples was about 65 °C above that of the oil due to the Joule heating. In order to compensate the effect of the differences of the electrical and thermal resistances of the samples, the current of each sample was suitably adjusted around the nominal value. In this way all of the samples were stressed at the same temperature, 221 °C, whereas the current densities of a given type of samples were in a range of about 10% around the nominal value.

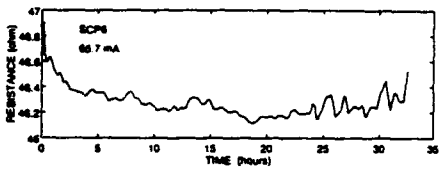
Fig. 1a and 1b show typical resistance vs time curves of two samples belonging to the SCP group (aluminum-silicon-copper, passivated) plotted until failure. All the passivated samples showed a nearly flat resistance time evolution with variations around the average value which did not exceed 1% and without a continous increase before failure. This behaviour, due to the effect of the passivation layer which strongly reduces the surface metal atom motion, is typical of passivated narrow metal stripes [6].

On the contrary, all the unpassivated samples showed a progressive resistance increase until breakdown.

Furthermore, as can be clearly seen in fig. 1b, the SCP lines showed an initial resistance decrease (about 1% during the first 20 hours of stress). This decrease has been also observed in another group of equal samples which had been annealed at the same temperature without current, during the preliminary operation necessary to tuning up the measurement system. Therefore it seems likely that the observed resistance decrease is not related to the phenomenon of electromigration, but it is due to an annealing effect, for instance a redistribution of Cu precipitate at the grain boundaries, promoted by the high temperature. In unpassivated samples this phenomenon is probably masked by the rapid resistance increase caused by the EM damage.



a)



b)

Fig. 1 Resistance behavior vs. time: a) the failure occurs without any preliminar resistance increase; b) initial resistance decrease in a SCP sample.

In order to verify this hypothesis, four samples, one from each of the four different types, were kept for 24 hours in a temperature controlled oven at 230 °C. The resistance R_0 at 0 °C and the temperature coefficient of the resistance α were measured before (R_{0I}, α_I) and after (R_{0F}, α_F) the temperature stress. The results are shown in Tab. 1 and can be summarized as follows : i) the changes observed in S and SP samples (a few part per ten thousand for R_0 and a few part per thousand for α, are comparable with the accuracy of the measurement system [7]; ii) the variations measured for SC and SCP samples are more pronounced (between 1 and 2% for R_0, and between 1.4 and 2.5% for α). Therefore the hypothesis that the resistance decrease observed during the first part of the MTF test of the SCP samples is due to an annealing of Cu precipitates is confirmed.

Table 1.

	S	SC	SP	SCP
R_{0I} (Ω)	23.71	22.31	22.96	23.58
α_I (K^{-1})	4.38 10^{-3}	4.22 10^{-3}	4.35 10^{-3}	4.17 10^{-3}
R_{0F}	23.72	22.18	22.95	23.14
α_F	4.39 10^{-3}	4.28 10^{-3}	4.37 10^{-3}	4.27 10^{-3}

In Fig. 2, the results from the MTF test are reported in a lognormal plot. The results related to the group of S

samples were not included in the figure because these samples showed a very low time to failure: the interruption of the line occurred in a few minutes, therefore it was not possible to adjust the current and the temperature at the prefixed values, as was done for the other samples.

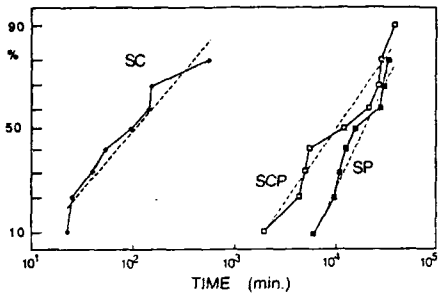


Fig. 2 Cumulative failures in a lognormal plot.

The MTF resulted $2x10^1$, 10^2, $9x10^3$ and $1.5x10^4$ minutes for S, SC, SCP and SP samples respectively. It is evident from the plots of fig. 2 the ability of the passivation layer in dramatically increasing the metal stripe lifetime.

As far as the role of copper in the metal matrix is concerned, the results of MTF tests showed an evident discrepancy between passivated and unpassivated samples. While in the unpassivated resistances the presence of copper causes an increase of the average lifetime, among passivated samples the SP group had a slightly higher lifetime and a lower standard deviation than the SCP one. This can be seen in fig. 2 from the comparison between the behavior of the cumulative failures graphs.

2.2 Noise Measurements.

Spectral Analysis of Resistance Fluctuations (SARF) technique has been employed to characterize the samples also under not accelerated stress conditions [2]. The main aim of this type of characterization was the evaluation of the activation energy E_a of the phenomenon. To this end, the low frequency voltage fluctuations induced by EM at the ends of the samples, supplied with a constant current, were measured versus the stress temperature. A fixed value of the current density ($j = 1.9 \times 10^6$ A/cm^2) was used in this case, whereas the temperature was changed in the range between 60 and 140 °C. A purposely designed micro-oven was used, whose temperature stability was sufficiently high to not impair the eccellent noise characteristics of the measurement system.

An attempt was made to evaluate E_a for each type of sample by using the procedure described in [2]. This procedure has been demonstrated to be valid when the frequency exponent γ of the noise spectra is close to 2. In Fig. 2 the Arrhenius plots needed for the evaluation of E_a are reported for S, SP and SCP samples. The quantity S_F on the y axis is related to the power spectral density of the resistance fluctuation of the line, S_R, through the relationship [2]

$$S_F = S_R \frac{kT}{\rho j} \qquad (1)$$

where ρ is the material resistivity and k the Boltzmann constant.

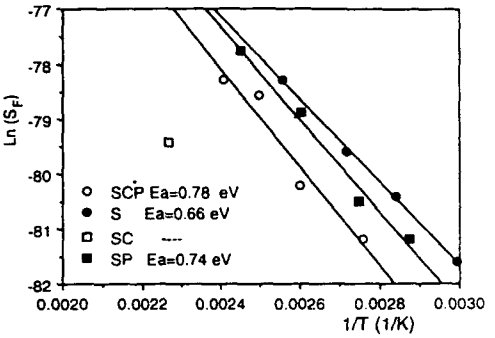


Fig. 3 Arrhenius plot for the evaluation of the activation energies.

Only one point is reported for the SC sample. In fact the model for the evaluation of the activation energy applies only if $\gamma \cong 2$, therefore only the experimental points obtained under stress conditions at which $1.8 \le \gamma \le 2.2$ are normally used to contruct the plot. In the case of SC samples, the frequency exponent rised from 1.5 to 2.4 in the explored temperature range and only for T = 440 K γ lay in the aforementioned range. This means that it has been not possible to evaluate the activation energy for this sample: the point in fig. 2 has been reported only in oder to allow a comparison between the noise level of SC samples and that of the other ones.

In the figure the values of Ea for S, SP and SCP samples are reported: they lie in the range 0.72 ± 0.06 eV. The first conclusions that can be deduced from the analysis of this figure are:

i) No apparent correlation exists between the noise level at a given temperature and the MTF; in fact, for instance, the SC samples which show the lowest noise level, are characterized by an MTF about two order of magnitude shorter than SP and SCP. In other words,

the widely accepted empirical rule *the higher the noise the lower the reliability* in this case fails.

ii) The samples for which the activation energy is measureable show a value of Ea about 0.2 eV lower than that of similar samples previously characterized in a lower temperature range [3].

iii) The passivating layer.seems to induce a ligth increase of the activation energy as already observed previously [3]. This is in agreement with the great increase of the MTF observed in passivated samples.

These first results were unsatisfactory because the initial target of this investigation was to search for a correlation between the MTF test and the low-frequency noise measurements. Unfortunately, unlike previously presented experimental studies [3,8], both the used noise parameters, noise level and activation energy, do not show any evident correlation with the reliability of the stripes. The reason is probably due to the fact that the degradation processes and, perhaps, even the microstructure of the samples, are significantly modified when the temperature and the current rise from the values used for the noise measurements to those used for MTF test.

For this reason, the noise measurements have been repeated at the same stress conditions used for life tests. The results obtained in this case must be considered as preliminary. In fact, whereas the low frequency noise measured at weack stress conditions ($j<2x10^6$ A/cm^2 and T<160 °C) was practically constant, at a given j and T, for all the samples of the same group, the same quantity at heavier stress conditions ($j=3.5x10^6$ A/cm^2 and T = 225 °C) show significant differencies among samples of the same group. Therefore, a complete analysis would have required the examination of a number of samples, for instance 10, for each group, as was done for MTF test. Such an analysis is now in progress and only the preliminary results will be reported here and summarized as follows:

i) type S samples can not be measured at these stress conditions because they break in a few tens of minutes;

ii) type SC samples are the most noisy;

iii) type SP and SCP initially show comparable level of noise, but after about one hour of stress the latter show an evident increase of the noise level, which is normally the signature of an irreversible damage, whereas the noise produced by the former remains constant for at least 3 hours. Therefore, the aforementioned empyrical rule in this case seems to be applicable.

2.3 SEM Observations.

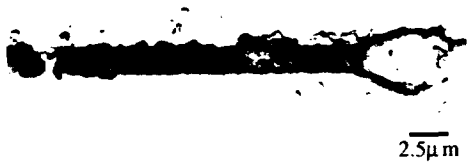
SEM observations have been performed on several samples after that the interrupion of the metallization line was detected during the MTF test.

The purpose of this investigation was to look for possible differencies in the failure mode of passivated and unpassivated AlSi and AlSiCu samples.

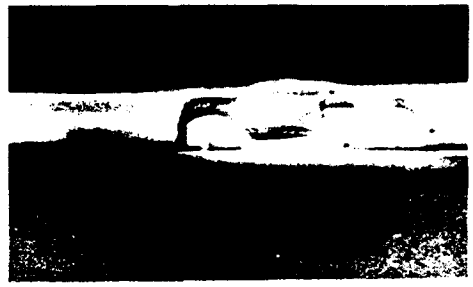
Each test pattern has been mechanically decapsulated and washed in cold acetone before SEM observations. In some cases, a thin gold layer has been evaporated on the sample in order to reduce charging effects.

This procedure has allowed us to perform accurate observations on unpassivated samples. In the case of passivated samples, however, only rather straightforward observations were possible because of the presence of the Si_3N_4 passivation layer which could not be removed with the available sample preparation facilities.

No apparent differencies were found in the failure modes of unpassivated AlSi and AlSiCu samples. In all the cases, extended voids and hillocks were present all along the stripe. Often, but not always, a quite big hillock is observed a few microns away from the location in which the interruption of the stripe is observed (Fig. 4a).



a)



b)

Fig. 4 SEM pictures from samples S (a) and SCP (b) (contrast is reversed).

In the case of passivated lines, the point in which the interruption of the metallization occurs is not easy to detect. However, cracks on the passivation are often observed. These cracks probably correspond to regions in which formation of hillocks has occurred. In some cases, as in Fig. 4b, the passivation appears to have "exploded", probably because of the excessive pressure generated by underlaying hillocks. The voids which can be noted in the picture, probably correspond to volumes previously occupied by single grains.

A more detailed microstructural characterization could have been performend by means of TEM observations. Howerver, because of the peculiar structure of the test pattern, the preparation of TEM samples is not straightforward. We are currently experimenting a few promising preparation procedures which should allow us to perform preliminary TEM observations in the very near future.

3. DISCUSSIONS AND CONCLUSIONS.

A first attempt to correlate the results of SARF technique with those of MTF test in Al based micrometric lines has failed. The reason of this failure is probably due to the fact that noise measurement and MTF test have been performed at quite different stress conditions, therefore the degradation process of the sample microstructure was deeply different. Particularly, in the case of Al-Si-Cu samples, which showed the most controversial results, it is well known that, when the temperature rises above a threshold, a diffusion and a redistribution of the Cu precipitate occurs. TEM analysis is now in progress to check this hypothesis. These results, if confirmed by further analysis, would demonstrate that it is not justified an extrapolation of the results of the noise measurement as well as of the MTF test at stress conditions significantly different from those at which the characterization has been carried out.

Finally, the preliminary results of a SARF characterization at the same stress conditions used for MTF test, seem to indicate that the noise level is greater for the groups of samples characterized by the lower values of the MTF. Further analysis are necessary, but a confirmation of these results would allow to substitute the extremely time consuming MTF test with much less time consuming noise tests.

REFERENCES

[1] V.D'Haeger et al., Proc. of ESREF 93, pp.141-145, Bordeaux, France, Oct. 1993

[2] B.Neri, A.Diligenti, P.E.Bagnoli, IEEE Trans. on Electron Devices, ED-34, pp. 2317-2322, 1987

[3] A.Diligenti, P.E.Bagnoli, B.Neri, S.Bea and L.Mantellassi, Solid State Electronics, 32, pp. 11-16, 1989.

[4] S.Chicca, C.Ciofi, A.Diligenti, A.Nannini, B.Neri, to be published on IEEE Trans. on Electron Devices.

[5] C.Ciofi, A.Diligenti, F.Giacomozzi, A.Nannini, B.Neri, Journal of Electronic Materials, 22, p. 1323, 1993.

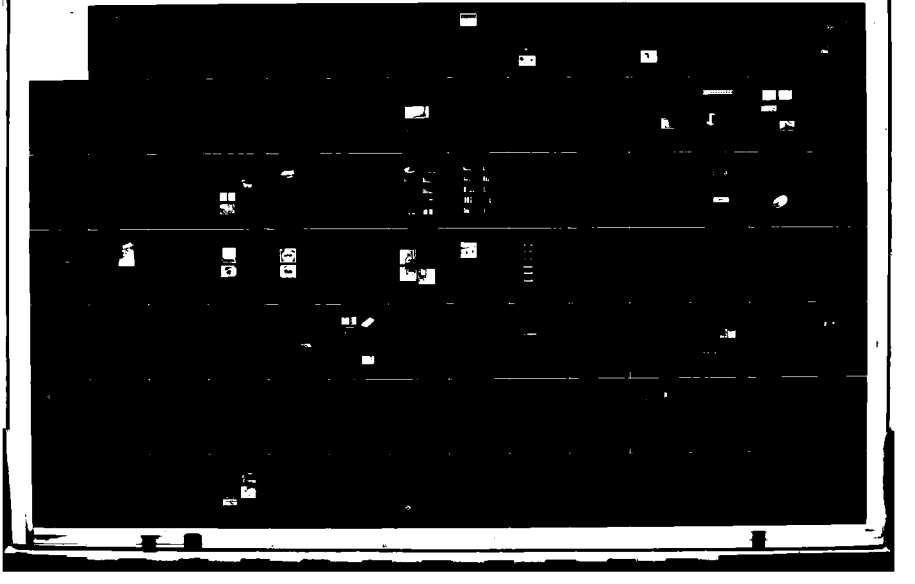
[6] G.Specchiulli, F.Fantini, Proc. of ESSDERC 87, p.217, Bologna , Italy, Sept. 1987.

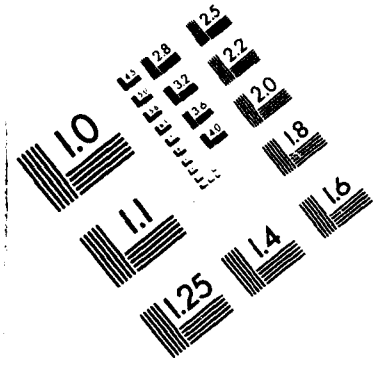
[7] A.Diligenti, B.Neri, A.Nannini, S.Ciucci, Journal of Electronic Materials, 20, pp. 559-565, 1991

[8] J.G.Cottle, N.S.Klonaris, Journal of Electronic Materials, 19, pp. 1201-1206, 1990.

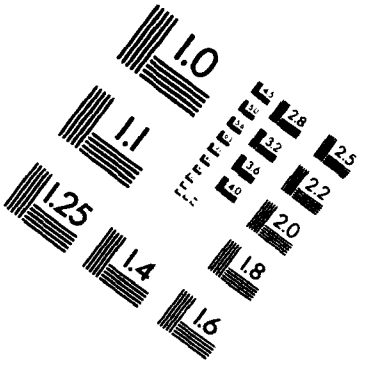
AD-A286 614 EUROPEAN SYMPOSIUM ON RELIABILITY OF ELECTRON DEVICES 5/7
FAILURE PHYSICS AND..(U) GAREL ASSOCIATES NORTHAMPTON
(UNITED KINGDOM) G M BRYDON 7 OCT 94 XB-ONREUR

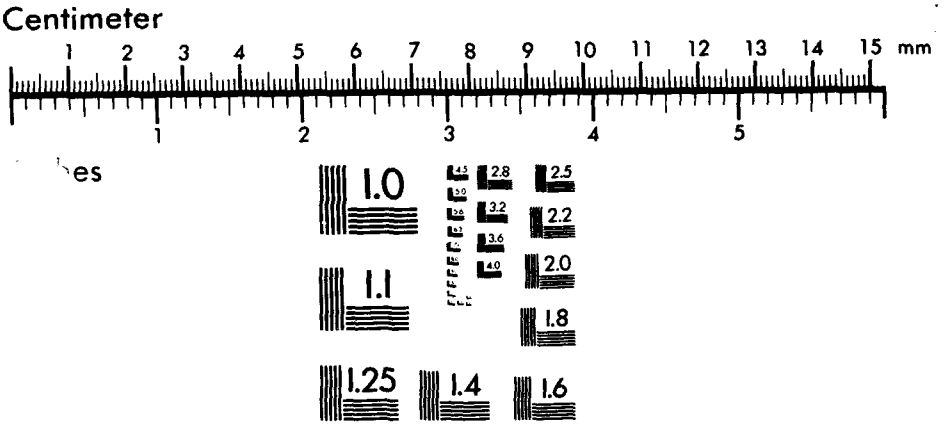
UNCLASSIFIED NL

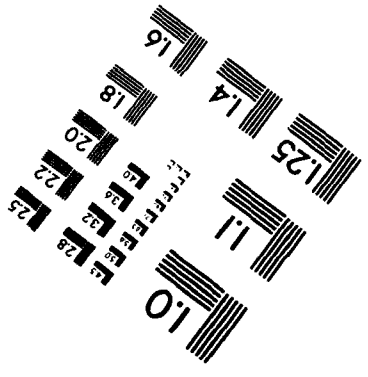


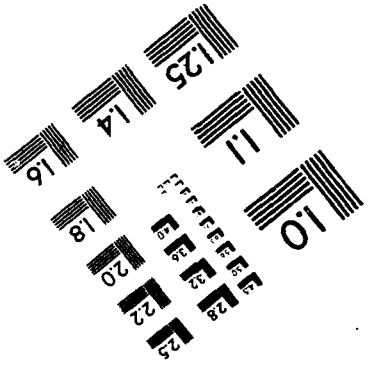


Association for Information and Image Management
1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202









MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

ELECTROMIGRATION MODELLING AS TOOL FOR PREDICTION OF THE FAILURE BEHAVIOUR OF METALLIZATION LINES

H. Hoebbel, V.Breternitz

Technical University of Ilmenau, Institute of Materials Science
PF 327, D- 98684 Ilmenau ; Tel.: +(3677) 69 3135, Fax.: +(3677) 69 3171

1. FUNDAMENTALS

In order to characterise the quality of a microelectronic device, it is necessary to know the median time to failure (mtf) as well as the type and the form parameters of the distribution of the individual time to failure. The lifetime depends on a great number of influence factors such as the material for metallization and substrate, line geometry, grain structure and stress conditions (current density, temperature, mechanical stress). In general, it is impossible to investigate completely the influence of an individual influence factor by means of experiments as well as to test systematically all possible parameter combinations. For that reason, it is useful to attempt a computer simulation of the electromigration processes.

The simulation model elaborated is based upon the solution of the transport equation for every grain boundary within a grain boundary network. At the location where the particle transport shows divergences there is a void or a hillock formation causing an increase in the electrical resistance of the stripe. The time leading to a defined percentage of relative resistance increase, e.g. 10%, is used as criterion of time to failure (ttf). A median time to failure (mtf) can be deduced if this procedure is effect for a certain number of comparable structures. In our case to each set of independent (stress and structural) parameters 100 simulations were done basing upon the same structure where the variation is performed by a random generation of the grain orientations.
Our work is based on a simulation model previously presented (Ref.1). The concept of calculation is oriented to short computing times. This fact and the opportunity of testing the behaviour of identical structures under different stress conditions make the model especially useful for a statistical analysis.

The modelling starts from the well-known facts that for a single component system (no diffusion term) and neglecting thermodiffusion the particle flux in a grain boundary can be described as a result of the action of an electron wind force and the mechanical stress as driving and backdriving mechanisms, respectively.

If the geometry of the grain boundary and the conditions at its ending points are known, the particle flux can be expressed in terms of changes of the numbers of particles at the triple points. The net number of particles in each triple junction is determined by superposition of the results of the adjoining grain boundaries.
It is assumed that the change of numbers of particles causes a generation of mechanical stress. After exceeding typical limits there are irreversible deformations resulting in the formation of hillocks and voids.
The grain structure is taken into account in terms of the angle α between the direction of electron flow and the grain boundary, the individual grain boundary diffusion coefficient D depending on the difference of orientation between adjacent grains (angle Φ), and the length l of the grain boundary.

2. COMPUTATION CONCEPT

For each time step it is necessary to know the local temperature and the electrical resistance of a defined line segment. Normally, this calculation, e.g. by finite elements methods, is implemented in the main procedure and done in each time step. In our program the temperature and the resistance of a void containing segment is calculated as a function of void diameter and the results are saved in the curves $T_{segm} = f(r_{void}$, line width) and $R_{segm} = f(r_{void}$, line width). During the execution of the essential calculation the temperature and the resistance can be read from these data. This procedure leads to an appreciable decrease of the program run time. This is especially important regarding the fact that the resistance has to be calculated by a complete integration in all segments in order to improve the precision. This optimisation of the program schedule permits to execute time expensive statistical analysis. For example, the calculation of the dependence of the standard deviation on the current density and temperature for 100 comparable structures, 20 current densities and 8 temperatures needs about 24 hours on a workstation IBM-R6000.

3. APPLICATIONS

The procedure presented above allows to discuss the influence of grain structure, line geometry and stress conditions on the failure behaviour in terms of the distribution function of ttf as well as of its parameters, namely mtf and standard deviation. The following results should serve as examples and give some insights into the capacity of the simulation and its applicability.

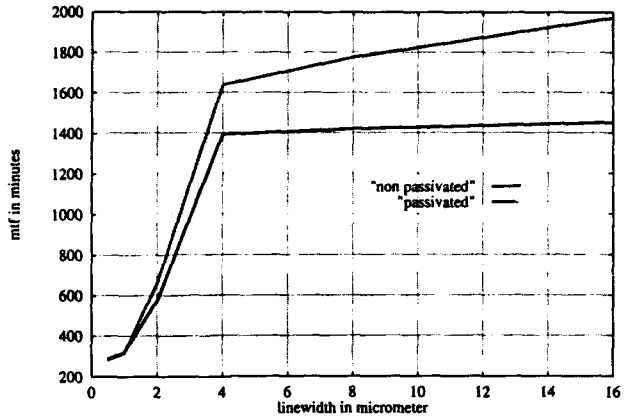


Figure 1: Dependence of mtf on linewidth for two different cases of mechanical stress

Figures 1 and 2 show the results of the simulation of the failure behaviour of a lot of comparable test stripes as a function of the line width as well as the influence of a passivation layer. In the model the case of non-passivated and passivated lines are distinguished by assuming different values for the critical stress limit of the hillock formation. It is supposed that this limit is equal to the surface tension of aluminium (approximately 2 MPa) or given by a supplementary stress representing the (mechanical) action of the covering layer (assumed to be 1000 times higher, i.e. 2000 MPa). The slopes of the curves fit rather well experimental results.

Although for the simulations the mean grain size is assumed to be approximately a half micron the known increase of the mtf for sufficiently small decreasing line widths is not observed since the special effects of the mass transport in bamboo or near bamboo structures are not yet considered.

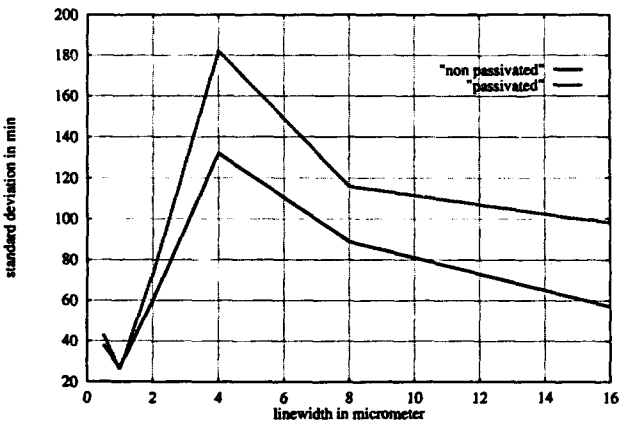


Figure 2: Dependence of standard deviation of the mtf (according to fig.1) on linewidth for two different cases of mechanical stress

The following two figures show as another example of the capacity of the program the dependence of the standard deviation on the current density and the temperature especially at low temperatures and low current densities where the tests are impossible because of the long execution time.

As already shown (Ref.1) the local temperature has an important influence on the time to failure. In the most of the industrial test facilities only the temperature of the substrate or of the furnace (T_s) is measured and used for the recalculation, e.g. by application of Black's formula.

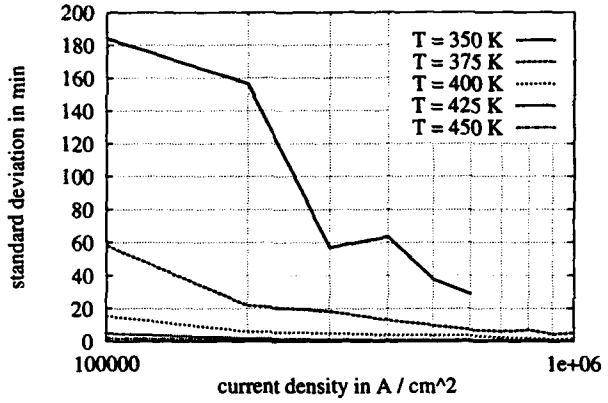


Figure 3: The dependence of the standard deviation on current density for several temperatures

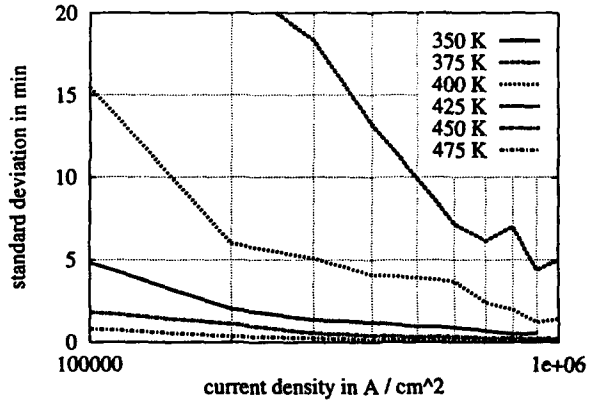


Figure 4: Zoom of figure 3 at low temperature range

In practice both mtf and the standard deviation of ttf have to be available in order to describe completely the failure behaviour of a test lot of stripes. The recalculation from accelerated tests to operation conditions gives only useful information if the distribution parameters are known for all stress conditions.

But our calculations have pointed out that the real temperature of the stripe can deviate distinctly from T_s. This fact applies in particular to the local temperature in the surroundings of a void which depends among other things on the size and the shape of the voids (Ref.2). During the calculation the temperature acts as accelerating factor by increasing the number of moving species

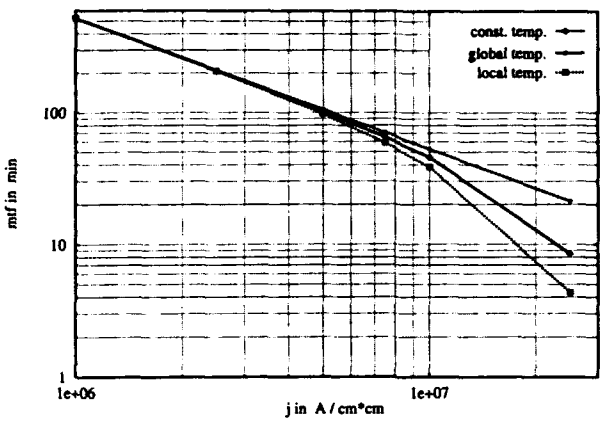


Figure 5: Simulation of mtf as a function of current density for several cases of temperature assumptions (see text)

as well as the diffusion coefficient according to usually applied Arrhenius set up.

Figure 5 shows the simulated slope of the function mtf = f (j) for three different cases.

1) assuming a constant temperature which implies an ideal heating transfer
2) considering the global temperature constant over the major part of the stripe as result of additional Joule self heating
3) involving the local temperature increase in the surroundings of the voids

The first effect leads to a linear dependence of ln(mtf) on ln(j) for the entire current density range. In the other two cases there is a deviation from the linear slope which is the strongest when all temperature effects are considered. Now the slope is in a good agreement with the results received from the measurement.

4. CONCLUSIONS

Although the Black's formula was not used explicitly as an input of the model the calculated dependencies confirm experimental results especially the range of values for the current density exponent n. The deviation of n can be described as consequence of the temperature profile including Joule heating and local temperature increase.

Our calculations show that, for test conditions typically used for accelerated test, a current density exponent n=2 can be assumed. But it is not constant over the full range and approaches n = 1 at sufficiently low current densities. The recalculation to operating conditions using a constant n = 2 involves the worst case for the prediction of the life time. However, present circuit technologies which are characterised by high claims of the customers to the quality standards require both a high life time and a prediction of the life time under normal conditions as exact as possible.

Because of the short computing times it is possible to calculate the mtf at operation conditions as well as under accelerated test conditions and, by that, to test the applicability and correctness of a recalculation method such as Black's formula.

5. REFERENCES

1. H. Hoebbel, V. Breternitz, Ch. Knedlik; "Computer simulation of the influence of stripe geometry and grain structure on the electromigration behaviour of aluminium metallization lines" proc. ESREF 93, Bordeaux 1993; pp.183-188

2. H. Hoebbel, "Ein Computermodell zur Simulation des Einflusse der Kornstruktur auf das Elektromigrationsverhalten von Leiterbahnen in integrierten Schaltkreisen, thesis, 1994

PLASTIC PACKAGING IS HIGHLY RELIABLE

Dr Nihal Sinnadurai
Professor of Electronics Technology,
Middlesex University Microelectronics Centre, England .[Fax:(+44) 473 211576]

1. THE NEED FOR A PACKAGE

The history of integrated circuits contains many chapters of unreliability arising from the exposure of the integrated circuit chip to harmful effects of the external environment. Early problems with gel filled packages led to the defined need for hermetic packages, and detailed reliability physics studies identified a hermeticity requirement of $5x10^{-11}$ millibar.litres/sec. However the inability of commercial leak testers to measure such low leak rates influenced the hermeticity specifications to be set at much higher levels, which have progressively been tightened to the present level of 10^{-8} millibar.litres/sec; but is that good enough? Read on for the answer. Meanwhile the inevitable pressure for cost reduction and increasingly automated manufacture, led to the development of plastic encapsulation by transfer moulding and other methods. Early suspicions that contaminants in the plastic and the permeation of moisture would cause failures were amply confirmed and led to an enthusiastic condemnation of the use of plastics by defence and telecommunications reliability authorities. Regrettably, the highly influential defence authorities retained that condemnation for decades despite having no new evidence to support that archaic viewpoint, while the rest of the world moved on. Today more than 80% of IC packages in the world are plastic and many are use in high reliability applications, with considerable evidence not only to support the high reliability obtainable with plastic encapsulations, but also the increasing evidence of field failures of hermetic packages satisfying the MIL883 hermeticity standards.

2. LABORATORY EVIDENCE OF RELIABILITY OF HERMETIC AND PLASTIC PACKAGED DEVICES

Reliability engineering of plastic encapsulation was pioneered during 1966-1985 at British Telecommunications Laboratories (BT Labs), England, with the active participation of major plastics materials suppliers. The evaluations were facilitated by the invention of the HAST reliability assessment technique by BT Labs in 1968. The high reliability capabilities were published in the early eighties, and plastic encapsulation was adopted for terrestrial telecommunications use at that time. The outcome of the work over many years, led to two developments:
(i) proven relationships for accelerated ageing by non-saturated damp heat within distinct validation limits for systematic acceleration[2].
(ii) strong evidence that discrete devices junction coated with particular plastic encapsulants were highly reliable[3].

The first such experiments which produced convincing evidence of high reliability were conducted on test vehicles comprising an assembly of npn and pnp transistors and specially designed moisture sensors, assembled onto ceramic substrates[3] (Figure 1). These were then coated with one of 15 different plastic coatings. The test vehicles were then subjected to high temperature and separate high humidity tests with electrical bias applied to the devices. Evidence from over 4000 hours duration of overstress testing on some 500 test vehicles showed that reliability equivalent to 25 years in tropical climates was achieved by some four silicone plastic encapsulations, while most of the other materials continued to be hazardous to device reliability (Experiment 1, Table 1, Figure 2). A second series of experiments (Experiment 2) used the better junction coating materials from the first experiment, selected for their compatibility with a range of top coatings which were added to provide "screwdriver" protection. The results from the dual coated test vehicles showed that, of the materials tested, two combinations - a silicone plus filled silicone, and the silicone and a filled phenolic - demonstrated high reliability also equivalent to over 25 years in tropical climates (Table 2, Figure 3, Figure 4). On the basis of such work, British Telecommunications adopted the use of plastic encapsulated integrated circuits (ICs) in high reliability applications and opened the door in Europe to the acceptance of plastic encapsulations for high reliability applications. Others followed this example and also subsequently published promising results on the reliability of plastic encapsulations[4,5,6,7,8].

The identification of high reliability plastic coatings led to the invention of the EPIC, a low cost and high performance, plastic chip carrier (Figure 5) which was

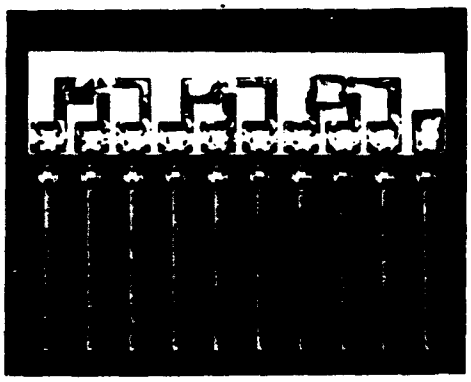


Fig 1. Test Vehicle for Junction Coating Evaluation npn, pnp and moisture sensors

Coating Code	Coating Type	Curing Temp °C
A	Rigid silicone	175
B	Semi-flexible silicone	175
C	Flowing silicone	175
D	Thixotropic silicone	175
E	Silicone-epoxy specially adapted	160
F	Poly-paraxylene	—
G	Heat-resistant polyimide	175
H	Unfilled epoxy (junction or top coat material)	160
J	Filled epoxy	160
L	Filled silicone (junction or top coat material)	175
M	Thixotropic epoxy dip (top coating material)	120
D+H	Silicone+Epoxy	160
D+M	Silicone+Epoxy	120
R	Proprietary combination: junction+top coatings	—
S	Proprietary combination: junction+top coating	—

Total of 15 batches of 16=240 test vehicles

Table 1. Junction Coatings Tested in the First Experiment

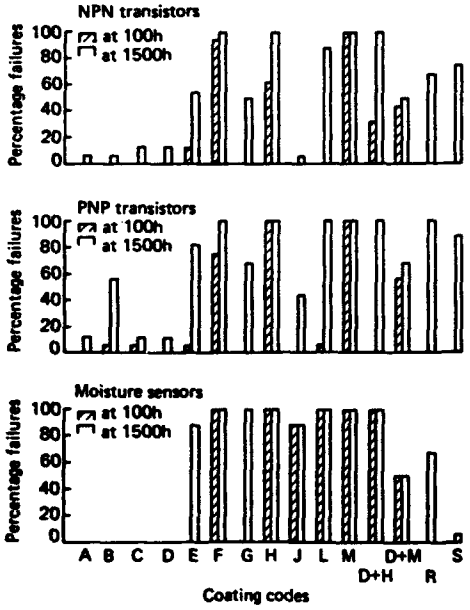


Fig 2. Bar Charts of Percentage Failures of the Junction Coated Test Elements

Coating Codes	Coating Types	Final Curing Temp °C
C+L	Thin silicone+Filled silicone	175
C+M	Thin silicone+Thixotropic epoxy	120
C+N	Thin silicone+Fusing epoxy powder	100
C+P	Thin silicone+Phenolic dip	160
D+L	Thixotropic silicone+Filled Silicone	175
D+M	Thixotropic silicone+Thixotropic epoxy	120
D+N	Thixotropic silicone+Fusing epoxy powder	100
D+P	Thixotropic silicone+Phenolic dip	160
J+L	Filled epoxy+Filled silicone	160
J+M	Filled epoxy+Thixotropic epoxy	120
J+N	Filled epoxy+Fusing epoxy powder	100
J+P	Filled epoxy+Phenolic dip	160
K*+L	Silicone RTV+Filled silicone	175
K*+M	Silicone RTV+Thixotropic epoxy	120
K*+N	Silicone RTV+Fusing epoxy powder	100
K+P	Silicone RTV+Phenolic dip	160

Total of 16 batches of 16=256 test vehicles
*K was substituted as a very promising replacement for A which had been discontinued by the plastics manufacturer, during the exercise.

Table 2. Combination of Good Junction Coatings With Added Top Coatings Tested in Second Experiment

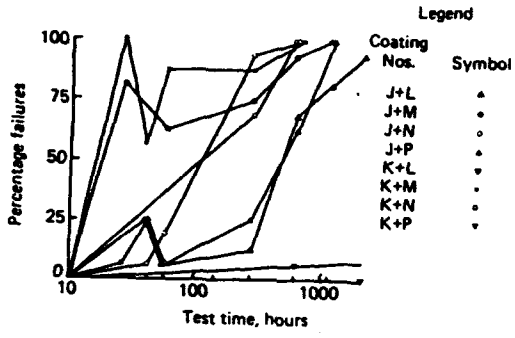


Fig 3. Cumulative Percentage of Parametric Failures of Test Vehicles in Second Experiment

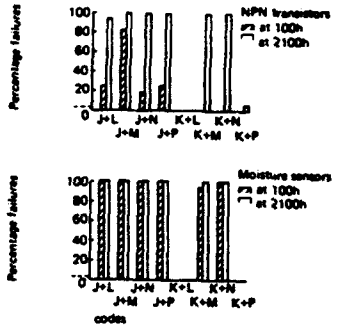
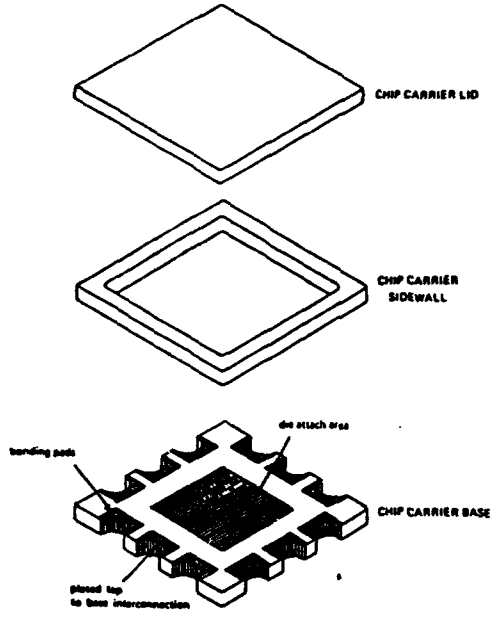


Fig 4. Bar Charts of Percentage Failures of the Junction + Top Coated Test Elements



Principle of EPIC chip carrier construction.

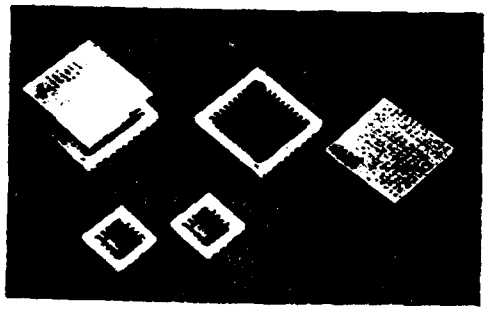


Fig 5. EPIC Chip Carrier

HIGH TEMPERATURE WITH BIAS (HTB) (150°C)

Source Code	Batch Size	Package Type	IC Type	IC Source	Metallisation + Passivation
C	20	SOIC	741	B	Al + Oxide
	20	SOIC	748	B	Al + Oxide
G	20	SOIC	741	B	Ti/Pt/Au + Si_3N_4
	20	SOIC	748	B	Ti/Pt/Au + Si_3N_4
Q	40	SOIC	741	E	Al + Oxide
H	40	SLCC	741	B	Al + Oxide
K	40	SLCC	741	B	Ti/Pt/Au + Si_3N_4
R	20	HCC	348	V	Al + Oxide
S	20	HCC	348	V	Al + Oxide

Table 3 High Temperature & Bias Testing of Micropackages

HIGH HUMIDITY TESTING IN NON-SATURATING AUTOCLAVE (108°C, 90%%RH)

Source Code	Batch Size	Package Type	IC Type	IC Source	Metallisation + Passivation
A	40	SOIC	741	B	Al + Oxide
	40	SOIC	748	B	Al + Oxide
D	40	SOIC	741	B	Ti/Pt/Au + Si_3N_4
	40	SOIC	748	B	Ti/Pt/Au + Si_3N_4
P	40	SOIC	741	E	Al + Oxide
F	40	SLCC	741	B	Al + Oxide
J	40	SLCC	741	B	Ti/Pt/Au + Si_3N_4
N	20	SLCC	741	L	Al + Oxide
	20	SLCC	748	L	Al + Oxide
T	20	SLCC	348	V	Al + Oxide
U	20	EPIC	348	V	Al + Oxide
W	20	SLCC	3096	Y	Al + none
X	20	EPIC	3096	Y	Al + none
M	20	DIP	741	L	Al + Oxide
	20	DIP	748	L	Al + Oxide
Z	15	EPIC	348	V	Al + Oxide

Table 4 High Temperature, Humidity & Bias Testing of Micropackages

fabricated by standard printed wiring board techniques[9].

Reliability assessments of the EPIC were carried out alongside other commercially available packages containing the same species of IC[10]. The results of both dry heat and damp heat tests (Table 3 and Table 4) show that a certain commercially manufactured surface mount micropackages were capable of achieving high reliabilities in both temperate and tropical climatic conditions, while there still remained a number of suppliers unable to achieve adequate reliability even in temperate climates (Figure 6 and Figure 7). It is particularly important to note that the silicone junction coated ICs and one manufacturer's plastic moulded small outline IC package (SOIC) outperformed the hermetically packaged ICs. The mechanism of chemisorption of silicones onto the silicon passivation surfaces was identified as the adhesion and protection mechanism with silicone encapsulations, while Van der Waal adhesion forces provided the adhesion mechanism for epoxy encapsulations. The fact that hermetically packaged ICs failed in the high humidity tests clearly showed up the false assumption that MIL883 tested hermetic packages do not need to be subjected to humidity tests (the vulnerability of MIL883 hermetic packages to external moisture is described later in the paper). In fact all packages must be subjected to all the tests if they are to prove their ability to survive in the climates represented by the accelerated test conditions (the accelerated test conditions are described later in this paper). The reliability tests were accompanied by full failure analyses and showed weaknesses in some of the plastics materials, such as ionic and halide contamination, and also the excellent and intact survival of silicone junction coated ICs (Figure 8). The reliability tests also established the credentials of the materials specification M219F[11].

The IEEE Gel Task Force took up the challenge later that decade to undertake evaluations of polymer gel coatings for integrated circuits[12]. The Task Force, formed under the auspices of the Computer Packaging Committee of the IEEE Computer Society, comprised representatives from 24 companies in the USA. Some 1440 IC chips containing a specific test patterns and protected by different glassivations, were tested with five gel types and one silicone RTV. The test patterns were intended to reveal simple corrosion and breaks in the wires. The evaluation tests conducted on the encapsulated test vehicles comprised thermal shock, salt spray and HAST (non-saturated autoclave).The outcome was that the more rigid and thicker coatings caused wires to be broken. It was clear that thick silicone RTV coatings, applied without dilution, caused strain and damage to the wires (Figure 9). This was consistent with the recommendation from the original work[3] that junction coatings should be applied thin, barely to cover the wirebonds on the chip, i.e. about 25 microns thick. However three of the thinly applied silicone gels did indeed achieve good protection of the test vehicles (Figures 10 & 11), the test vehicles without any glassivation over the metalisations performing the poorest of these. Of the better performing glassivated test vehicles, it appeared that the nitride passivations on the test vehicles in the Task Force exercise were of poor quality and showed stress fractures through which corrosion products had penetrated under the coatings. The Task Force undertook no determination of the physical or chemical adhesion or protection mechanisms of the encapsulations. The evaluation did establish that gel coated devices would withstand acceleration forces up to 15000 G.

Recent work by British Telecommunications on low cost packaging for optoelectronic components[13] took encouragement from the IEEE Task Force results on gels and the pioneering work by the predecessors at BT[3], and proceeded with the evaluation of the use of the better gels as reported by the Task Force report and applied them to PIN and laser diodes and GaAs ICs. The accelerated ageing conditions and the acceleration factors to predict longevity were based on the earlier work[3]. Tests of the coated PIN diodes in damp heat conditions, showed that three of the four gels provided excellent reliability protection (Figure 12) equivalent to many decades of longevity, provided the acceleration factors for non-hermetic silicon devices from the earlier tests were applicable. The lasers showed inconsistent behaviour (Figure 13), with the degradation occurring cumulatively in the small population of devices. The investigator was encouraged by the survival of 3 of the 4 gel coated lasers beyond 3500. The damp heat tests on gel coated GaAs ICs showed a remarkable stability of the ICs up to 6000 hours (Figure 14). These encouraging observations led to the development of low cost assembly of an optoelectronic circuit board with the individual active devices (PIN, Laser diode and GaAs logic ICs) encapsulated in gel.

Silicone and indeed some epoxy encapsulations of the multiple chips of multi-chip modules (MCMs) provide high reliability protection not only for longevity in temperate climates but also in tropical climates and the high reliability demands of satellite communications[14]. Of course space is quite a different environment, with the hostility arising from launch and manoeuvre stresses and from radiation. Nevertheless, plastic encapsulation by the use of junction coatings (compliant thin coatings directly applied to the surfaces of the dice) offer very high protection of semiconductor devices. Glob Top is an unfortunate terminology, because it has been associated with low technology low reliability applications.

The potential for space applications arises because reliable dice can be protected by an intimate

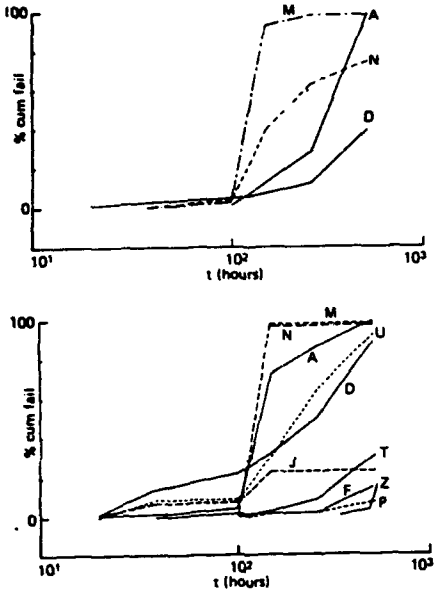


Fig 6. Cumulative Failures with Test Time of Micropackages Tested According to Table 4.

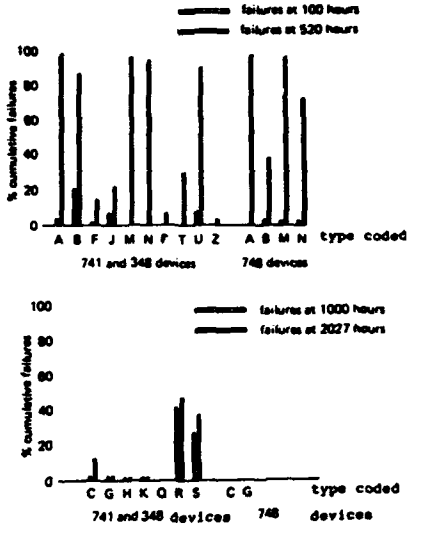


Fig 7. Bar Charts of Percentage Failures of the Micropackages According to Tables 3 and 4.

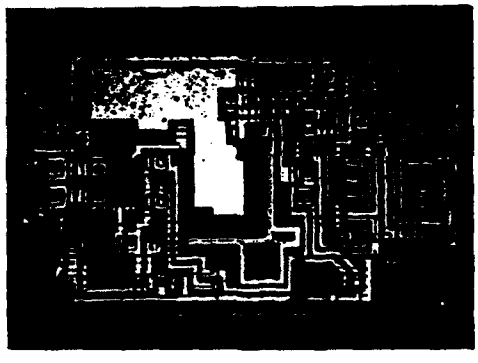


Fig 8. Plastic Decapsulated Reliable IC.

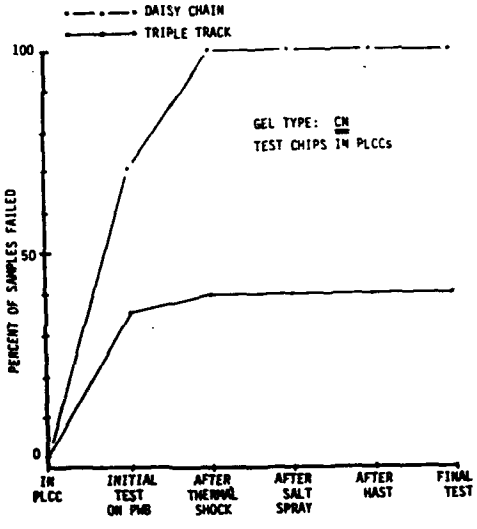


Fig 9. Cumulative Failures of Thick Silicone Coated Vehicles of IEEE Gel Task Force[12]

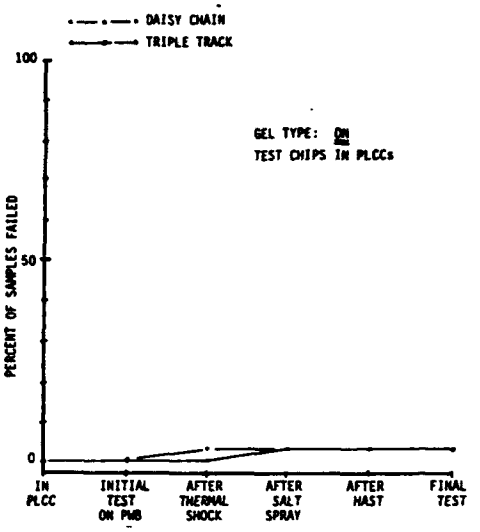


Fig 10. Cumulative Failures of Thin Silicone Gel DN Coated Test Vehicles of IEEE Task Force[12]

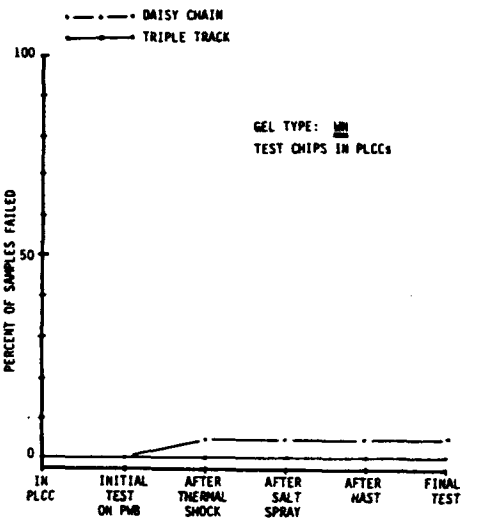


Fig 11. Cumulative Failures of Thin Silicone Gel WN Coated Test Vehicles of IEEE Task Force[12]

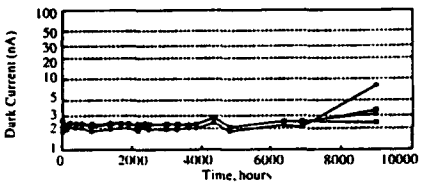


Fig 12. Dark Current Parametric Behaviour of Reliable Junction Coated PIN Photodiodes[13]

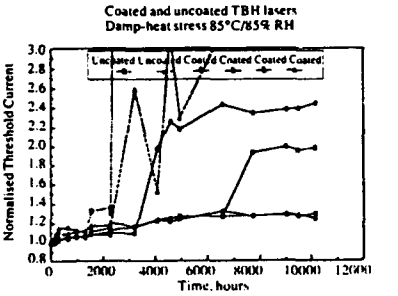


Fig 13. Threshold Current Behaviour of Junction Coated Laser Diodes[13]

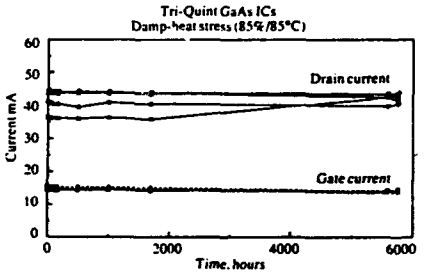


Fig 14. Supply Current Behaviour of Reliable Junction Coated PIN Photodiodes[13]

application of a junction coating. There are now many excellent silicones which form intimate bonds with the surfaces of semiconductor dice by chemisorption. This requires the surfaces of the dice to be chemically clean, which can be achieved by either wet processes or ideally by a final plasma clean. Application in vacuum facilitates intimate contact. Coatings must be thoroughly cured, which means beyond the manufacturers' recommended curing times (which are defined for a high percentage of polymerisation). Effective curing times are, e.g. 150°C for 30 mins + 3 hours at 180°C. Such a coating achieves an intimate and robust seal which will safeguard the existing reliability of the dice against external climatic attack. For space applications, the encapsulation adds mechanical support of wire or TAB bonds, and adds the major advantage of attenuating particles which can give rise to soft errors. α particle activity of silicones is very low, of the order of 0.002 per cm^2 per hr for 1-8 MeV energies. On the other hand, silicone effectively attenuates α particles, for instance, resulting in zero errors in 10^6 device hours of operation, compared with 200 soft errors if the silicone protection was not present. As α particles can be sourced also from alumina substrates and lead solder terminations, the silicone adds versatility to the range of materials that may be used in the assembly of the MCMs.

The options for MCM materials for use in space are still wide open. The payload advantage currently remains with silicon substrates using thin film multilayers with either SiO_2 or polyimide dielectrics. Silicon substrates also provide advantages of heat removal. Nevertheless, a number of low dielectric constant polymers are providing exciting opportunities for high performance. The Radiation Hardness of a number of MCM polymer materials have been tested to differing degrees of severity. For instance, polyimide has been found to retain essential properties with dosages in excess of 100 MRads of γ radiation, while BCB materials have been tested successfully with dosages in excess of 100 MRads of γ radiation. Therefore there are few fears for the inherent useability of polymers in satellite communications applications.

3. FIELD EVIDENCE OF THE RELIABILITY OF HERMETIC AND PLASTIC PACKAGED DEVICES

3.1 Failures of installed telecommunications equipment in India due to hermetic packaged IC failures.

The major investment by the Indian Department of Telecommunications (DOT) in acquiring digital switching systems for the planned expansion and modernisation of their telecommunications network was achieved by importing systems and large scale technology transfer. The digital system was supplied as a mature technology using components, including mandatory hermetically packaged ICs, to MIL883. However, no attempt was to adapt the technology to the conditions of supply and use in India. Fortunately, an United Nations Development Programme (UNDP) expert advisor scheme was set in place at the same time, which led to an effective scheme being set up in Bangalore for reliability assessment, field failure data retrieval, failure analysis, and analysis of the climatic conditions relevant to the use of electronics in India.

DOT set up SCRAM (System for Component Reliability Analysis and Measurement)[15]. Controlled retrieval of information from the installed base of digital telephone exchanges analysed by SCRAM, has produced the distribution of failures by component type shown in Figure 15, in which it can be seen that the predominant contributor to failures were ICs, followed by hybrid modules, transistors and capacitors. The data was further analysed into classifications of defects for all components, and an example is shown for ICs (Figure 16). For the period in question, the IC failure rate alone amounted to 1755 Fits (1 Fit = 1 failure in 10^9 device hours) in a population of 2.35 million ICs.

Failure analyses confirmed the IC hermeticity met the MIL883 criterion of 10^{-8} millibar.litres/sec with only 3 exceptions. The analyses also showed that the dewpoints within the packages ranged from -20°C to +30°C, corresponding to relative humidities that will have reached as high as 95% during operation. Detailed physical analysis of some 10000 components, including more than 2000 ICs, were carried out over five years at the Failure Analysis Laboratory. The failure mechanisms of the active devices (ICs, hybrid microcircuits, transistors, diodes) were distributed as shown in Figure 17. Apart from those failing due to electrical overstress, attributable to a poorly suppressed electrical environment, the most consistent failure mechanism observed was corrosion of the aluminium metalisations.

The analyses of the devices and the environment showed that, after manufacture, the packages gained moisture sourced from the high levels encountered during transport and storage in India, entering through the "acceptable" leaks in the package body. The moisture retained within the package will have started a train of corrosion even during storage, and accelerated when the equipment was electrically operated as the equipment was brought into service.

The manner in which hermetic packages can gain moisture is illustrated by calculations of the water penetration into packages having a leak rate equal to the MIL883 specified level. Even a modest external ambient moisture of 25°C, 70%RH can cause a gain of 5000 parts per million of water into such a hermetic package within 2 months[16] (Figure 18). The hermeticity required to properly safeguard active

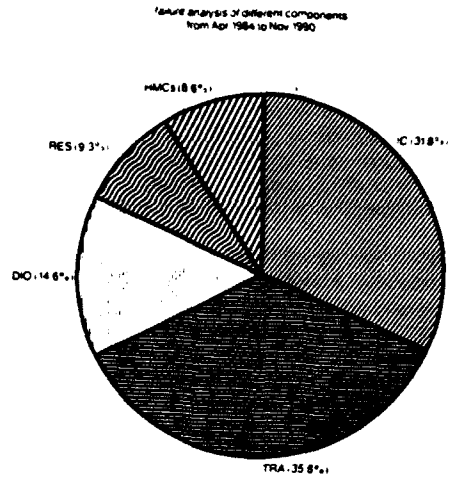


Fig 15. Pie Chart of Distribution of Failures According to Component Type

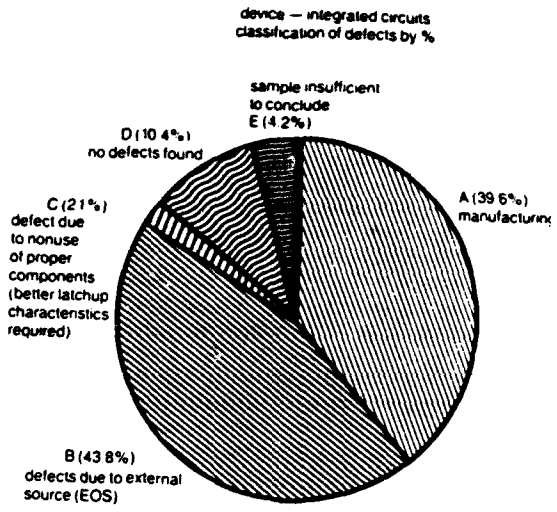


Fig 16. Pie Chart of IC Defects

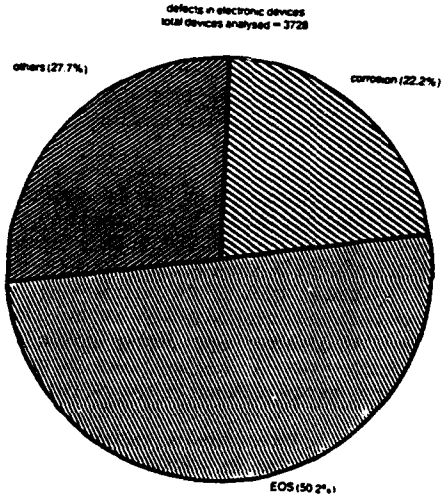


Fig 17. Pie Chart of Cause of Component Failures

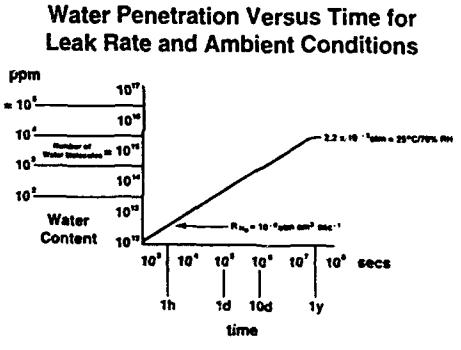


Fig 18. Moisture Entry Into Package At MIL883 Specified Leak Rate

devices for high reliability applications must be less than 10^{-11} millibar.litres/sec.

3.2 Analysis of the climates and conditions for assuring reliability in severe climates

The problems of the more severe climatic conditions were anticipated well before the failures actually occurred. A programme to analyse the climatic distribution, average climatic conditions, and conditions representing reliability safeguards covering 95% and 99% of the occurrences likely in India, was therefore already well in hand. Meteorological data obtained over 30 years from 162 locations, weighted according to the geographic area they represent, has been analysed. The outcome of this work has been reported[17]. This data converts into Table 5.

Table 5. Conditions for reliable operation, within specification, averaged over the required life

Climatic Condition	Covering 95% Times & Places	Covering 99% Times & Places
Hot Humid	29°C, 83%RH (33 mbar)	29°C, 86%RH (34 mbar)
Hot Dry	40°C	42°C

The information from the climatic analysis has been applied to determine test conditions for both equipment and components. The particular specification for environmental testing conditions for telecommunications equipment is the DOT specification QM-333[17].

The climatic conditions given in Table 5 have been applied to the well established equation for reliability assessment by damp heat testing of non-hermetically packaged ICs[18], in order to calculate the detailed stress test conditions to simulate 20 years operation. Inserting the BT Labs HAST stress test conditions into the equation gives:

$$t_s = 175000/\exp\{0.00044[(90)^2-(RH_{amb})^2] + 7000(1/T_{amb}-1/381)\}$$

where:- t_s is the required duration of the stress test
- RH_{amb} is the ambient humidity from Table 5
- T_{amb} is the ambient temperature from Table 5 converted to absolute units

the applied stress condition of 108°C, 90%RH, in a non-saturating autoclave HAST test[2] gives more dependable but more severe damp heat acceleration of ageing than is obtainable with humidity chambers.

Hence, the test durations for THB overstress tests to simulate 20 years (175000 hrs) operation in various climates have been calculated (Table 6).

Table 6. THB Reliability Tests for 20 Year Survival in Different Climatic Conditions

Climate	THB Overstress Condition 108oC,90%RH Time (hours)	85oC,85%RH Time (hours)
Temperate General (12°C,72%)	..100	500
Temperate Equipment Room (30oC,25%)	60	300
Tropical Coverage 95% India (29oC,83%)	850	4100
Tropical Coverage 99% India (29°C,86%)	1000	5100
Tropical Severe (35°C,90%)	2000	10000

Clearly the non-saturating autoclave HAST test is considerably more efficient than the 85°C,85%RH humidity test. The extensive base of experimental work that has been conducted with the apparatus has also established its credentials for robustness and precision of control.

It is an unfortunate fact that the nations which can least afford the problems of severe climatic conditions are those which require greater and therefore more costly safeguards of component and equipment reliability. The evidence presented here, accumulated over many years, has clearly revealed the vulnerability of supposedly reliable components used in modern equipment. The Indian (and other tropical) climate requires more severe testing of electronics components than is applied abroad.

3.3. Field results from indigenous equipment using plastic packaged microcircuits.

In order to establish systems more relevant to the dispersed rural needs of India, its diverse and severe climatic conditions, and the need to achieve a high storage and operational reliability of its equipment, the Indian Government established the Centre for Development of Telematics (C-DOT)

C-DOT set to with significant determination and ability to develop a modular digital telecommunications switching system, starting from a building block suitable for rural locations not having air conditioning. The C-DOT developers made the bold decision to use plastic packaged ICs and that the component technologies should be able to operate without air conditioning and survive the rigours of the diverse climatic conditions to be encountered in India.

Components were evaluated using accelerated ageing tests appropriate to assuring 20 years reliability in uncontrolled climatic conditions, determined during the course of the work described herein. The equipment was tested in walk-in chambers to the conditions specified in QM-333. The choice made by C-DOT thereafter was to approve the manufacturers of the components that met the evaluation criteria.

The switching systems were successfully developed and have been in the field for two years. Modularity has enabled switches with 10000 line capacity to be built. The reliability results in the field have been promising with one glaring exception.
- so far, the IC failure rates are less than 1 Fit out of 154000 line circuits.
- hybrid microcircuit failure rates amounted to 600 Fits during $3x10^9$ device hours
- the alarming exception was the rogue failures of a particular type of memory IC which passed the functional tests upon assembly into the memory cards, but then failed on installation into the equipment. Allowing a generous amount of 10 hours for survival upon installation, yields a component failure rate of $8x10^6$ Fits! Because a single IC failure caused the memory card to fail, 40% of the memory cards failed. The manufacturer of these ICs was ostensibly a major international manufacturer, as the package logo and code showed. However, tracking down the source proved otherwise! The logo was an imitation, and the components, while having the same function, were fake. The finding was a severe lesson that manufacturer approval alone, and the lack of routine vigilance and assessment, is a dangerous gamble. Nevertheless, the excellent behaviour of the authentic ICs proved the bold decisions to use known good plastic packaging, was wise. Thus there is now an awareness, alertness and responsiveness to the needs of more demanding climates, and for cost-effective ways of achieving them.

3.3. Field data from non telecommunications installations using plastic packaged microcircuits.

Benefits are already being realised, for instance, in an 8 billion Franc project in France in which 1 billion Francs is being saved by using reliable plastic packaged ICs in portable military communications. Defence applications already have a 5 year history of the use of plastic packaged devices in missiles with no field failures. The reported[20] increasing military use shows that the final bastion of resistance to the sensible use of low cost high reliability plastic packaging, has been penetrated.

The relative reliability of plastic packages at 55°C can be scaled as follows:

simple devices	1 Fit
memory devices	10 Fits
ULSI logic & µP	100 Fits

Packages using the purer[11] moulding compounds have an average reliability of 40 Fits. This can be improved to 30 Fits by the use of good quality Si_3N_4 passivation.

Automotives are also increasingly using plastic encapsulation even in the hostile environment under the bonnet. A good example is the Ignition Control hybrid microcircuit of the Fiat Cinquecento. The active integrated circuit is a bare-chip-and-wire assembly which is junction coated with a silicone encapsulant. Some 500000 vehicles have had no system failures in the field in two years.

4. DISCUSSION AND CONCLUSIONS

The traditional requirement for hermetic packaging may now be overturned in favour of plastic packaging on many grounds:
(i) the inadequacy of the hermeticity specification.
(ii) the full range of laboratory evidence of high reliability of silicone junction coated ICs undertaken by a number of researchers, including the author, showing that plastic encapsulations may now be used for microelectronics and optoelectronics in telecommunications and also in automotive, military and space applications as the better option in many instances.
(iii) the demonstrated field failures of "MIL883" hermetic packaged devices, and the massive ingress of moisture, which caused many system failures.
(iv) the demonstrated better field reliability in rural tropical climates, of assessed, standard plastic packaged ICs from major manufacturers.

The invention of HAST has proved to be an invaluable tool in the drive to develop high reliability plastic packaging. Combined with the climatic analyses, there are now clear methods for evaluating plastic packaged device reliability for various climates. Of course, tests only provide the basis for selecting the right technologies. The essence of reliability engineering is to develop the technologies and build in the reliability improvements.

Fortunately the good prospects of achieving the required reliability at low cost followed from the initiatives described earlier in this paper. These initiatives have been followed and recognised in the influential work which followed[12,15].

It is a profound fact that each day's production of plastic packaged ICs is sufficient to provide sufficient data on parts per million defects, whereas the tedious qualified components route for hermetic package evaluation can take three years to obtain the same magnitude of data. The inexorable momentum will make the undisputed encapsulation of future ICs will be the plastic package.

REFERENCES

1. F N Sinnadurai, "The Accelerated Ageing of Semiconductor Devices in Environments Containing a High Vapour Pressure of Water", *Microelectronics and Reliability*, Vol. 13(1), pp23-27, 1974.

2. N Sinnadurai, "More Than a Decade of the Non-Saturating Autoclave as a Highly Accelerated Stress Technique for Evaluating the Reliability of Non-Hermetic Microelectronics Components", *Microelectronics and Reliability*, Vol. 24, No. 5, 1983.

3. Nihal Sinnadurai, "An Evaluation of Plastic Coatings for High Reliability Microcircuits", *Microelectron. Journal*, Vol 12, No 6, pp30-38, 1981.

4. H Baudry and G Kersuzan, "Individual Encapsulation for Integrated Circuits: How to Use Thermal Analysis to Optimise Curing Conditions Towards Reliability", *Proc. European Hybrid Microelectronics Conference*, pp111-121, 1983.

5. S M Boyer "Plastic Encapsulation of Active Devices on Hybrid Microcircuits", *Proc. European Hybrid Microelectronics Conference*, pp122-132, 1983.

6. J J Granger, J C Basset, N Vimont and P Viret, "Wire Bonded Chips Encapsulation Using Resin Droplets", *Proc. European Hybrid Microelectronics Conference*, pp146-154, 1983.

7. O Mallem and J Lantaires, "The Plastic Composite Package Provides an Economic Alternative to LSI-VLSI Packaging for High Density Microcircuits", *Hybrid Circuits*, pp5-12, No.8, September 1985

8. Paul Collander, Marja K H Huhtanen, Raimo A Mäkelä, Sauli M Palo, "Humidity Testing of Plastic Coated Integrated Test Circuits", *Proc International Microelectronics Symposium*, pp249-254, 1987

9. Nihal Sinnadurai, "EPIC: A Cost-Effective Plastic Chip Carrier for VLSI Packaging", *IEEE Trans. Components, Hybrids and Manufacturing Technology*, Vol CHMT-8, No.3, pp386-390, 1985.

10. N Sinnadurai and D Roberts, "Assessment of Micropackaged Integrated Circuits in High Reliability Applications", *Microelectronics Journal*, Vol 14, No. 2, pp5-25, 1983.

11. British Telecom Specification M219F for "Type Approval of Encapsulating Resins, Coatings and Adhesive Materials for Semiconductor Devices and Microcircuits", *BT Group Technical Services.*

12. J W Balde, "The Effectiveness of Silicone Gels", *IEEE Transactions on Components, Hybrids and Manufacturing Technology*, Vol 14, No. 2, June 1991.

13. I P Hall "Novel Low Cost Techniques for Low Cost, High Performance, Optoelectronic Component Assembly", *Hybrid Circuits*, No.34, pp44-47, May 1994.

14. Nihal Sinnadurai, "MultiChip Module Application in Satellite Communications", *NATO Advanced Research Workshop on MCM/Mixed Technologies*, Islamorada, Florida, May 1994

15. T K Ramaswamy and T S Kuppuswamy, "Quality Aspects of Telecom Equipment", *Telecommunications (the Indian Journal of Telecommunications)*, pp14-18, February 1991.

16. D Stroehle, "On the Penetration of Gases and Water Vapour into Packages with Cavities, and on Maximum Allowable Leak Rates", Proc. IEEE 15th Annual Reliability Physics Symposium, pp. 101, 1977.

17. "Report on Climatological Conditions Encountered by Telecom Equipments in India", *General Manager TQA Circle, Department of Telecommunications*, Bangalore, India, Issue 1, (Sept 1988)

18. "Specification for Environmental Testing of Electronic Equipment for Transmission and Switching Use", Telecom Quality Assurance Circle, Department of Telecommunications, Bangalore, India, QM-333, Issue 1 September 1990.

19. F N Sinnadurai, "Handbook of Microelectronics Packaging and Interconnection Technologies", *Electrochemical Publications Limited*, 1st ed., pp. 6-7, 150, 153-155, 1985

20."Plastic-packaged ICs in military equipment", *IEEE Spectrum*, pp62-63, February 1991.

OPTICAL CURRENT DENSITY PROBING AT INTERFACES IN BONDINGS AND SOLDER JOINTS: INVESTIGATION OF AGEING MECHANISMS.

W. Claeys, S. Dilhaire, D. Lewis, V. Quintard, T. Phan, A. Fenech*.

W. Hasse+ and K. Weide+

Centre de Physique Moléculaire Optique et Hertzienne (CPMOH)
URA-CNRS 283
*Laboratoire d'Etude de l'Intégration des Composants et Systèmes Electroniques(IXL)
URA-CNRS 846
University Bordeaux I and CNRS
351 cours de la Libération, F-33405-TALENCE Cedex, France.

+Institut für Halbleitertechnologie und Werkstoffe der Elektrotechnik
University of Hannover
Appelstrasse, 11a, 30167 Hannover, Germany.

1. INTRODUCTION.

This work presents an original technique of current density measurement in metallic assemblies like solder joints. This development is to study the evolution of current flow inside connection assemblies during ageing tests. The measurements are performed upon microsections showing the different metals. Electrical measurements are quite difficult upon such structures because the variation of potential between close points is very small and because a contact measurement may influence the result. We propose a contactless method based upon a laser reading. The current density amplitudes are indirectly derived from measurements of thermoelectrical effects generated in the structure.

The thermoelectrical effects which take place in metal assemblies are the Joule, the Peltier and the Thomson effect. The first is very small in metals as the local resistance is negligible. The Thomson effect needs high temperature gradients to be significant, this is not the case in the situation we are interested in. Our experimental approach is therefore focused upon the detection of the Peltier effect which is produced at the interface of metals.

This thermal effect is experimentally observed through the thermal expansion of the structure. This is done by measuring the surface displacement with a high resolution interferometric laser probe (Ref 1). We have shown in earlier work, (Refs 2, 3), that it is possible to measure the Peltier heat expansion selectively from the Joule one, with our laser probe.

The basic idea of this paper is to measure the Peltier thermal expansion generated at the interface of two metals crossed by an electric current. This is only valid for assemblies of metals of different kind. As the Peltier coefficient depends only upon the two metals involved, the heat deposited by this effect at a given location, will be proportional to the current density crossing the interface of the two metals at this point. So by mapping the interface Peltier expansion, we map the current density crossing the interface.

In this paper we show the feasibility of the method upon a microsection of a solder joint. A solder joint is a pile up of three different materials, lead, solder, and a copper line. We present a Peltier response map of the microsection and profiles of current density crossing the interfaces.

A finite element numerical simulation method (ANSYS) has been used to calculate the current density map of the solder joint section. Good agreement with the measurements is observed, showing the method to be valid.

We have shown, in previous papers (Refs 2 and 3) concerning solder joints, the Peltier expansion response to short current pulses to be a good early ageing detection method. The purpose of this work is to provide quantitative experimental data in order to investigate ageing mechanisms inside solder joints. The future goal of this approach is to study how ageing does influence the current flow inside bondings, joints and other connecting devices.

2. PRINCIPLE OF CURRENT DENSITY MAPPING.

In order to explain the principle of the optical current density mapping we propose, let us consider a situation like the one sketched in figure 1. Two pieces A and B of different metals, having same thickness z_0 are linked together over a certain length x_0 (the interface size is $x_0 \times z_0$). The ends of both metals are connected to a current generator. The current has to flow through the interface of the two metals and it is the Peltier thermal expansion produced there that we detect.

The current densities are, due to symmetry reasons, independent of z for a given x. If we assume a contact potential Φ_{AB} between A and B, then for cosine current excitation $I_0 \cos(\omega t)$, the Peltier surface power density $P_P(t, x)$ given off at location x upon the interface will be:

$$P_p(t,x) = \Phi_{AB} J(x) \cos \omega t \quad (1)$$

with $J(x)$ the current density.

This energy exchange is mixed up with the one resulting from the Joule power density $P_J(t,x)$ at (x, y=0):

$$P_J(t,x) = \rho[J(x)\cos\omega t]^2 = \rho[J(x)]^2\left[\frac{1}{2} + \frac{\cos 2\omega t}{2}\right] \quad (2)$$

with ρ the metal resistivity.

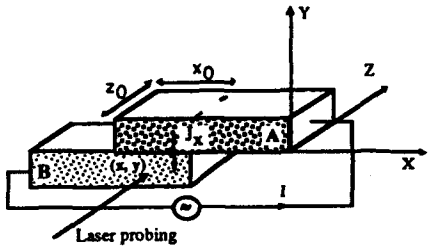


Figure 1: Sketch of experimental situation for Peltier current density probing at the interface of two metals A and B.

We see from relations 1 and 2 that the Peltier contribution can be distinguished from the Joule one by detecting the signal at the first harmonic. The power variation over time (relation 1) of the Peltier effect produces temperature variations in the structure at the same frequency. These variations are the same for all z for a given x. We take advantage of this by measuring the thermal expansion resulting from the temperature variations of the component in the z direction. We point the laser probe in the z direction and use the front surface as reflector. The expansion amplitude at location x upon the y = 0 line will be proportional to $\Phi_{AB} J(x)$ according to relation (1), and as a consequence the profile along x will map the current density profile.

3. MEASURING METHOD.

The expansion measurements are performed with our interferometric laser probe presented in earlier work (Ref 1).

3.1. Experimental test structure.

We have chosen to study the same test structures as in previous work (Refs 2, 3), which were Daisy chained components mounted on an Isolated Metallic Substrate. We have made a microsection so as to have access to the metallic interfaces separating different materials. It can be seen as a pile up of different elements: lead, solder, copper, insulator and substrate (Figure 2).

3.2. AC thermal excitation.

The electrical connections are taken between the top of the lead and the copper line near the tip. We applied to this test structure AC current excitation of about 200mA amplitude. We use a synchronous detection and select the first harmonic response. A good sensitivity is needed because the expansion amplitudes are less than one picometer. The choice of the excitation frequency is very important as we want to get a high dilatation amplitude and a very short diffusion length. The diffusion length of a thermal wave is the distance over which the wave amplitude has dropped by a factor e^{-1}. The two requirements above are antagonist with respect to the frequency as the expansion amplitude depends on $\frac{1}{f}$ and the diffusion length depends on $\frac{1}{\sqrt{f}}$. We have chosen the frequency equal to 10 kHz as a good compromise, the diffusion length is then of the order of 30 μm.

3.3. Synchronous detection at harmonic one.

The use of a lock-in amplifier allows a good selection of the Peltier response with respect to the Joule one (2f). In fact we note that under the same excitation conditions the Joule expansion is too small to be observed.

4. RESULTS.

This current density measuring method holds also for components having stacked metals like a solder joint. Figure 2 shows a sketch of a solder joint section. Current flows from the lead L through the solder S to the copper C. Peltier sources, proportional to the local current density, are located along the lead-solder and solder-copper borderlines

We have measured first harmonic thermal expansion amplitudes at different locations upon the solder joint section. Figure 2 shows a map of these expansions upon a solder joint. From this map it is clearly seen that the current mainly flows at the tip and heel of the foot representing the solder joint section. This 3D map shows the general trend of current flow together with other features difficult to understand at the present state of our investigations. We therefore have undertaken more detailed measurements along the metals interfaces.

Figure 3 shows a picture of a section of the solder joint. As the Peltier detection allows to measure the current density along the borderline of two different metals, we have measured the first harmonic thermal expansion along line AC (lead-solder) and along AB (solder-copper). Figures 3b and 3c show the results. They show with more details the observations already seen in the map of figure 2. There are two main locations where current flows, at the end of the lead and at the curved part of the lead. The existence of two main routes for the current is rather surprising. We show in the next section that this is predicted by modelling.

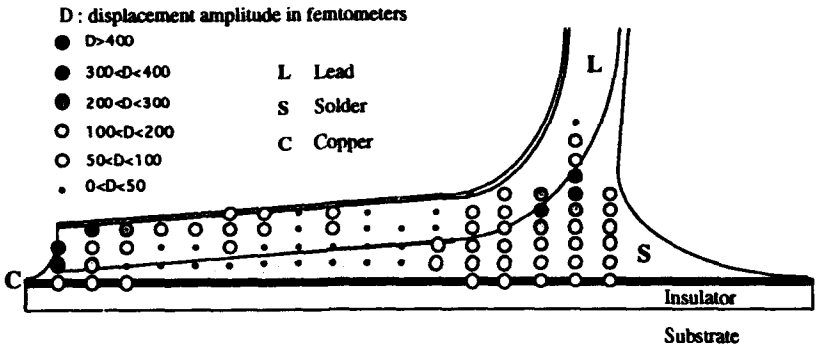
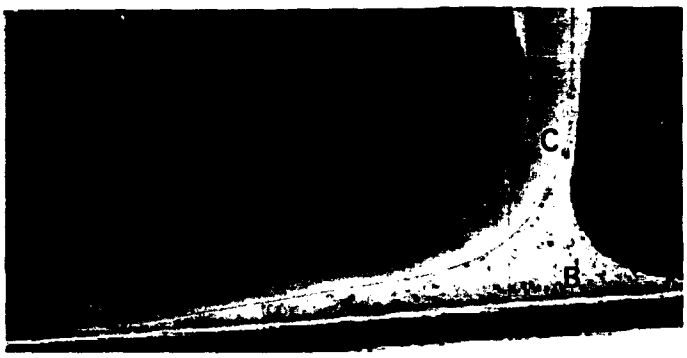
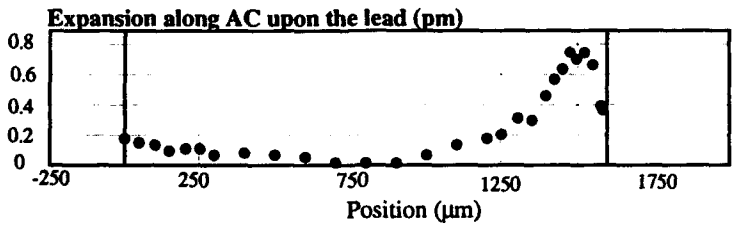


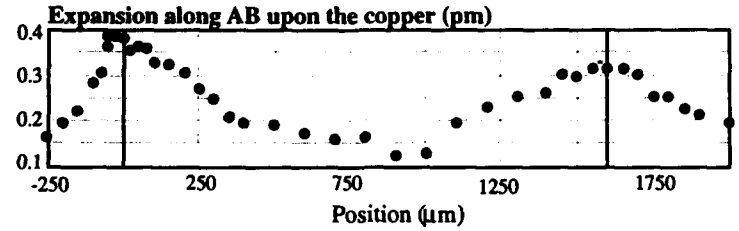
Figure 2: Peltier expansion map of solder joint indicating current flow



a/



b/



c/

Figure 3: a- Solder joint section, b- Expansion amplitude profile along AC interface, c- Same along AB interface. Expansion units are picometers.

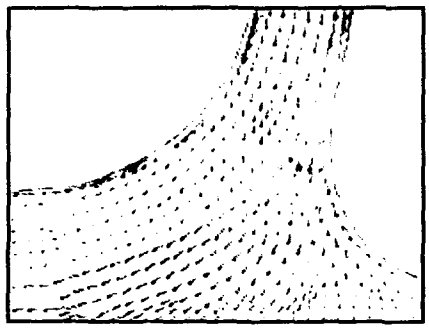
5. SIMULATIONS.

A two dimensional Finite Element (FEM) simulation of the solder joint was performed using the commercial FEM program ANSYS. We used the thermal-electric element type PLANE 67 for the model. A potential of 0V was applied on the left border of the model at the copper plate (see Figure 2). On the top border of the model a voltage of 0.8mV was applied at the lead. This boundary conditions do correspond to a induced current of 170mA. The bottom of the copper plate was considered as a heat sink at room temperature. The used material parameters are given in table 1.

	unit	copper	lead	solder
Thermal conductivity	W/mK	370	15	51
electrical resistivity	μΩcm	1.7	76	17

Table 1: Material parameters used in the FEM simulation.

The temperature dependence of the material parameter was neglected, because a temperature rise of less than 1°K was observed for the lead case. The upper part in figure 4 shows the calculated current density map in the "heel" of the solder joint, while the lower part shows the same map at the "tip" of it. The overall trend measured in figure 2 agrees with this calculations. It is important to remember that the Peltier thermal response only measures the current density component normal to the metals interface.



a

b

Figure 4: calculated current density map inside a section of the solder joint presented in figure 3 Map a- shows "heel" of the joint. Map b - shows "tip" of the joint.

Therefore an important Peltier signal is seen at the "tip" and "heel" of the "foot" of the solder joint. This is clearly seen in figure 4.

We show in figure 5 the calculated absolute values of the component of the current density normal to the interface AB (copper-solder). These results have to be compared with the experimental results of figure 3c.

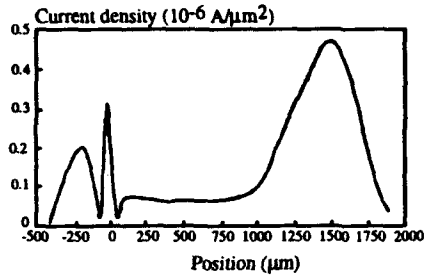


Figure 5: calculated absolute values of the component of the current density normal to the interface AB.

The sharp structures seen around position zero in the calculations are not reproduced in the measurements. This is due to the fact that our measurements are indirect, we measure periodical thermal expansion consecutive to Peltier energy exchange. This means that we measure an average of the thermal effect over the diffusion length of the thermal wave. This produces broadening of the structures. We plan however to try measurements at higher frequencies where the diffusion length is shorter. Besides this, the agreement is very satisfactory and gives confidence in the experimental method.

6. CONCLUSION AND PERSPECTIVES.

We have shown that the measurements of the Peltier thermal expansion upon microsections of metallic assemblies, like solder joints, are capable to provide current density profiles along the metals interfaces. This new technique gives a good insight upon the functionality of solder joints. The good agreement with simulation gives confidence in this measuring method. The main power of this method will be in the study of the evolution of the current density map during thermal cycling ageing tests. We plan to explore how the current density profiles evolve during ageing tests, the study will be particularly interesting in the region where structural changes occur (Refs 2, 3, 4).

7. REFERENCES

1. Claeys W., Dilhaire S. and Quintard V. "Laser probing of thermal behaviour of electronic

components and its application in quality and reliability testing", Microelectronic Engineering 24, 1994, pp 411-420.

2. Claeys W., Dilhaire S., Quintard.V., Hijazi A. and Danto Y."Early detection of ageing in solder joints through laser probe thermal analysis of Peltier effect" ESREF 93, Bordeaux -Arcachon 4-7 October , 1993, pp 513-517.

3. Claeys W., Quintard V., Dilhaire S., Lewis D. and Danto.Y. "Early detection of ageing in solder joints through laser probe thermal analysis of Peltier effect."Quality and Reliability Engineering International, 1994, accepted for publication.

4. Hijazi A., Danto Y., Ousten Y. and Allamargot B. "Thermomechanical failure analysis of SMD solder joints" Proceedings of ESREF 92, Schwäbisch-Gmünd, Germany, 1992, pp327-330.

DETERMINATION OF THERMOMECHANICAL BEHAVIOUR OF MICROELECTRONICS PACKAGING BASED ON MICROSTRUCTURAL ANALYSIS

A. Fenech, A. Hijazi, Y. Danto
Laboratoire de Microélectronique IXL
Université Bordeaux I, URA 846-CNRS
351 Cours de la Libération,
33405 TALENCE, FRANCE

Abstract : The thermomechanical stress in large leaded packages submitted to thermal cycling were studied, from both experimental ageing and ANSYS finite elements simulations. A degradation law, based on the observation of the microstructure evolution in the SnPbAg solder joint was derived and used as an early degradation indicator.

1. INTRODUCTION

Due to the thermal expansion mismatch between the different materials of the electronic assemblies, cyclic thermal loads induce, in the structures and specially in the solder joints, cyclic stress and strain, leading progressively to degradation of the solder alloy. The major failure mechanism is caused by the thermal expansion mismatch between componants and PCB: it leads to a phenomenon of cumulative fatigue during the cycles in the solder joints (Ref.1). On the other hand, creep generated during dwell time and atmospheric oxydation accelerate the apparition and the propagation of the cracks induced (Ref.2). In specific conditions, such as automotive applications, thermal cycling has become one of the main causes of degradation and failure (Ref.3).
In order to define the regions where cracks have appeared, we have performed thermal cycling tests on electronics assemblies. The samples under test were the following :

components : PQFP 100 A42 leads
33,2 x 33,2 x 3,4 mm
1mm pitch

substrate : SMI 3mm thick

solder material : Sn62Pb36Ag2

thermal cycles :
- type A : lower dwell time : 10 min
  upper dwell time : 60 min
  -40 / +150°C and -40 / +125°C
  transition time : 30s
- type B : lower dwell time : 10 min
  upper dwell time : 10 min
  -40 / +150°C and -40 / +125°C

transition time : 30s

The assembling technology used in this study, SMT, is one of those currently found in automotive electronic circuitry. At the same time, the thermal cycling configuration is representative of typical thermal cycles produced by car utilization.
After ageing, degradations were always observed in the upper side of the solder joint near the rear curved side of the lead. Crack initiation begins to appear between 400 and 500 cycles as a function of the type of the dwell time (Ref.4), and after a while, a damaged area is observed in the solder material near the lead tip (Fig.1).

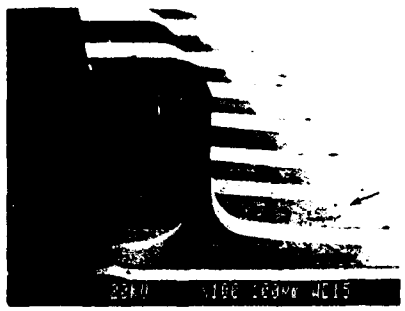


Fig. 1 : damaged solder joint after thermal cycling ageing

All these degradations are first observed near the corners of the packages (Ref.3). This zone does in fact correspond to the most distant point from the centre of the assembly, inducing thus the highest thermal expansion mismatch (Fig.2).

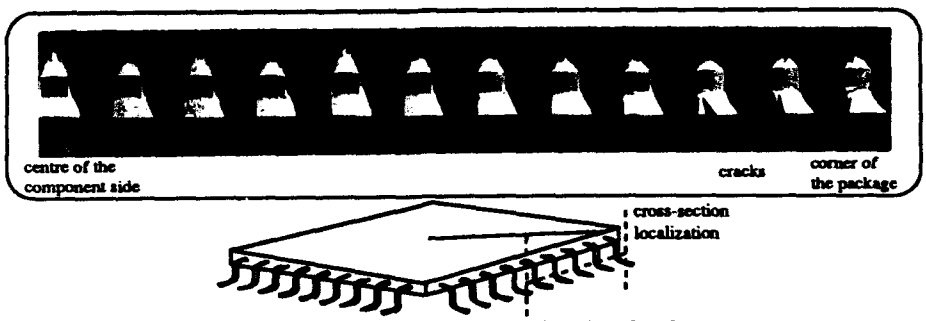


Fig. 2 : cross-section of the leads along a half side of a PQFP package

2. FINITE ELEMENTS SIMULATION

In order to determine the critical areas of the assemblies under test, we performed a finite element simulation with ANSYS software on a PQFP/SMI assembly (Ref.4). The solder behaviour is represented by a stress-strain relation based on a "bilinear model", that includes both elastic and plastic regimes varying with temperature. These simulations were 2D or 3D type and show local stressed areas in the solder joint (Fig.3); simultaneously, maximum stress, strain and plastic work per unit volume were evaluated. The reference temperature was 25°C and the highest one 125°C.

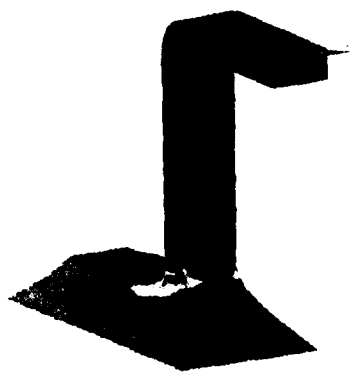


Fig. 3 : 3D gullwing lead simulation

The physical constants values used in this simulation to characterize the materials of the assembly are listed in table 1.

Material	Young modulus (GPa)	Poisson ratio	Yield strength (MPa)	CTE $x10^{-6}/°C$
A42	138	0.30		5.1
PCB	18			13.5
SnPbAg	45		29.1	23
Al	70	0.33		24
Package	10	0.25		25

Table. 1 : Material properties of the assemblies under study

From the obtained simulated stress map (Fig.4), we can observe a good correspondance between the highest stress zones, i.e. near the curved area of the lead in the solder and near the lead tip, and the experimental observation of crack initiation.

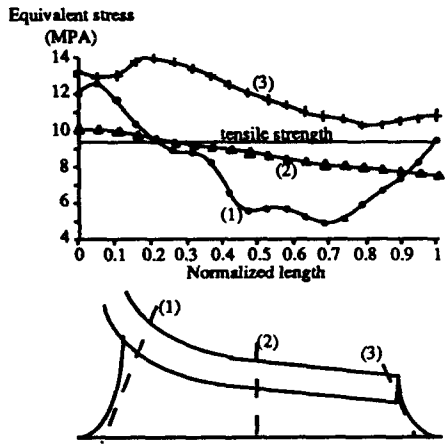


Fig.4 : simulation of stress in the solder joint (PQFP on SMI)

The curves plotted (Fig.4) show the calculated stress along three different cross-sections of the solder, in the regions where cracks are observed after ageing.
It has been shown that in the critical areas of the solder joint, the calculated stress exceeds the tensile strength, leading to plastic strain.

3. MICROSTRUCTURAL ANALYSIS

As the main physical mechanisms, which lead to degradation of the solder joint, have been observed in the coarsened Pb phase regions, the mean size of Pb grains seems to be a good degradation indicator, sensitive enough to exhibit local stressed areas. Indeed, the study of the external aspect of the solder, including the possible cracks is not reliable enough and often too late to have any utility.

Fig. 5 : SEM cross-section view of a PbSnAg SMT solder joint
(a) initial sample ; (b) aged sample after 1500 hours
at constant temperature (150°C)

Fig. 6 : SEM cross-section view of the degradation
along the intergranular interface

We have chosen the study of grain coarsening in order to establish a relation between solder joints life prediction and their microstructural evolution.

During thermal cycling, generation of cracks results principally in cleaving of the solder material along the intergranular interfaces (Fig.6) (Ref.5).

As for the other metal alloys, the mechanical characteristics of the solder material depends strongly on the grain size of the secondary phase: the smaller the mean size, the higher the tensile strength limit (Ref.5). As the main reason for cyclic fatigue is due to the plastic strain per cycle (Coffin-Manson law), high tensile strength (small Pb precipitates) will decrease the level of plastic strain (Ref.7). Coarsening phenomena, highly accelerated by temperature and mechanical stress concentration, is irreversible and stored all the constraints submitted by the solder joint during the past. Before cracks generation, an initial period is observed during which the solder material stores plastic strain energy without microstructural degradation (cyclic fatigue and creep only). After a critical time, cracks appear and the period before total degradation is then very short. We have studied grain coarsening in samples under constant temperature (Fig.5) and under thermal cycling, in order to establish a law relating grain size, time, temperature and stress level.

The experimental analysis has been performed as following: the samples have been cross-sectioned through the solder joints. After polishing, they have been observed with a SEM and the Pb grains population were statistically evaluated by the use of a CRYSTAL analyser. The analysed region was localized near the curved zone of the lead (highly stressed area in cycling conditions) (Fig.8).

3.1 Thermal cycling ageing

In order to exhibit the main parameters inducing Pb grain evolution, we have performed ageing at constant temperature and after that, under thermal cycling. During this second experiment, the assemblies are submitted to both temperature and mechanical stress (induced by thermal expansion mismatch). The cycles applied to the samples were of A and B types.

We have considered that the main phenomenon of coarsening will occur at upper dwell time of the cycles and in the highly stressed areas; thus, the physical analysis of the Pb grain has been localized in the curved region of the lead (Fig.7).

Fig.7 : grain coarsening under thermal cycling

So, by comparing the evolution at constant temperature (150°C), and for equivalent cumulated upper dwell time during thermal cycling (Fig.8), the aggravating influence of the thermomechanical stress, provided by thermal cycles, appears clearly and leads to a higher increase of Δs in the mean grain size for the longest ageing of our test (Ref.3).

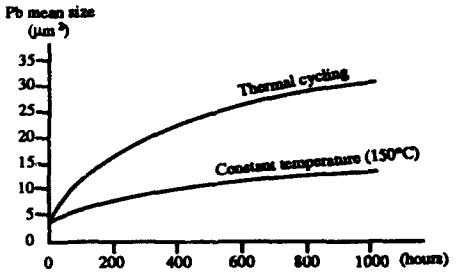


Fig. 8 : constant temperature (150°C) and thermal cycling (-40 / 150°C) grain coarsening evolution with time, showing the mechanical stress effects

4. PHASE EVOLUTION MODELLING

4.1 Time and temperature contribution

The dwell time influence can be easily confirmed by the curve of the figure 9.

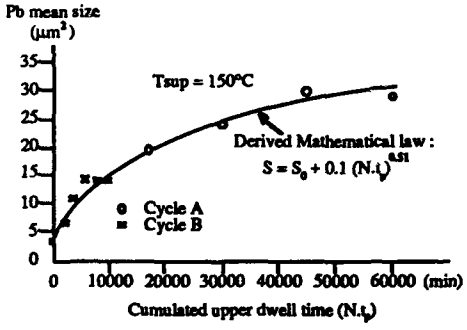


Fig. 9 : Pb grain evolution with cumulated upper temperature dwell time (Tsup = 150°C)

If we report to the results in figure 9, where the grain increase is reported as a function of cumulated upper temperature dwell time ($N.t_p$) for the two types (A and B) of thermal cycles, a continuous curve is obtained, well represented by a time square root evolution as shown by the following equation :

$$s(\mu m^2) = 0.1\left(N.t_p\right)^{0.51} + S_0 \qquad [1]$$

where S0 ($\approx 3.8 \mu m^2$) is the mean lead surface at the beginning of the experiment.
So, as expected, the mean Pb grain size evolution is dependent only on the time spent at the upper temperature and follows a classical diffusion law ($\approx \sqrt{t}$) in the solids.
On the other hand, it is obvious that grain coarsening is also temperature activated. From the two different upper temperatures chosen for the cycles (125 and 150°C), we can deduce an estimation of the activating energy, found to be around 0.45 eV; this value is in agreement with other results (Ref.1). Finally, the temperature and time dependent evolution of the lead domain can be represented by :

$$S = S_0 + A\exp\left(\frac{-E_a}{kT_{sup}}\right)\sqrt{N \times t_p} \qquad [2]$$

where N is the number of cycles, t_p the upper dwell time, $N.t_p$ the cumulated time at high temperature and T_{sup} the highest temperature of the cycle. The numerical evaluation of the parameters gives, Ea = 0.41 eV, A = 8330 $\mu m^2/\sqrt{min}$, with $S_0 = 3.8 \mu m^2$ measured at the beginning of ageing test.

4.2 Thermomechanical stress contribution

It is well known that the fatigue phenomenon appears as soon as the solder material enters plastic strain. To take into account the increase of grain size under cycling condition as shown in figure 8, we have introduced a multiplier factor where the maximal plastic constraint in the solder joint, $\sigma_{(T)}$, is represented with respect to its tensile strength, $\sigma_{L(T)}$:

$$\left(\frac{\sigma_{(T)}}{\sigma_{L(T)}}\right)^{\alpha} \qquad [3]$$

α is a physical constant depending on the sample and determined by identification with experimental curves (Fig.8).
At this step of the analysis, it must be highlighted that the local plastic strain in the solder can be only estimated from the finite elements simulation.
Hence, we have established the following law for grain coarsening in which both the thermal and mechanical effects are considered :

$$S = S_0 + B\exp\left(\frac{-E_a}{kT}\right)\sqrt{N \times t_p}\left(\frac{\sigma_{(T)}}{\sigma_{L(T)}}\right)^{\alpha} \qquad [4]$$

where α (= 5.24) and B ($\approx 18400 \mu m^2/\sqrt{min}$) corresponds to the particular case of our samples configuration and experiment conditions.

5. LIFE PREDICTION

The preceding study has highlighted the strong dependency of grain coarsening on temperature, time of ageing, and maximal stress level: under thermal cycling, the time of ageing is equivalent to the cumulated time spent at the upper temperature. From a general point of view, the mechanical studies of the alloys showing a strong dependency of the ultimate strength value on the microstructural characteristics, it can be considered that, for a typical thermal cycle, a local maximal stress will correspond to a critical phase size for which the probability of cracks initiation is high. The previous study clearly shows a relation between number of cycles and Pb phase size evolution. Thus, a relation between the maximum Pb domains size and the probability of cracks initiation should lead to a life prediction probability as a function of the number of cycles.

Up to now, only experimental observation has been available to determine micro-cracks initiation around the lead domains.

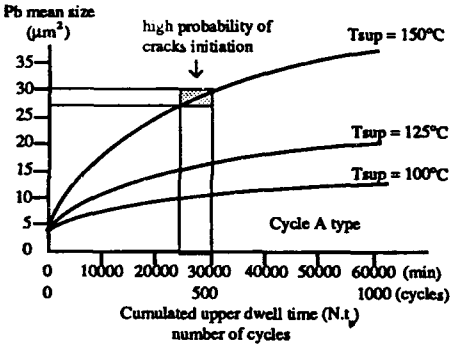


Fig. 10 : Pb grain size evolution for a typical thermal cycling in function of the cumulated dwell time or the number of cycles

As an example, from equation [4] we have represented (Fig.10) the evolution law of Pb domains size as a function of the cumulated upper dwell time, for different maximum temperatures. If we consider the evolution for Tsup=150°C, the physical analysis of the samples under test (upper dwell time=1hour) has shown a cracks initiation for 400/500 cycles; it corresponds to a cumulated dwell time of 24000/30000 min and therefore a critical Pb size about $30\mu m^2$. This value corresponds well with the observed mean size of lead domains at the cracks initiation. The definition of the critical size, as a function of a maximal stress corresponding to a precise upper temperature of a cycle, could give a good prediction of life of an assembly under this cycle.

6. DISCUSSION

It has been evidenced that Pb grain coarsening is strongly dependent on the main parameters influencing the solder joint degradation (temperature, time, cyclic plastic strain). Hence, a simple microsection of an aged sample with a microstructural analysis can easily show the level of cumulated plastic fatigue in the solder joint. The analysis of the spatial distribution of this phenomenon can also provide information on the type of stress undergone by the sample and on the preferential crack initiation zones. As expected, the mean size of Pb grains works as a good indicator of stored plastic strain and thus, of the state of fatigue in the material.

In order to optimize the assemblies with respect to their ability to bear a high number of thermal cycles, many approaches are possible. Firstly, a reduction of the mechanical stress induced in the solder joint can be obtained by an optimization of the geometrical characteristics and of the physical properties of the materials: smaller packages and minimal thermal expansion mismatch between PCB and package (Table.2) will induce lower constraints in the solder joint. On the other hand, reduction of lead stiffness (material, geometry) can strongly decrease the cyclic plastic strain of the solder alloy during the thermal cycle.

Package / PCB	ΔCTE 10^{-6}/°C
plastic / Al_2O_3	19
Al_2O_3 / Al	18
plastic / epoxy	11.5
Al_2O_3 / epoxy	7.5
Plastic / Al	1

Table 2 : Typical assembly mismatch

The Pb mean size increasing during the ageing time, it should be also interesting to obtain the minimal size of the Pb precipitates just after the assembly process. In order to reduce the main size of Pb grains after reflow, we performed different types of processes corresponding to three different cooling times (Fig.11). This operation was performed with components of type DIL 16 Gullwing leads on FR4 board. The different profiles are shown :

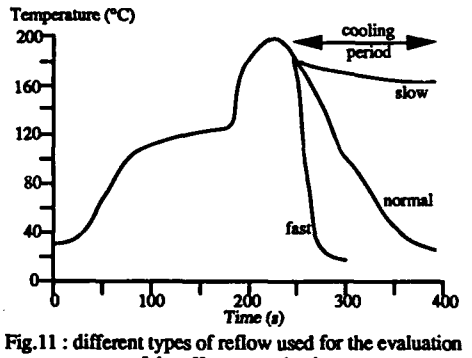


Fig.11 : different types of reflow used for the evaluation of the effect on grain size

After reflow, the different samples were analysed and we obtained the following results for the mean size of the Pb precipitates.

long time cooling : $1.5\mu m^2$

normal cooling : $1\mu m^2$

fast cooling : $0.5\mu m^2$

The results obtained in figure (11) show that a fast cooling period at the end of the reflow process could reduce the Pb grain size and hence increase the life time of the assembly under thermal cycling.

7. CONCLUSION

We have proposed a microstructural evolution law based on lead domain coarsening of a SnPbAg solder submitted to thermal cycling. The Pb domain size is increased under the joined influences of temperature, the number of cycles, through the upper temperature cumulated dwell time, and the stress incursion above the tensile strength. Due to the fact that the maximum strain is deduced from simulation only, the relation between grain coarsening and the number of cycles can only be used as an approximate basis to describe the lead domain behaviour. Life prediction or the determination of number of cycles in operation needs to correlate to the domains size with the cracks initiation. This has only been possible, up to now, by direct observation of solder cross-section. Note that, in any case, a statistical incertitude of the failure time is always provided by the distribution of the lead domain size, either in a given solder joint or from one joint to another.

The following part of the study will deal with the establishment of a law relating grain size and the actual ultimate strength value of the solder material in order to estimate the time before crack generation with a good confidence level.

8. REFERENCES

1. J.H.Lau,"Thermal Stress and Strain in Microelectronics Packaging", Van Nostrand Reinhold

2. J.H.Huang, J.Y.Pei, Y.Y.Qian and Y.H.Jiang, "Life Prediction of SMT Solder Joints under Thermal Cycling", University of Science and Technology Beijing, China, Soldering and Surface Mount Technology N°16, February 1994

3. European Programm EUREKA-PROMETHEUS-PROCHIP report, June 1993

4. A.Hijazi, Y.Danto, Y.Ousten,"Thermomechanical Failure Analysis of SMD Solder Joints", ESREF 1992

5. M.F.Ashby, D.R.H.Jones,"Materials", Pergamon Press, Oxford, 1980

6. R.Tummala, J.Rymaszewski, "Microelectronics Packaging Handbook", Van Nostrand Reinhold

7. S.S.Manson,"Thermal Stress and Low Cycle Fatigue", Mc Graw-Hill, New-York,1966

FOCUS ON FAILURE MODES OF PLASTIC DEVICES AFTER TEMPERATURE AND HUMIDITY ACCELERATED SEQUENTIAL TESTS

N. BARLASSINA, MF. BEDOS, M. BRIZOUX, G. DELEUZE
THOMSON-CSF/SCTF, Department of Technology
l'Orée de Corbeville - BP 56, 91401 ORSAY Cédex France

ABSTRACT

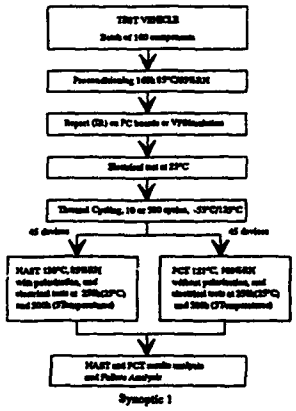
We present in this paper results of long duration temperature and humidity tests made on standard plastic devices.
After a description of the test sequence and the population put in test we present, and analyse the results in term of failure proportions, failure modes...
We summarize in conclusion the results of the analysis about :
- failure modes evolution from 88 to 93
- correlation between HAST and PCT
- set up of a duration and severity for a long term HAST and PCT test.

1. INTRODUCTION

This study is a part of Plastic Components Program supported by DGA/STEI, and others industrials, which aim is to allow the use of plastic encapsulated devices in military equipments. It is a part of a work done to set up a duration for sequential tests, used in a standard project "Specification for qualification of plastic devices in severe environment".

2. PRESENTATION OF THE TEST VEHICLES AND ENVIRONMENT TESTS

Plastic encapsulated products can be very sensitive to ambient humidity, temperature and thermomechanical stresses. So we developped sequential tests in order to simulate the most severe stresses applied on a SMD during its whole lifetime (synoptic 1). The synoptic of the sequential tests is as follow :



Synoptic 1

Different package families PLCC, SO, SOJ, SOT, SSOP, PQFP, TSOP, TO220 from different manufacturers were put in test representing a wide range of functions like memories, logic ICs, ASICs, micros, transistors and so on.
To do that, a large number of components (about 8000) were tested in three periods of time (see fig 1,2,3) :
the first one : date codes 88/89
the second one : date codes 90/91
and the last one : date codes 92/93.
As it can be seen in these figures, the proportion of plastic devices in HAST is more important than in PCT, this disparity comes from the fact that there was not systematically PCT made on devices on date codes 88 and 89. The choice of devices in test and tests applied is directly linked to the interest of THOMSON-CSF equipments applications, thus explaining the large variety of plastic devices families, and unequal quantities put in tests.

3. DEFINITIONS

The classification used in this paper is :

3.1 Package classification

Package families are : PLCC, SO, SOJ, PQFP, TSOP, SSOP and for the following ones, a technology groupment is done : TO220 (and DPAK), SOT23 (and SOT223), SOT89 (and SOT323).

3.2 Function classification

For the same reasons as above, we consider the following distribution : transistor (small signal and power), DAC (and CAD), PAL (and GAL), FPGA (and EPLD), PROM (and EPROM, OTPROM), SRAM, ASIC, microprocessor, linear, logic, Driver, resistance.

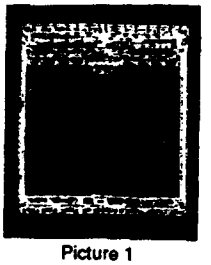
3.3 Electrical failure modes

OC = Open circuit and SC = short circuit.
FD = Functional defect : an important derive of an entry parameter cause the non functionality of the circuit.
PD = Parametric defect : one or more parameter is out of the specification (whenever the circuit is functional).
LC = Leakage current : leakage current out of sanction causing an important bias current.
MC = Memory content : memory source content lost or changed.
NT = Non tested

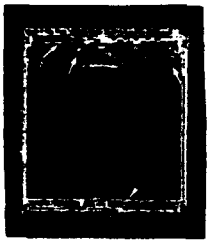
EOS = Failure of the circuit caused by an electrical constraint : flash of metallization, dielectric breakdown.
NC = defect non confirmed
UK = Unknown

3.4 Physical failure modes

CO = Corrosion of metallizations on aluminium lines or pads by electrochemical way (HAST) or galvanic way (HAST and PCT).
IM = Intermetallic : degradation of Au-Al interface caused by temperature or humidity or both. The Au-Al phase is modified and electrical and mechanical characteristics are degradated.
H_2O = Moisture penetration H_2O : this physical defect is defined by an absence of corrosion, intermetallic degradation, or other visible failure and an electrical failure that sometimes disappears after at 125°C/24hrs storage. This mechanism is caused by ion contamination though resin or resin leadframe interface (see picture 1 and 2).
FE = Front-End : Weakness passivation on front-end process, allowing moisture penetration (see picture 3).
RE = Report : failure of the device report on PCB.
BO = Bonding : failure of the assembly process, in front-end process (see picture 4).
OX = Oxide : pinholes defects in oxide.
PC = Pop Corn : cracks in the resin after report phase (see picture 5)
NA = Non analyzed.
NC = Unconfirmed failure.
UK = Unknown.



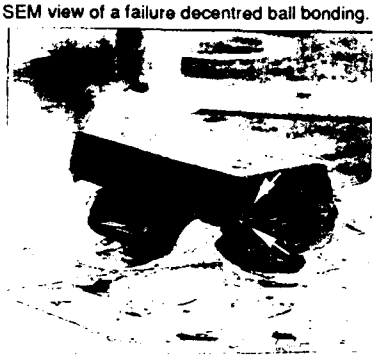
Picture 1
Emission microscopy image (address x = 0, y = 0) of a reference device.



Picture 2
Emission microscopy image (address x = 0, y = 0) of a failure device.

Picture 3
SEM view of cracks in the passivation of a TO 220 plastic device.

Picture 4
SEM view of a failure decentred ball bonding.



Picture 5
SEM view of cracks in the resin after report phase on a SOT 23 : POP CORN effect.

4. ANALYSIS OF RESULTS

4.1 Failures proportions in HAST and PCT

We consider first the failure proportions in HAST and PCT (fig 4 and 5). For HAST, the total failure percentage is stable and possibly decreases with year of manufacturing (drop from 12% in 90-91 to 6% in 92-93).
In PCT, we observe an increase of failure proportion from 15% in 88-89 to 26% in 92-93 that is a reverse evolution comparing to HAST and the common opinion that reliability of components in accelerated test increases with year of manufacturing. We propose later an explanation.

4.2 Packages sensitivity

For packages, TSOP and SOJ have the higher failure proportion in HAST ; mainly because they are dedicated to specific functions (PROM and SRAM) sensitive to HAST test (cf fig 9 and § 4.3) and effects of charge trapping. But the general proportion of defects is quite low and in term of distribution among all package families (fig 7), there is no evident trend. Sensitive

packages in PCT are SO (logic families) and TO220 (fig 8 and 10). Effect of delaminations or corrosion in these particular cases is predominant (see the picture 6 below as example).

Picture 6

SEM view of a pad cross section showing resin/die delamination on a SO plastic device.

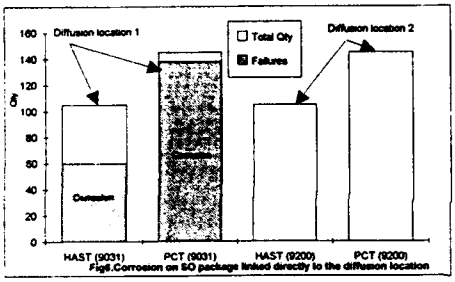


Fig6.Corrosion on SO package linked directly to the diffusion location

On the other hand, we can also see the influence of the diffusion location for example on SO packages (see fig 6). Concerning the high proportion of failures for recent manufacturing, we can explain it with the failure rate of TO220 devices.
Indeed, a noticeable part of date codes 92-93 is formed by TO220 on this file test, that is particularly sensitive in fig. 5 bis, we have the failure proportion of non TO220 package : that is stabilized around 15%.
Thus we conclude that behaviour in PCT is related with package integrity. PCT is a package sensitive test. Therefore we see that PCT has a selective severity on some functions or packages. It is also important to notice that there is non direct relation between package complexity or size and resistance to HAST and PCT test.

4.3 Functions sensitivity

In term of functions, the most critical ones in PCT are FPGA, logic transistor and linear (in accordance with sensitivity of TO220 and SO packages (see fig 12 and 14) for the three last families. The presence of FPGA is due to the fact that test had been made on few products, including a sensitive one, that had an influence on statistics.
In HAST, we find four families : logic, PROM, microprocessor and SRAM. The analysis of failure modes gives an explanation to this fact.

4.4 Failure modes

Different behaviours are observed in HAST and PCT : great variety of modes (electrical and physical) in HAST (cf fig 15 and 16), with no preponderant mode. The most important physical failure mode in HAST is moisture penetration, not corrosion. When we consider the function associated with this failure mode, three of these are sensitive to moisture penetration on passivation layers (es'pecially EPROM) (fig 11 and 13). HAST is a "function sensitive" test.
In PCT, open circuits related with corrosion represent around 70% of failures (cf fig 17 and 18). We have here a complementarity of HAST and PCT in term of failure modes : when we consider the sum of failure modes observed in HAST and PCT, we have the three main modes related with moisture effect on packaging : corrosion, moisture contamination, intermetallics degradation.

4.5 Evolution of failure modes with date-codes

In HAST, we have a continuous improvement in term of resistance to corrosion (fig 19). Corrosion does not occur anymore on 92-93 date codes. Effect of H_2O penetration failures increases, in relation with the increasing complexity (memories, FPGA, EPLD...).
In PCT, there is no major evolution yet. The corrosion failure slightly regressing. Intermetallics degradation is quite now equal to zero (fig 20).
Moreover failures probably linked to new functions like EPROM appear (H_2O penetration).

4.6 Set up of a qualification target

Depending of the criticity of the device, and of the hardness of environment, the requirement for plastic devices in military equipment has to be a failure rate in the range of 10 to 200 $10^{-9}/\lambda$, and a lifetime with no wearout in the range of 10 to 20 years. As the experience gained by field return analysis needs still a few years to be significant. We rely on long term accelerated test to demonstrate (theorically) these figures. Thus, we fixed as a primary target, by calculations based on theorical acceleration laws :
- 500 hrs sequential HAST 130°/85%, 1 failure max. and
- 500 hrs sequential PCT 121°/100%, 1 failure max.
Although these severity are completely beyond the limits of traditional "reliability" tests mode by manufacturer, the majority of components succeed to reach it.
However after analysis of failure modes, we think that it is difficult to go beyond this limit for back-end qualification es'pecially because of wearout apparition

on some packages (TO220 in PCT and maybe some functions (PROM in HAST).

5. CONCLUSION

5.1 Failure modes evolution

The results confirm the improvement of plastic devices resistance to corrosion including intermetallics, at the exception of power packages. But we notice also an increasing of sensitivity to moisture penetration in passivation layers and oxides, es'pecially for PROM in HAST test.
Thus the global proportion of failures of the whole population of products in HAST and PCT at 500 hrs is a constant, from 1988 to 1993 (around 15%).

5.2 HAST and PCT correlation

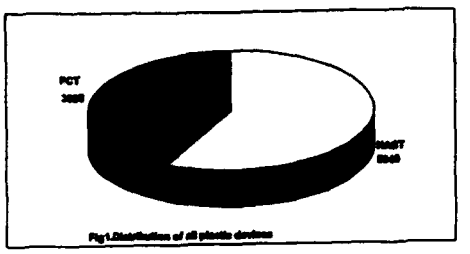
Concerning the sequential tests, no correlation between HAST and PCT results appear. In fact, they are complementary each of them adressing different failure modes. Mixing the failure modes observed in HAST and PCT gives all the modes that can be potentially found in real moist environment.

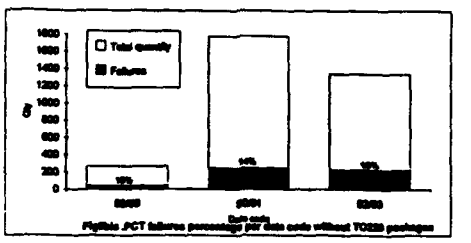
5.3 Test duration and severity

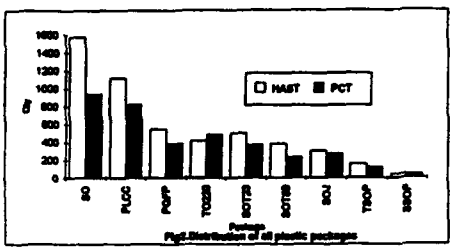
The results confirm the realism of duration and severity set up in the standard project to qualify plastic ICs in rugged environment see synoptic 1.However an adaptation has to be made for power packages (incl. TO220 and DPAK), because of their sensitivity to long term PCT.

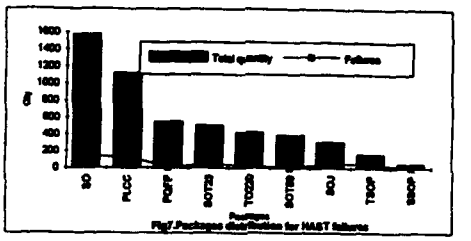
ACKNOWLEDGMENT

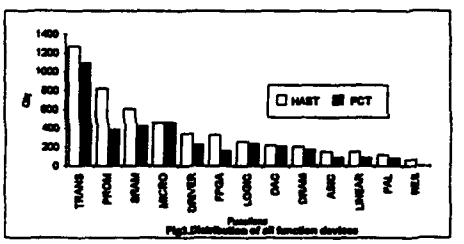
The authors wish to thank particularly R. EVEN from the DGA/STEI for his support in our general Plastic Components Program.
We also thank our colleagues, who contributed to this work, M. REFFET, M. MANSART for typing and arranging the final form.

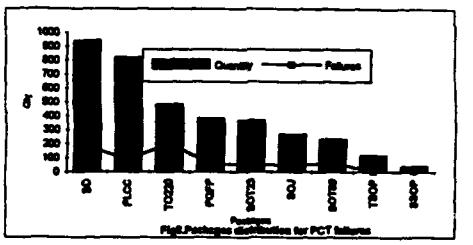

PCT
HAST

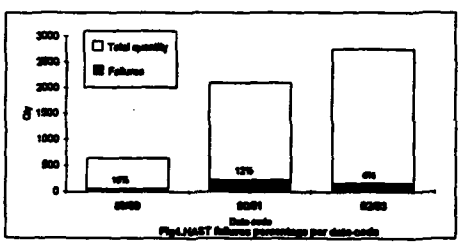

Total quantity
Failures
Qty

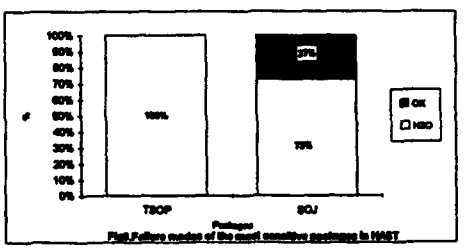

HAST
PCT
Qty
SO
PLCC
PDFP
TO220
SOT23
SOT89
SOJ
TSOP
SSOP

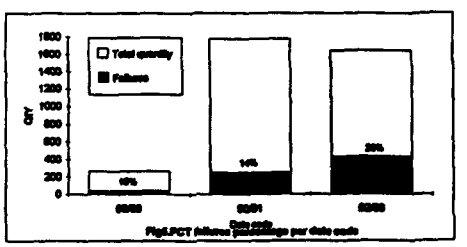

Total quantity
Failures
Qty
SO
PLCC
PDFP
SOT23
TO220
SOT89
SOJ
TSOP
SSOP

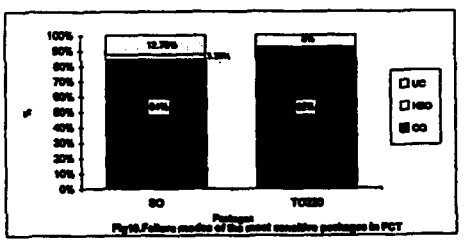

HAST
PCT
Qty
TRANS
PROM
SRAM
MICRO
DRIVER
FPGA
LOGIC
DAC
DRAM
ASIC
LINEAR
PAL
RES


Quantity
Failures
Qty
SO
PLCC
TO220
PDFP
SOT23
SOJ
SOT89
TSOP
SSOP


Total quantity
Failures
Qty


OK
HBO
TSOP
SOJ


Total quantity
Failures
Qty


UC
HBO
CO
SO
TO220

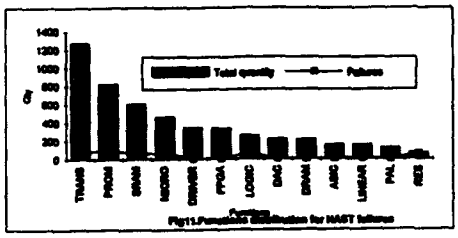


Fig11.Function distribution for HAST failures

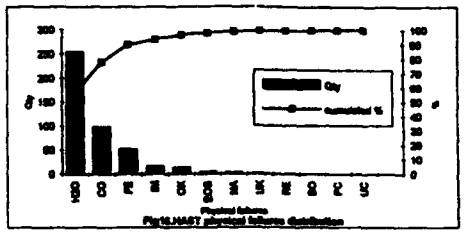


Fig16.HAST physical failures distribution

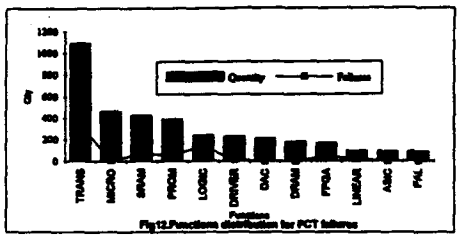


Fig12.Functions distribution for PCT failures

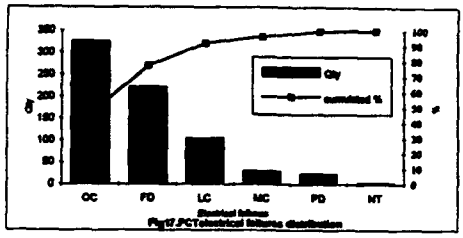


Fig17.PCT electrical failures distribution

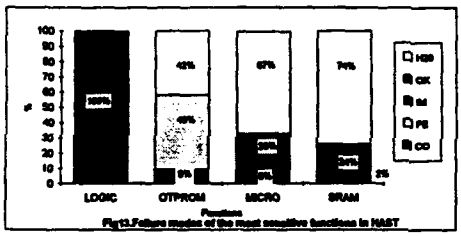


Fig13.Failure modes of the most sensitive functions in HAST

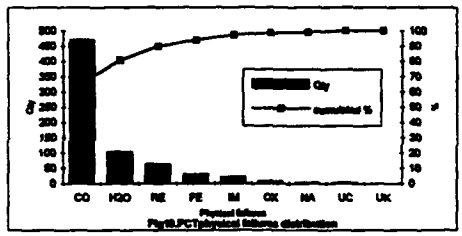


Fig18.PCT physical failures distribution

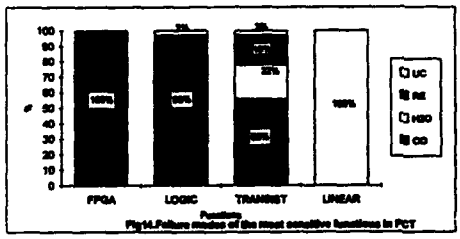


Fig14.Failure modes of the most sensitive functions in PCT

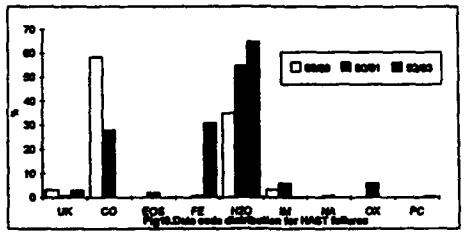


Fig19.Data code distribution for HAST failures

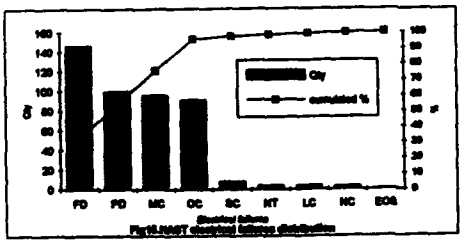


Fig15.HAST electrical failures distribution

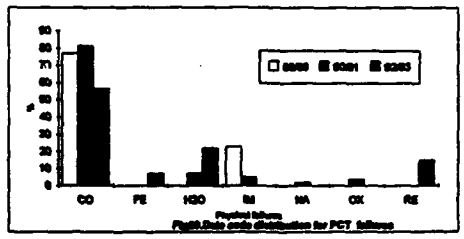


Fig20.Data code distribution for PCT failures

THE DEMISE OF PLASTIC ENCAPSULATED MICROCIRCUIT MYTHS

E. B. Hakim
Army Research Laboratory
AMSRL-EP-RA
Fort Monmouth, NJ 07703-5601

R. K. Agarwal
Corporate Technology Center
Delco Electronics
Kokomo, IN 46904-9005

M. Pecht
Calce Electronic Packaging Research Center
University of Maryland
College Park, MD 20742

SUMMARY

Production of microelectronic devices encapsulated in solid, molded plastic packages has rapidly increased since the early 1980s. Today, millions of plastic-encapsulated devices are produced daily. On the other hand, only a few million hermetic (cavity) packages (Figure 1) are produced per year. Reasons for the increased use of plastic-encapsulated packages include cost, availability, size, weight, quality, and reliability. Markets taking advantage of this technology range from computers and telecommunications to automotive uses. Yet, several industries, the military in particular, will not accept such devices. One reason for this reluctance to use the best available commercial parts is a perceived risk of poor reliability, derived from antiquated military specifications, standards, and handbooks; other common justifications cite differing environments; inadequate screens; inadequate test data, and required government audits of suppliers' processes.

This paper describes failure mechanisms associated with plastic encapsulation and their elimination. It provides data indicating the relative reliability of cavity and solid-encapsulated packaging, and presents possible approaches to assuring quality and reliability in procuring and applying this successful commercial technology.

1. INTRODUCTION

Plastic-encapsulated microcircuits (PEMs) have been used primarily in commercial, industrial, automotive, and telecommunications electronics. Consequently, they have a large manufacturing base. With their major advantages in cost, size, weight; performance; and near-instant availability, plastic packages have attracted widespread attention for government and military applications. Although this is a major opportunity for PEMs, they have met formidable challenges in adapting to the high-reliability demands of these markets. While the major impediment to PEM application has been a perception of lower reliability, problems also arise as a result of the military's small procurement and production volumes, the predominance of manual package-assembly operations used by military suppliers, and the defense department's outdated standards and handbooks.

Some of the first semiconductor devices were encapsulated in plastic. These early devices used molding compounds plagued by thermal intermittence problems.[1] Because of the difference in the coefficients of thermal expansion (CTE) of the bond wires and the encapsulant, these devices exhibited open-circuit failures at the bond pads at temperatures above ~100°C. As temperatures decreased, compressive forces restored the contact of wire to bond pad. Moisture-induced failures, like corrosion, cracking, fracture and interfacial delamination, were also significant.[2] This problem has been largely resolved; testing at 85°C/85% relative humidity in 1974 produced 25% cumulative failures at 1,000 hours, compared with 0.1 in 1990[3]. The nearly exclusive use of hermetically sealed microcircuits in military, aerospace, and other high-reliability, high-criticality applications is a direct result of the problems associated with early plastic packaging.

The decade of the 1980s brought revolutionary changes in electronics technology in general, and in plastic packaging in particular. Earlier plastic- encapsulated transistors and diodes were fabricated by dispensing a small amount of room-temperature vulcanizing silicone or flexible epoxy material over the die and bond wires (glob-topping). Later, various molding techniques were attempted, including transfer, injection, and potting. Hundreds of variations in epoxy, silicone, and phenolic materials were evaluated for cost, performance, implementation, shelf life, repeatability, flammability, and reliability. Also evaluated were various additives for heat removal, adhesion, viscosity, mold release, flame retardation, and appearance. Protecting the die surface prior to molding by coating it with silicone elastomers, varnish, or spun-on glass (SOG) was a popular procedure. To reduce voiding between encapsulant and package leads, silicone resin was forced into these voids under a vacuum, a process known as "back filling."

The progressive improvement in plastic packaging integrity has been effected by improved materials, increased plastic purity, high-quality device passivation, improved lead frame designs, and manufacturers' quality programs. In general, the failure rate of plastic packages has decreased from about 100 failures per million device hours in 1978 to about 0.05 per million device hours in 1990[4]. Hermetic cavity packaging does not appear to have kept up with these advanced requirements in either performance or cost, as is obvious from the curves in Figure 2; worldwide sales of commercial microcircuits in 1995 are projected to be $100B. Military sales projections are down to $1.6B, a decrease to 1.5% of the total market share from a high of 16% in 1975.

2. ADVANTAGES OF PLASTIC PACKAGING

2.1 Performance

A plastic package has advantages of light weight and small size, compared with its ceramic counterpart; commercial plastic packages generally weigh about half as much as ceramic packages. A 14-lead plastic dual in-line package (DIP), for example, weighs about one gram, versus two grams for a 14-lead ceramic DIP. Although there is little difference in size between plastic and ceramic DIPs, smaller configurations, such as small-outline packages (SOPs) are only available in plastic. These small packages result in higher packing density and shorter propagation delays. At the printed circuit board level, the use of SOPs allows smaller, fewer, and higher-performing circuit boards. Figure 3 illustrates the impact plastic encapsulation is having on microcircuit assembly and packaging. Today, 72% of ICs are produced by surface-mount technology (SMT)—non-cavity assemblies. With the introduction of ball-grid arrays (BGAs), the percentage of SMT packages is projected to increase significantly.

2.2 Cost

The cost of a packaged electronic part is determined by several major factors: die, package, volume, size, assembly and assembly yield, screening, pre-burn-in and its yield, burn-in, final test and its yield, and the specified qualification test. Because more than 90% of the IC market is plastic-packaged, cost has been lowered by

automated volume manufacturing and low off-shore labor expenses. Hermetic packages are usually fabricated using more expensive materials and labor-intensive manual manufacturing processes (JAN military requirements in the past specified on-shore manufacturing). Moreover, there is little cost difference between plastic surface-mount components and plastic DIPs, whereas ceramic surface-mount components are more expensive than ceramic DIPs. Thomson - CSF reports a 45% purchase cost reduction for each of twelve printed circuit boards (PCBs) in a manpack transceiver application implemented with PEMs rather than ceramic components [5].

It may be argued that hermetically packaged ICs may cost up to ten times more than plastic-packaged ICs because of the rigorous testing and screening required by the user for hermetic parts[6]. However, ELDEC[7] estimated that plastic ICs cost 12% less than their hermetic counterparts when both types were screened to customer requirements.

High yields and low assembly costs are achieved with plastic-packaged parts because they lend themselves well to automatic assembly techniques, thereby eliminating manual handling and operator error. On the other hand, automated pick-and-place machines reportedly can crack hermetic seals or chip the package. Moreover, costs of PEMs become lower with a higher level of integration and higher pin-count devices, because of the high price of the die in relation to the total cost of the packaged device. While these benefits may not be realized for complex monolithic VLSIs, great cost advantages may accrue for complex package styles, such as multichip modules.

The price per part to the user will include the above costs, and a significant price adder for the military. This adder, a built-in fact for the military market, includes: the cost of the supplier's military infrastructure; with the absence of competition, price will be what the market will bear; and a need for a favorable profit margin.

2.3 Availability

Plastic-encapsulated microcircuit part types are much more available than hermetic devices. First, because non-cavity plastic devices are assembled and packaged on continuous production lines, as opposed to the on-demand production of hermetic parts, acquisition lead times are significantly shorter and problems associated with the restart of a hermetic line are not encountered. Second, some parts are simply not available from major manufacturers in cavity form. Most designs are developed first as plastic-encapsulated microcircuits. Suppliers estimate that, at any given time, 30% more part functions are available in plastic than in ceramic. Hermetic packages are developed only in response to sufficient market interest, performance requirements, and cost benefits: the military, the major purchaser of hermetic parts, has become a small portion of the total electronics market. With the current technology transition toward SMTs, interest in ceramic devices has lagged in the market, making adaptation of plastic ICs to military applications more critical.

2.4 Reliability

The reliability gap between cavity-packaged devices and solid PEMs has decreased in the last decade. Figure 4 summarizes published improvements in plastic encapsulated microcircuit reliability since 1976[7]. Two major contributions to this trend are encapsulating materials and passivation. Modern encapsulating materials have low ionic impurities, good adhesion to other packaging materials, a high glass transition temperature, high thermal conductivity, and CTEs matched to both die and leadframe. Advances in passivation include better adhesion to the die, fewer pinholes or cracks, low ionic impurity, low moisture absorption, CTEs better matched to substrates, and the use of such techniques as spun-on glass.

Figure 5 presents comparative failure-rate data between 1978 and 1990 of plastic-encapsulated microelectronics and hermetically packaged devices from a commercial source[8]. The database for this figure is from first-year warranty information on commercial equipment operating primarily in ground-based applications (office, laboratory, and transportable equipment); these failure rates are for the same part (or part function) over time. As Figure 5 shows, during this period both types of packaged devices improved more than an order of magnitude in early-life failure rate. For PEMs, the current value for this type of reliability is ~ $0.02/10^6$ hours or 20 FIT. However, it should be noted that the use environment is not precisely known for either type of device and data cannot be isolated to the package without knowing the changes in die reliability during the period.

To compare common device types, Condra, et. al[9] tested the same mature custom bipolar IC in both plastic (commercial part) and hermetic ceramic DIP (military part) versions on twelve circuit-card assemblies. They ran 1,000 temperature cycles, from -55^0 to $+85^0$C, to compare the functional reliability of the two types of packages. No differences were observed in any of the twenty-six measured parametric values. They then added these parts to an untested group of about a hundred of the same devices, half plastic and half hermetic, in another set of circuit-card assemblies, along with an older discrete version of the card as a control. All these were subjected to 1,000 hrs. of 85^0C/85% relative humidity condition with 28 volts of intermittent bias (30m on, 30m off). The previously thermally cycled parts (both ceramic and plastic) could only be tested up to 650 hrs. before failing. Among the new group, no failures of either type of component were observed. Conservative lifetime estimates for both package types in avionic applications were well over thirteen years, even for combined severe testing.

The big question is why DoD isn't taking advantage of this superior technology and reaping the same benefits as the rest of the industry. The answer lies in history and tradition. In the 1960s, reliability and quality issues plagued the new electronic device called the integrated circuit. Government documents were subsequently generated to regulate ICs because military applications were the driver for these products. These documents included:

- 1962: MIL-HDBK-217, Military Handbook on Reliability Prediction of Electronic Equipment;
- 1963: MIL-STD-454 (Requirement 64-Microelectronic Devices), General Requirements for Electronic Equipment;
- 1963: MIL-STD-781, Reliability Testing for Engineering Development, Qualification and Production;
- 1965: MIL-STD-785, Reliability Program for Systems and Equipment Development and Production;
- 1968: MIL-STD-883, Test Methods and Procedures for Microelectronics;
- 1969: MIL-M-38510, General Specification for Microcircuits;

A decade later, the world market for military-approved microcircuits was well below the industrial/commercial demand for ICs. Current projections indicate the market for military and industrial hermetic devices will be only about 1% of the world production of microcircuits by 1995.

As did the integrated circuit, low-cost encapsulation processing required a period of learning and experimentation. By the early 1980s, the failure mechanisms that caused quality and reliability concerns had been researched and essentially reduced to background noise. Aluminum interconnect metallization corrosion was controlled by improvements in passivation composition and reduced defect density. Ionic contaminants on the die and in the encapsulant had been reduced to insignificant levels, eliminating bond-pad corrosion in field-use environments. Material and structural designs controlled thermal mismatch issues.

With technology advances, new failure mechanisms can evolve, but in high volume lines they are immediately addressed and controlled. This was the case with surface-mount technology delamination or "popcorning" resulting during assembly soldering from the vaporization of absorbed moisture in very thin packages containing very large dies.

Reliability data from many sources, including Texas Instruments, the ITT Research Institute, Honeywell, Rockwell International, Hamilton Standard, and Litton, indicate that PEMs are equally or more reliable than hermetic parts. This is not surprising, considering the microelectronics market trends depicted in Figure 2. Device manufacturers are improving their competitiveness in world markets by automating and modernizing their industrial and

commercial facilities at the expense of their military assembly and packaging facilities.

The DoD isn't taking advantage of this high-quality market because it is held captive by the military specifications generated three decades ago. MIL-HDBK-217, for example, is still used to predict the reliability of a system, even though it is thirty years old. These predictions have been biased so that only military-approved parts will meet specified reliability goals. Other factors—such as MIL-HDBK-217's dependence on steady-state temperature, which incurs system penalties of size, weight and cost; ignorance of design impact; and current field-return failures, which have no relevance to the 217 model—make this document a deterrent to sound scientific judgment. Program managers for system developments, which typically require use of the 217 model, fear a substantial risk if they use PEMs. Other military standards also support this perceived risk and bias program managers in their choice of parts. MIL-STD-454, Requirement 64, includes the order of precedence by which an equipment developer must select microcircuits; only a military-approved part is permitted.

MIL-HDBK-217 is based on the assumption that part failures are the cause of equipment failure. What it does not take into account is that, since the early 1960s, parts have become extremely reliable and are not generally the cause of equipment failure. Field-failure returns from OEM repair facilities, microcircuit suppliers, and DoD depots universally indicate that less than 5% of all failures are chip- or package-related. The vast majority of returns are retested and pass, while the remaining 30 to 40% fail because of printed circuit-board assembly or design-related reasons.

3. WHAT THE ARMY IS DOING

Since the late 1960s, the Army has been using plastic-encapsulated devices (transistors, diodes, microcircuits) in systems for which program managers realized the advantages of the technology. In one case study, an Army Panama Canal Zone field study was initiated in 1970 to assure that helicopter radios using these devices would not have reliability problems. Ten years later, with 5,000 transistors and integrated circuits on test and a quarter-billion device hours accumulated, the verdict supported the use of PEMs[10].

The Army has since used this technology in a few select development systems. These include the Platoon Early Warning System—where one million PEMs were used—and numerous electronic fuze applications, the most recent being the M762, which is now in production using surface-mount technology.

The Army is taking advantage of cost savings resulting from the procurement of non-developmental items (NDI). This equipment is procured without the restrictions imposed by military documents. Most microcircuits for this NDI equipment are PEMs purchased to assure that the system meets performance and reliability requirements. The DoD, in most cases, does not specify or even know what devices are used or how they are procured. The environmental uses for these NDI systems range from temperature/humidity controlled enclosures to hand-held, uncontrolled, anywhere-in-the-world applications.

Army document HDBK-179(ER) - Microcircuit Application Handbook dated October 1993, lists among other things the equipment environments in which PEMs can be used and how an OEM can assure the government that selected suppliers and microcircuits will surpass system quality and reliability requirements.

The newest DoD qualification system, MIL-I-38535, the Qualified Manufacturer List (QML), embraces many of the ideas of best commercial practices (BCP). Commercial suppliers of high-volume PEMs produced for users who demand quality, reliability, and low cost have incorporated a methodology which assures these goals: control of incoming materials, in-line process control, statistical process control (SPC), in-line process monitors, continual periodic testing, and so forth.

Today, the six certified QML manufacturers produce wafers on lines used by both their military and industry customers. In addition, the DoD now permits off-shore QML lines after vendors argued that maintaining on-shore production was financially untenable. However, QML still demands that the IC supplier support a military organization in addition to its commercial organization, which is a non-value added proposition.

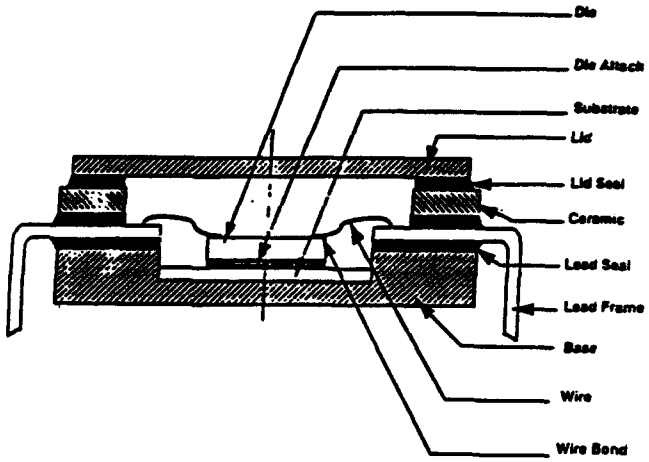
While liberalizing QML is a move in the right direction, maintaining separate military lines is no longer justifiable, even to satisfy the environmental requirements of the military customer. Packaging and assembly for non-military industrial markets is done on high-volume plastic-encapsulation lines that are more efficient and cost-effective than their military counterparts. These lines are high-yield, high-quality, have a low defect rate, and provide a highly reliable product to demanding customers.

4. SUMMARY

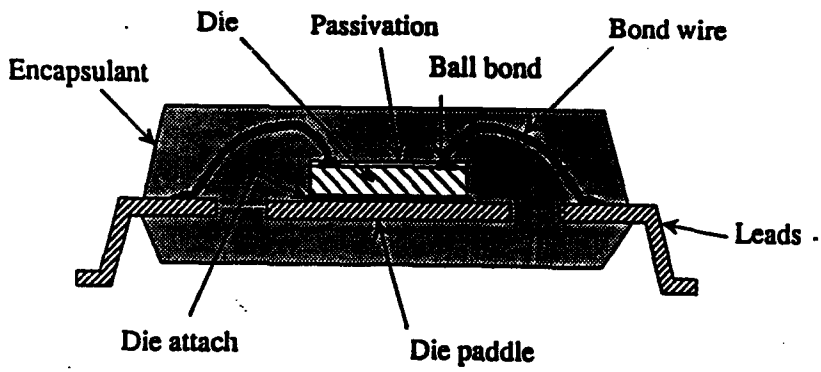
The quality and reliability of high-volume, best-commercial-practice parts are no longer an issue. Data is available showing that this technology is equivalent to traditional hermetic cavity packages. The challenge is how to procure them to meet equipment requirements cost-effectively. The Army has developed HDBK-179(ER), Microcircuits Application Handbook and has applied this methodology to several key programs, including the Comanche helicopter, the Battlefield Combat Identification System (BCIS), and the Single Channel Ground & Airborne Radio System (SINCGARS). The projected savings secured for each program by the use of PEMs are substantial.

References

1. Motorola ICD, Reliability Engineering Report, "Plastic Package Reliability Report, Bipolar Integrated Circuits 1974," I/C Reliability Engineering Dept., Motorola.
2. J. T. Flood, "Reliability Aspects of Plastic Encapsulated Integrated Circuits," Proceeding of the International Reliability Physics Symposium, IEEE, 1972, pp. 95-99.
3. B. E. Kurtz, "Implications of Non-Hermetic Chip Packaging for Military Electronics," 1990 International Packaging Conference, September 1990, pp. 848-57.
4. G. F. Watson, "Plastic-Packaged ICs in Military Equipment," IEEE Spectrum, February 1991, pp. 46-48.
5. M. Brizoux, et al, "Plastic ICs for Military Equipment Cost Reduction Challenge and Feasibility Demonstration," 40th Electronic Components and Technology Conference, 1990, pp. 918-24.
6. Reliability Analysis Center, "IC Quality Grades: Impact on System Reliability and Life Cycle Costs," Report # SOAR-3, Rome, NY, 1985.
7. L. W. Condra, and M. Pecht, "Options for Commercial Microcircuits in Avionic Products," Defense Electronics, July 1991, pp. 43-46.
8. W.K. Denson, "Ask the RAC," RAC Journal, 1st Quarter 1994, Vol 2, No. 1, pp. 15.
9. L. W. Condra, S. O'Rear, T. Freedman, L. Fiancia, M. Pecht, and D. Barker, "Comparison of Plastic and Hermetic Microcircuits under Temperature-Humidity-Bias," IEEE Transactions on Components, Hybrids, and Manufacturing Technology, 15(5), 1992, pp. 640-650.
10. E. B. Hakim, "US Army Panama Field Test of Plastic Encapsulated Devices," Microelectronic Reliability, Vol 17, 1978, pp. 387-392.



(a) cross-section of a typical ceramic single-chip package



(b) plastic package construction

figure 1. comparison of cavity vs. non cavity (pem) package construction

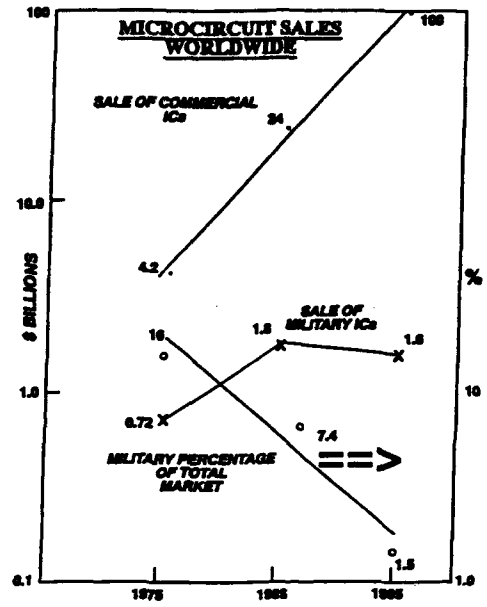


figure 2. market sales of commercial vs. military microcircuits (data provided by texas instruments)

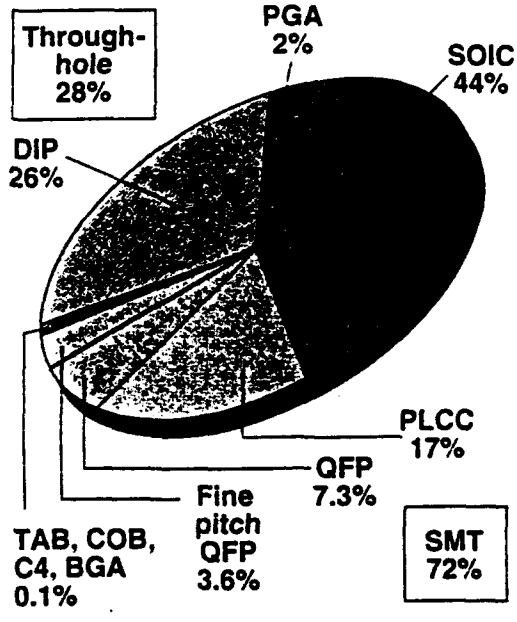


figure 3. package mix for integrated circuits (source: "design benchmarks and activity metrics," ceeris international, old lyme, connecticut)

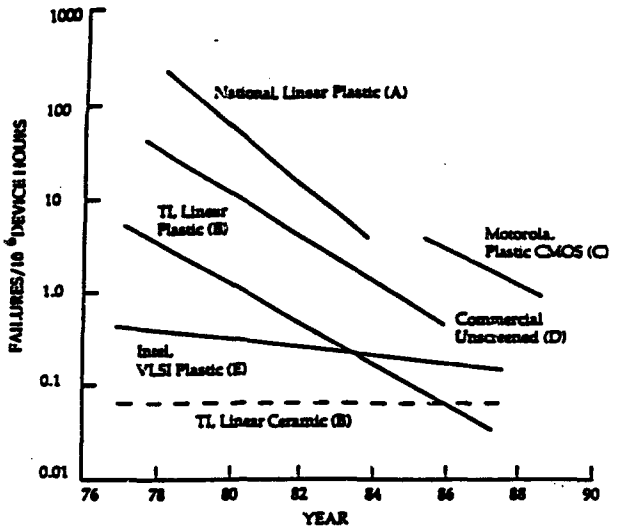


figure 4. microcircuit reliability improvement trends (condia, l. and pecht, m., "options for commercial microcircuits in avionics products," defense electronics, july 1991)

(reference letter is in parenthesis)

references:

(a) pantie, d. m., "maturity factors in predicting failure rate for linear integrated circuits," ieee transactions on reliability, vol. r-33, no. 3., /august 1984, pp. 208-211.

(b) texas instruments, ceramic/plastic comparison report to eldec, october 24, 1988.

(c) fergus, k., "a discussion on the effectiveness of better level i (temperature cycling) and level ii (burn-in) screening," motorola reliability report ric-6086, january 1990.

(d) frank, h., "reliability and mil-hdbk-217 prediction," reliability review, vol. 10, september 1990, pp. 7-9.

(e) crook, d. l., "evolution of vlsi reliability engineering," proceedings of the 28th annual ieee reliability physics symposium, 1990, pp. 2-11

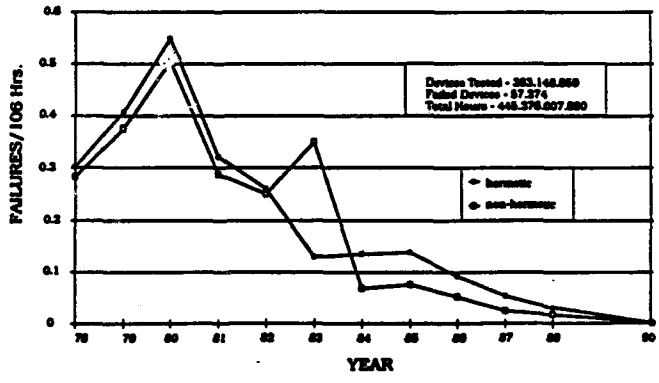


figure 5. ic failure rate as a function of year

FRACTURE MECHANICAL CHARACTERIZATION OF IC-DEVICE-INTERFACES

K. Burger
TEMIC TELEFUNKEN microelectronic
Theresienstr. 2
D 74072 Heilbronn

1 INTRODUCTION

The reliability of plastic packaged IC-devices, concerning solder-dips, thermal cycling or humidity tests is predominantly governed by the adhesion between the different components and their thermomechanical properties.
A "first law of plastic packaging" was given by Alpern et. al. /1/: "Perfect adhesion at any interface minimizes failure", taking account the results of failure analysis /2/ and finite element calculations /3/.
In plastic encapsulated integrated circuits three main interfaces can be distinguished. First, the interface between the die and the plastic (A in Fig. 1), second, the die-attach to the die-pad (B in Fig. 1) and third the phase-boundary between the plastic package and the lead frame (C in Fig. 1).
To characterize the bond-strength or adhesion of such interfaces, several destructive and non-destructive methods are known.
The scanning acoustic microscope (SAM) is commonly used to detect delaminations /4, 5/ along the interfaces. But, no information about the bond-strength of interfaces is available from SAM.

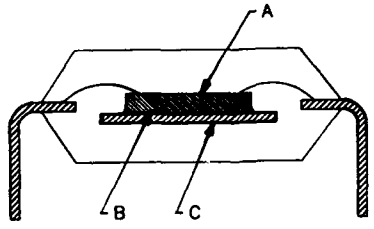


Fig. 1: Schematic cross-section of a plastic-packaged IC-device

Peel-, pull- and shear-tests are destructive methods to measure the adhesion between two materials /6/. The pull- and peel-tests are mainly used to measure the adhesion of the epoxy-resin on the lead frame or the Si-chip. Special test-specimen and -geometries were necessary and therefore, the thermomechanical stresses along such interfaces were not comparable to those existing in the real IC-devices.
The shear-test is usable to characterize the bond strength of the die-attach in the unmoulded state.

In the following a new fracture mechanical test method will be presented, which allows the measurement of the bond-strength or adhesion of each

of the three interfaces present in plastic encapsulated devices. The change in adhesion along this interfaces caused by reliability tests (thermal cycle test, humidity tests or solder dips) will be measured and correlated to the failure behaviour of the devices.

2 THE FRACTURE-MECHANICAL PRINCIPLE

To investigate the physical nature of the bond strength along interfaces between heterogen materials many theoretical /7, 8, 9/ and experimental work /10 - 13/ has been done.
Several different probe geometries and fracture-test philosophies have been developed to measure the strength of adhesion along heterogen phase-boundaries.
The four-point-bend-test-method /11/ will now be used to characterize the bond-strength of the interfaces, present in plastic-encapsulated ICs (Fig. 1).

2.1 The preparation technique

Fig. 2 shows the preparation steps necessary to get a four-point-bend-test-sample from a plastic packaged device.
The first preparation-step is cutting and polishing the IC-device from each of the four sides until the corresponding belonging chip border is reached. The specimen, resulting from this procedure has the size of about the chip-size and a thickness of the original device (Fig. 2b).

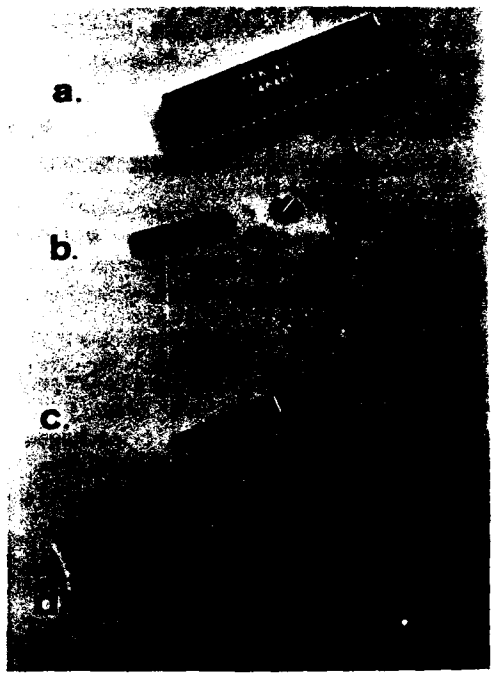


Fig. 2: Preparation steps from the IC-device to the fractured four-point-bend-test-sample

This core-sample consists of the plastic package above the chip, the chip itself, followed by a piece of the die-pad and the plastic material below. It contains all the three interfaces to be characterized.
To get a testable four-point-bend-test-sample, extensions of 25 x 5 x 2.5 mm^3 had to be sticked to both of the unpolished surfaces of the core-specimen (Fig. 2b u. 2c).
The phase-boundary under test has to be notched to be sure, that the crack runs along this interface.

2.2 The four-point-bend-test

This test was originally developed for the fracture-mechanical characterization of metal-to-ceramic-compounds /10, 11, 12/.
The three-point-bend-test, used for homogenous samples, has the disadvantage, that the upper support is located in the middle of the sample and therefore presses, in the case of heterogen compounds with one or more phase boundaries, on the interface itself.
The four-point-bend-test-geometry (Fig. 3) avoids this disadvantage, giving more reproducable results.

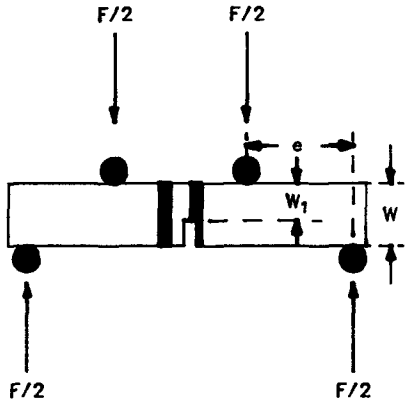


Fig. 3: Geometry of the four-point-bend-test

The critical fracture energy of an unnotched four-point-bend-test-sample can be calculated /14/ by the equation

$$G_c = \frac{3 F_c \cdot e}{B \cdot W^2} \qquad (1)$$

where F_c is the fracture-load and e, B and W described the geometry of the specimen (Fig. 3).

The width W_1 of the notched interface had to be used for calculating G_c of a notched sample.
Considering the place d and the depth $(W-W_1)$ of the notch, a correction-function Y_G can be calculated /10, 15/ and used for the determination of the fracture energy G_c of the notched heterogene interface /15/

$$G_c = F_c^2 \cdot \left(\frac{9e^2 \cdot \pi}{B^2 W^3}\right) \frac{Y_G}{E^*} \qquad (2)$$

where E^* is the effective modulus of elasticity.
If the elastic properties of the interface region are known, the fracture resistance K_c of the interface can be calculated /15/ by

$$G_c = \frac{1-\beta^2}{E^*} \cdot K_c^2 \qquad (3)$$

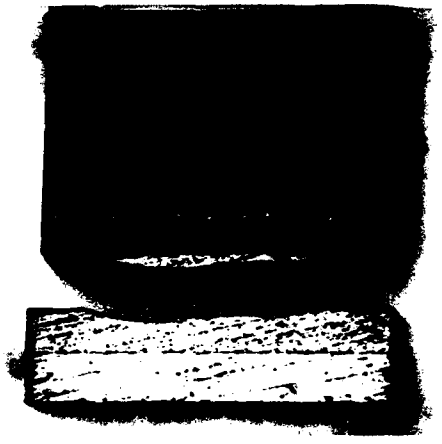
where ß is a composite parameter.

3 THE EXPERIMENTAL RESULTS

Because of the many elastic-, notch- and composit-dependent parameter, necessary for a quantitative calculation of the fracture-energy G_c or the fracture resistance K_c (chap. 2) of the various interfaces present in IC devices, a qualitative investigation of the respective adhesion has to be prefered.

Primarily, a four-point-bend-test-specimen will fracture in any case on its weakest point or interface.

a.



b.

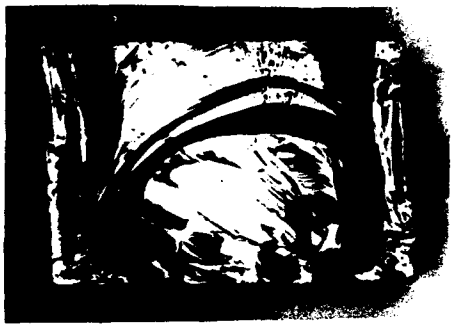


Fig. 4: Fracture behaviour of an untested IC-device
a. Side view of the fractured sample
b. Top view on the fractured surface

If the adhesion along one of the interfaces is higher than the fracture strength of one of the materials joined together, the crack will run through this material (see Fig. 4). In plastic packaged ICs the silicon is the material with the lowest fracture strength given by /16/

$$\sigma_c^{Si} = 1{,}2 \cdot K_c \cdot \frac{1}{\sqrt{\pi \cdot a}}$$

where K_c is the fracture toughness of the silicon (K_c^{Si} = 26 N/mm 3/2) and a is the surface crack depth in the Si.

The results of the qualitative analysis of the fracture behaviour of four-point-bend-test-samples were characterized as follows:

\+ the adhesion along each of the three interfaces is higher than the fracture-strength of the Si.
The fracture runs through the silicon chip (see Fig. 4a, 4b).

\- no or nearly no adhesion along one of the phase-boundaries. The sample fails along this interface during or after the preparation process with no or nearly no additional external force.

o the fracture strength of the weakest interface is smaller than the fracture strength of the Si, but high enough to be measurable using the four-point-bend-test-method.
The crack runs along the interface.

+/o the crack runs partly along the interface and partly through the Si-Chip (see Fig. 5). The fracture strength to be measured is a mixture of the Si-frac-

ture-strength and the fracture-strength of the interface (Fig. 5a, 5b).

a.

b.

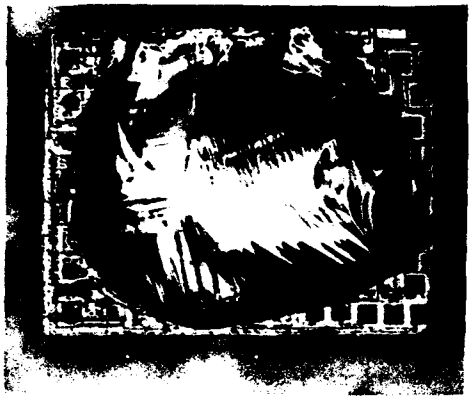


Fig. 5: Fracture behaviour of a temperature-cycled device
a. Si-chip side of the fracture
b. Mould part of the fractured sample

Which kind of fracture behaviour will be observable depends on the production parameters, the materials jointed together, the geometry of the interface and the reliability test done with the device.

The experimental results of the qualitative fracture-mechanical characterization of different types of ICs before and after several life-time tests are summarized in Tab. 1.

The adhesion along the interface die-pad/mould compound has been measured to be very low already before any test treatment of the IC-devices. The mechanical stability along this interface can be improved using dimples or anchor-holes in the die-pad, in which the mould compound can be pressed /16/.
A high pop-corn stability of the IC devices depends mainly on a good adhesion between the die-pad and the mould material.
The adhesion along the phase boundary die/die-pad of untested devices is stronger than the fracture resistance of the Si-chip itself (+ in Tab. 1) and independent of the die-attach technology (soldered or glued).
A similar fracture behaviour is observable after temperatur-stress-tests (temperature-cycles or solder-dips) and also after 1000 h in the standard 85 °C/85 % RH-humidity-test.
In the pressure-cooker-test (120 °C/ 100 % RH) it takes only about 50 h - 100 h to destroy (- in Tab. 1) the adhesion of the silver-epoxy adhesive die-attach.
The soldered die/die-pad-interface can be stressed more than 2000 h in the PC-test without any observable change of the high adhesion (+ in Tab. 1).

Manufacturer of the Device (S-DIP28)	Treatment before analysis	Die attach	Interface Adhesion: Moldcom./ die	Die/ die-pad	Die-pad/ Moldcomp.
A	--	silver-epoxy	+	+	-
A	85 °C/85 % RH	silver-epoxy	+	+	-
A	3x30 sec, 260 °C	silver-epoxy	+/0	+	-
B	500 h PC	silver-epoxy	-	-	-
A	500 h PC	silver-epoxy	-	-	-
A	3000 h PC	mod. adhesive	+	0	-
B	--	solder	+	+	-
B	3000 h PC	solder	+	+	-
A	--	solder	+	+	+ (anchor holes)
A	3000 h PC	solder	+	+	+

Tab. 1: Qualitative results of the fracture-mechanical characterization

The adhesion of the mould compound to the passivated chip surface is of essential interest for the reliability of the IC-devices.

The untested devices of standard production have a fracture-resistance along this phase-boundary higher than the fracture-strength K_c^{Si} of the silicon. The crack runs into the Si-chip (Fig. 4, + in Tab. 1).

After temperature-cycles (-50 °C, + 160 °C) or solder-dips (3 x 30 sec, 260 °C) the delamination starts from the chip edges along the interface chip-surface/mould compound and runs, in the middle region of the chip, into the Si-chip (Fig. 5; +/o in Tab. 1).

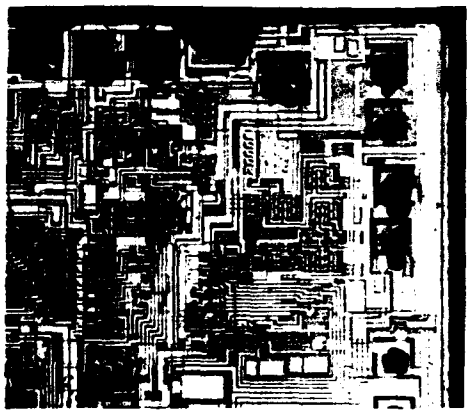
In the pressure-cooker-test the adhesion along the phase-boundary between the passivated chip surface and the mould compound fails already after about 50 h - 100 h test time, if a silver epoxy die-attach has been used. The fracture takes place along the interface itself (see Fig. 6). The fracture-strength is to low (- in Tab. 1) to be measured using the quantitative four-point-bend-test-method (see part 2.2).

The light-optical analysis of the separated surfaces (see Fig. 6 and 7) demonstrates, that, during the PC-test, the Al-metallization has been dissolved away in many of the bond-pads. The same is observable in the case of some of the conductor lines ending in this pads. The result of

this is an electrical failure (open), measurable after the PC-test.

Pads beneath the dissolved pads looks pretty good (see Fig. 6).

a.



b.

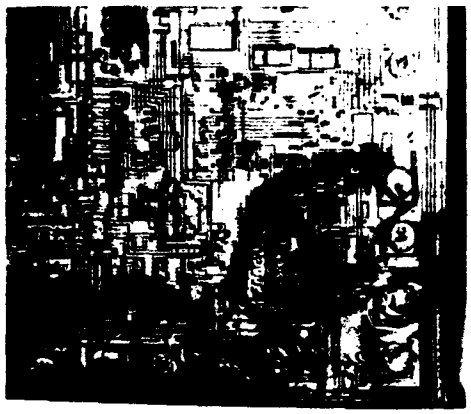


Fig. 6: Fracture behaviour of pressure-cooker tested devices
a. Si-chip surface
b. Mold part of the fractured sample

Further, a redeposition of the dissolved Al occurs in the delamination region, identical with the region where dissolved Al-pads can be observed on the chip surface (see Fig. 6). A redeposition film can be also found on the chip sides down to the meniscus of the silver epoxy adhesive (see Fig. 8). With WDX (wave length dispersive X-ray analysis) in the SEM oxygen and aluminium has be found in this redeposition films.
This indicates that deposition of $Al_x(OH)_y$ occurs in delamination-regions filled with water (see Fig. 7 and 8).
Using IC-devices with a soldered die-attach no decrease of the original high-adhesion along the interface chip-surface/mould-compound has been observed up to 3000 h pressure cooker test times (see Tab. 1). Also no electrical failure (short/open) and no dissolved Al-pads have been detected.

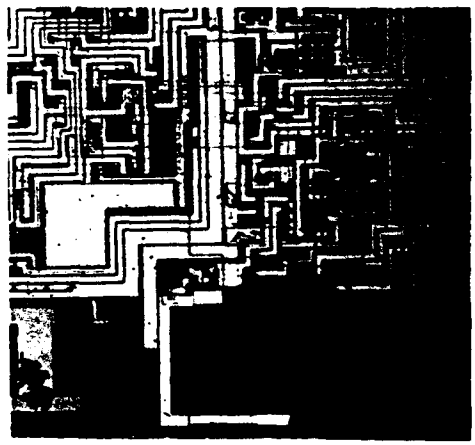


Fig. 7: Redeposition of $Al_x(OH)_y$ at the chip-surface

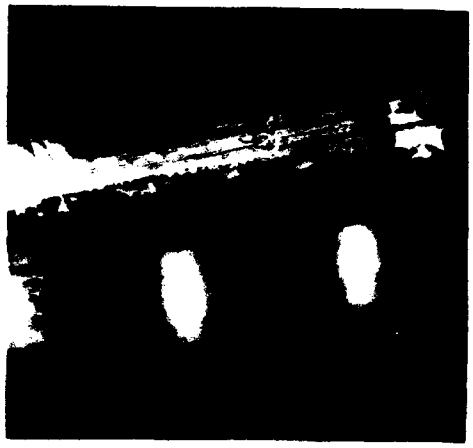


Fig. 8: Redeposition of $Al(OH)_x$ along the chip-side

4 DISCUSSION

4.1 Influence of the temperature treatments

The reason for the fracture behaviour observable after temperature cycles is the mechanical stress occuring during each temperature treatment because of the different thermal expansion coefficients of the Si-chip (α_{Si} = 2, 3 • 10^{-6} 1/K) and the mould material (α_M = 20 • 10^{-6} 1/K).
This mechanical stress σ_{MC} can be quantified by the following formular /16/

$$\sigma_{MC} = (\alpha_m - \alpha_{Si})(T - T_G)(1 + \frac{E_C \cdot A_C}{E_M \cdot A_M})$$

with: T_G = glas transition-temperature
E = elastic modulus
A_C = chip area
A_M = moulded area

The force across the interface because of this stress has its maxima in the edges of the chip.
In the unrelaxed state a warpage (see also Fig. 9b) of the bi-material system is the result of this stress across the interface.
Is this thermomechnical stress stronger than the fracture-strength along the phase-boundary, a delamination, beginning at the chip edges, occurs. The delamination region stops to increase, if the thermomechanical force at the distance x from the chip center is lower than the bending force between the chip-surface and the mould-material at the same place. On this place x the crack can run inside the silicon chip, if the bending across the interface is higher than the thermomechanical stress, which itself had to be higher than the fracture strength of the Si.
The mechanical stress applied during a four-point-bend-test had be added to the present internal stress and will fracture the pre-stressed (partly delaminated) system in the observed manner (see Fig. 5).
A way to prevent such delaminations and the following crack of the Si-chip during temperature cycle tests or solder-dips may be the use of modified mould compounds, the so called "low stress" - or even better "super low stress" mould-masses having better elastic - and thermomechanical-properties to compensate or lower the thermomechanical stresses occuring during the temperature treatments.

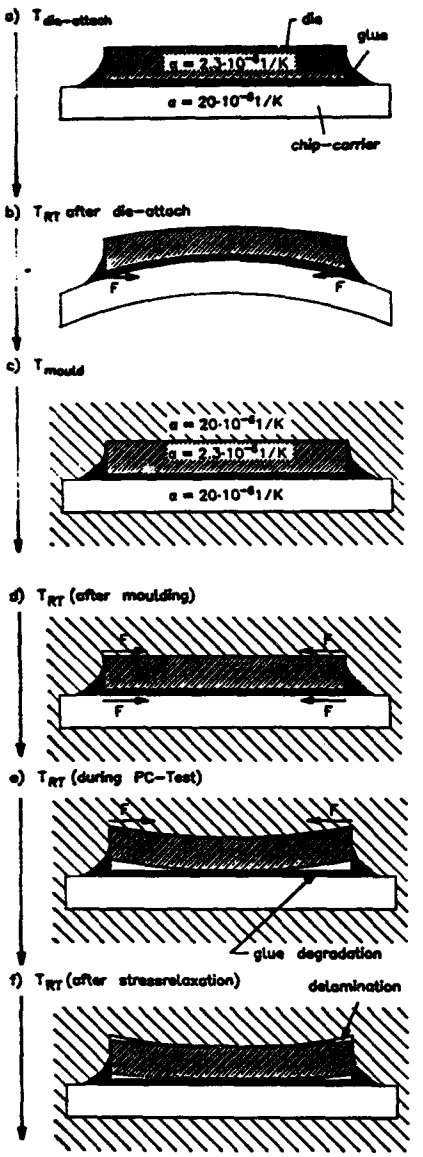


Fig. 9: Chip warpage as a function of process- and test-steps

4.2 Failure mechanismus in the pressure-cooker-test

The observed corrosion-phenomina occuring during the PC-test (dissolving of the pad-Al and of the Al out of the conductor lines ending in this pads and redepositing of $Al_x(OH)_y$ along the delamination region at the chipsurface and the chip sides (see Fig. 7, 8 and part 3) can't be found after the other common humidity tests (85 % / 85 % RH and HAST).
This phenomena are pressure-cooker specific material-transport mechanisms.

In the case of the common corrosion processes (surface corrosion, grain-boundary corrosion) the corrosion products ($Al_x(OH)_y$) were located direct at the place of the corrosion-reaction.
In the case of the PC-specific phenomena, the chemical dissolution of the Al-pads (dissolving of the Al out of the conductor lines) basing both on the electrochemical potential difference between the Al-metallization, the electrolyte (H_2O with solved ions (Cl^-, F^-...) and the epoxy (Ag)- die-attach material, a electrochemical-driven material-transport of the dissolved Al into the whole delamination region, filled with ionized water, takes place. The electrical potential of the active Si-regions, the conductor-line ends, can enforce the dissolving-process of the Al out of those metallization lines.
The growth of the redeposition film in the delamination region is a result of the saturation of the water in it with Al.

Water in a condensated state is possible in the PC-recipient because of the 100 % relative humidity (RH) condition.

The rapid decrease of the adhesion along the interfaces chip/mould compound and also along the phase-boundary chip/epoxy/chip-carrier also can't be observed after the other humidity tests (HAST, 85 °C/85 % RH).

The presented experimental results allowed a describtion of the PC-specific failure mechanism of IC-devices with an epoxy die-attach.

In the PC-test the following mechanisms takes place:

1. moisture diffusion through the plastic package
2. degradation (hydrolization) of the epoxy-die-attach
3. reduced or failed adhesion along the die-attach
4. decrease of the adhesion along the interface chip/plastic package
5. delamination along the interface chip/plastic
6. condensation of water in the delamination region
7. chemical dissolution process and battery effect
8. redeposition of the dissolved Al along the delaminated interfaces

This ends in a measurable electrical failure (open) after the PC-Test.
The process from the degradation of the epoxy die-attach to the delamination of the interface chip/mould compound is based on the mechanical stress between this two materials.

The Fig. 9 demonstrates the process-dependent change in the chip-warpage, which is the result of the mechanical stress between the different materials (different thermal expansion coefficients) present in the IC-device. After the die-attach process a warpage of the chip takes place (see Fig. 9b) because of the larger shrink-rate of the die-pad. After moulding (see Fig. 9d) nearly no warpage occurs because of the same thermal expansion coefficients of the die-pad and the mould material. The degradation of the die-attach adhesion (hydrolization of the epoxy) allows a warpage of the chip under the force F of the mould compound (see Fig. 9e). The decrease of the adhesion along the interface chip/mould-compound (point 4) is also a result of the epoxy-hydrolization and induces, together with the force F (warpage of the chip) the delamination along the chip surface (see Fig. 9f).
This delamination along the chip surface is necessary for the chemical dissolution of the pad-Al as well as for the battery-effect in this region, filled with condensed water.

5 SUMMARY

In plastic encapsulated integrated circuits three main interfaces can been distinguished. To characterize the bond strength or adhesion of this phase-boundaries a fracture-mechanical test method, the four-point-bend-test, has been used. Therefore, a new preparation technique was developed to get the four-point-bend-test-samples out of real-IC-devices.

A qualitative fracture-mechanical investigation of the fracture behaviour allows to distinguish four different possible failure modes for each interface.

The adhesion along the three interfaces of IC-devices before, during and after additional stress-tests (PC, HAST, solder dips, thermal cycle test) has been characterized.
The origin of two PC-specific corrosion phenomena has been described.
By preventing the source of the PC-failure-mechanisms, the life-time of the IC-devices under PC-conditions has been increased up to 3000 h and longer.

REFERENCES

/1/ P. Alpern, O. Selig, R. Tilgner, IEEE Trans. CHMT (1992), 926-929

/2/ K. van Doorselaer, K. de Zeeuw, Proc. IEEE 40th ECTC (199), 813-817

/3/ A. Nishimara, S. Kawai, G. Murakami; IEEE Trans. CHMT 12 (1989), 639-645

/4/ V. B. Jipson, Appl. Phys. Lett. 35 (1979), 385-387

/5/ T. M. Moore, Proc. ISTFA '89, ASM Int. (1989), p. 61-67

/6/ S. Kim, Proc. 41st ECTC, May 13-15, 1991, Atlanta, Georgia (1991), 1-9

/7/ H. Mykura, P. S. Bansal, M. H. Lewis, Phil. Mag. A. 42 (1980), 225 ...

/8/ A. P. Sutton, V. Vitek; Scripta Met. 14 (1980), 129 ...

/9/ A. G. Evans; Phil. Mag. A. 22 (1970), 841...

/10/ T. Saga, Dissertation, Univ. Stuttgart 1983

/11/ R. F. Pabst; Ber. Dt. Ker. Ges., 54 (1977), 392 ...

/12/ M. Rühle, K. Burger, W. Mader, A. G. Evans, Proc. of Seiken Symp. Dec. 2-4 (1985) Tokyo, Japan

/13/ J. T. Klomp, C. A. M. Mulder, g.d. Physique 46 (1985) C4-111

/14/ M. Turwitt, Disertation Universität Stuttgart, 1985

/15/ G. Pekow, T. Suga, G. Elssner, M. Turwitt, Sintered Metal-Ceramic Composites, edited by G.S. Upadhyaya, Elserier Science Publisher B.V. Amsterdam 1984, 3-18

/16/ H. Reichl, E. Zakel, in Conference Proceedings of the "ESREF 93" in Schwäbisch Gmünd v. 5.-8. Okt. 1993, VDE Verlag, p. 299-310

IONIZING-RADIATION AND HIGH-TEMPERATURE-INDUCED OXIDATION OF THE ELECTRONIC PACKAGING MATERIALS OF GE-HDI ELECTRONIC BOARD

Mohamad Al-Sheikhly & Aris Christou
Department of Materials and Nuclear Engineering
University of Maryland, College Park, MD 20742-2115, USA

Synopsis

Several methods have been developed for investigation of the thermal and ionizing radiation-induced oxidative degradation of polymers. These methods include the measurement of physical properties, as well as the detection of chemical changes. In the present work, the high-temperature and radiation-induced oxidations (chemical changes) were analyzed by detecting the non-volatile oxidation products and the unstable free radicals in GE-HDI electronic board. It is well known that the presence of ketones, aldehydes, hydroperoxides, and carboxylic groups is a direct indication of the oxidation in many polymeric materials[1].

In addition to the oxygen effects, it has also been demonstrated that some metals (e.g. Cu, Cr, Co, Si, Ti, Fe) and their oxides can markedly enhance the degradation of polymers and generally lower the activation energy for hydroperoxide decomposition[2,3] which ultimately leads to further degradation. Of particular concern is the acceleration of thermal degradation of the many polymeric packaging materials in the presence of copper[4]. Extensive research by some laboratories has demonstrated that cuprous and/or cupric ions can diffuse into a polymer and cause the decomposition of hydroperoxides which are formed when polymer films on copper begin to oxidize at elevated temperatures[5,6]. The catalytic action of copper oxides is due to an initiation reaction consisting of abstraction of tertiary hydrogen atoms. This will produce R° carbon-centered radicals and reduce copper oxide during the process. The hydroperoxide (ROOH) groups are decomposed by Cu^{+} ions which are oxidize to Cu^{2+} ions in the polymer films. The Cu^{2+} ions then decompose more ROOH groups and are reduced to Cu^{+}. The alkoxy (RO°) and peroxy (ROO°) radicals produced during these reactions then lead to the accelerated oxidative degradation of many polymeric electronic packaging materials[7].

Experimental Technique

In the present work, Fourier Transform Infrared (FTIR) spectroscopy and Electron Paramagnetic Resonance (EPR) spectroscopy were performed on the GE-HDI electronic board and its polymeric base material polyimide [KAPTON].

By utilizing the FTIR technique, the oxidative products can be detected as aldehyde, ketone, carboxyl, and ketonic carboxyl groups which can be monitored at 1500-1600, 1660, 1673, and 1716-1720 cm^{-1} wave number, respectively. A Perkin Elmer 1600 series model spectrofluorimeter

---

[1]Allen, Norman S. "Fundamentals of Polymer Degradation and Stabilization", Elsevier Applied Science, New York, 1992.

[2]Leo Reich and Salvatore Stivala, "Elements of Polymer Degradation", McGraw-Hill Book Company" 1971.

[3]Schnabel W. "Polymer Degradation Principles and Practical Applications," Macmillan Publishing Co., New York, 1981.

[4]Hansen R. "Thermal Stability of Polymers", R. T. Conley (ed.), Dekker, New York, 1970, p 153.

[5]Chan M. and Allara D. "Infrared Reflection Shades of Metal-Polymer Interfaces", Polym. Eng. Sci., No 14, 1974, pp 12-15.

[6]Hawkins W., Chan M. and Linke G. "Factors Influencing the Thermal Oxidation of Polyethylene," Polym. Eng. Sci., No 11, 1971, pp 377-384.

[7]Jellinek H., Kachi H., Czanderna A., and Miller A. "Thermal Oxidative Degradation of Isotate Polypropylene Catalyzed by Copper and Copper Oxide Interfaces", J. Polym. Sci. Poly. Chem. Edn., No 17, 1979, pp 1493-1522.

with a Spectra-Tech horizontal attenuated total reflectance (ATR) attachment was used along with a 45° germanium crystal. The ATR sampling technique makes it possible to analyze totally absorbing specimens such as GE-HDI board[8]. The method minimizes the effects of specimen thickness and orientation and reduces the error propagation among samples. Another advantage is that the need for a nitrogen purge to remove atmospheric water vapor and carbon dioxide is eliminated since the sample compartment is protected from undesirable elements[9]. The samples were placed on the crystal and pressure was applied with a sample clamp at a level which produced the best reproducibility. The samples were tested for 256 scans at 4.0 cm^{-1} resolution. The germanium crystal permits measurements at wavelengths of 5500-830 cm^{-1} which is the range for polymer peaks of interest.

The EPR study was carried out using a Bruker Model ESP 300 EPR spectrometer. Long-lived free radicals can cause severe damage to polymers (degradation) and result in premature aging. On the other hand, short-lived radiation- and thermally-induced free radicals are often harmless since they react bimolecularly to produce stable dimers (crosslinking reactions) or other products. The number of unpaired spin per gram (or the free radical concentration) was measured by using $MnSO_4.H_2O$ as standard. The samples and the standard were measured under the same conditions. The free-radical concentration of the irradiated KAPTON was measured according to the following equation[10]:

$$[x] = \frac{[std]A_x R_x (scan_x)^2 G_{std} M_{std} (g_{std})^2 [s(s+1)]_{std}}{A_{std} R_{std} (scan_{std})^2 G_x M_x (g_x)^2 [s(s+1)]_x} \qquad (1)$$

Where:

x: sample index
s: states of electron orbital angular momentum
g: g factor
A: measured area under the first-derivative absorption curve
scan: horizontal scale in G per unit length on the chart paper
G: relative gain of the signal amplifier
M: modulation amplitude

$$R = \frac{\sum_i D_i}{D_k}$$

D_k: degeneracy of the most intense line
$\sum_i D$: the summation of the degeneracies of all lines in the spectrum.

A ^{60}Co gamma-ray source was used to irradiate the KAPTON. The total absorbed dose was 100 kGy and the dose rate was 4.5 Gy s^{-1}.

Results and Discussion

Figures 1, 2, and 3 represent typical FTIR spectra of the heated and the unheated GE-HDI board. In Figure 1, the major difference between the spectra is seen in the band at 1716 cm^{-1}. The absorption peak at 1716 cm^{-1} wave number is direct evidence of the presence of the ketonic carboxyl group (oxidation product) in the heated sample[11,12]. The relatively high peak intensity, absorbance $A_{1716} \sim 0.01$, and very small thickness of the polymeric materials (5-10 μm) clearly demonstrate the high yield of oxidation, since the

[8]Fanconi B. M. "Fourier Transform Infrared Spectroscopy of Polymers-Theory and Application", Journal of Testing and Evaluation, JTEVA, Vol. 12, No. 1, Jan. 1994, pp 33-39.

[9]Messerscchimdt R. G. "A New Internal Reflection Element Design for High Optical Throughput in FTIR", Appl. Spectrosc. 40 (5), (1986) pp 632-635.

[10]Wertz J. E. and Bolton J. R. "Electron Spin Resonance Elementary Theory and Practical Applications", Chapman and Hall, New York, 1986.

[11]Tabb D. L., Sevik J. J., and Koenig J. L."Fourier Transform Infrared Study of the Effects of Irradiation on Polyethylene", Journal of Polymer Science, Polymer Physics Edition, Vol.13, 1975, pp 815-824.

[12]Kelly K., Yuichi Y., and Ishida H. "FTIR Reflection Technique for Characterization of Polyimides Films on Copper Substrate" Paper Presented at the 14th International Conference on Metallurgical Coating, San Diego, CA., U.S.A., March 23-27 1987, pp 271-279.

peak absorbance A is proportional to the thickness of the polymeric materials, and therefore the absorbance value A must be normalized per unit thickness (mm) of the sample.

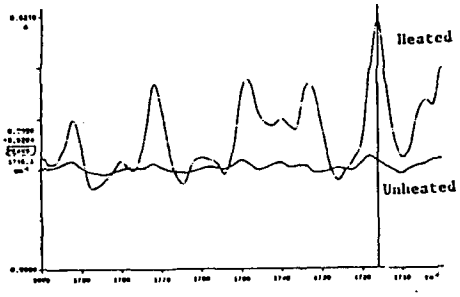


Figure 1: Absorption band at 1716 cm^{-1} (ketonic carboryl group).

Figure 2 demonstrates the difference between the FTIR spectrum of the heated and the unheated GE-HPI in the region 1500-1600 cm^{-1}. Once again, the multiple bands in this region give strong evidence of the presence of oxidation products in the heated sample [10,11].

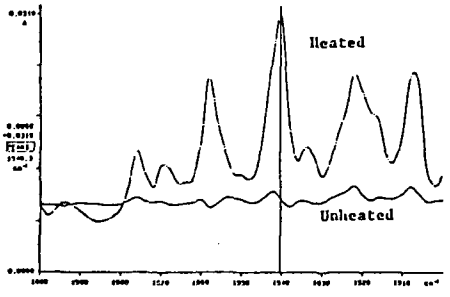


Figure 2: The presence of various oxidized products (multiple bands in the 1600-1500 cm^{-1} region)

In Figure 3, the presence of absorption bands at 1660 and 1673 cm^{-1} suggests the formation of COOH or C=O groups[11].

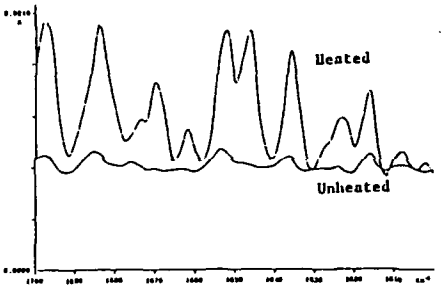


Figure 3: Absorption bands at 1600 and 1673 cm^{-1} (COOH and/or C=O)

It should be mentioned that severe oxidative degradation causes ring opening of the bismaleimid (e.g. KAPTON). This can be detected by monitoring increase in the FTIR peak at 2186 cm^{-1} [13]. Figure 4 shows no absorption band at 2186 cm^{-1} which suggests that no ring opening has occurred on the surface of the board.

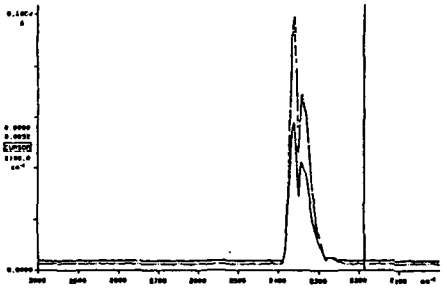


Figure 4: The absence of 2186 cm^{-1} peak (no ring opening on the surface)

Figure 5 represents the FTIR spectrum of the irradiated KAPTON (100 kGy) under aerobic conditions. The FTIR spectrum clearly shows the oxidation peak at 1716 cm^{-1}. A small oxidation peak at 1717 cm^{-1} was also detected in the FTIR

---

[13]Baldwin S. "Infrared and Ultraviolet Absorption Spectra of Enaminonitriles", J. Org. Chem., Vol. 26, 1961, pp 3288-3295.

spectrum of the unirradiated KAPTON sample. However, upon irradiation, a large increase in the oxidation peak was observed.

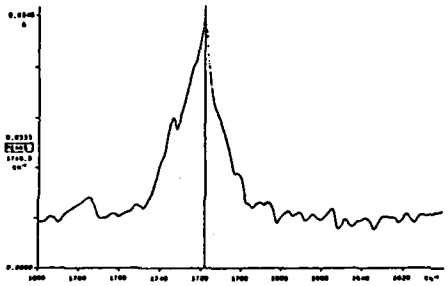


Figure 5: The formation of ketonic carboxyl group in the irradiated polyimide (KAPTON). Dose 100 kGy, dose-rate 4.5 Gy s^{-1}.

Table 1 summarizes the FTIR results of the heated GE-HDI board and the irradiated KAPTON samples.

Wave Number cm^{-1}	Heat/ Humidity	Radiation in the Presence of Oxygen 100 kGy	Oxidation Functional Group Assignments
1720	+	+	O ‖ R–C–R
1742	+	+	O ‖ R–C–H
1600-1500	+	+	Various Oxidation Products
1660-1670	+	+	R–COOH or R–C=O

Table 1: Oxidation products monitored by FTIR in the heated GE-HDI board and irradiated KAPTON[14]

Upon irradiation or treatment at high-temperature of the polymeric packing materials, carbon-centered free radicals will be produced. These carbon-centered radicals react either with themselves (bimolecularly) or with molecular oxygen to produce the corresponding peroxy radicals. These peroxy radicals will then initiate the oxidation process.

Figures 6 and 7 show the EPR first-derivative spectra of the irradiated and heated KAPTON (80°C), respectively. The shapes and the regions of absorption of these spectra are almost identical. These results suggest that the radiolytically produced free radicals and the heat-induced free radicals of the KAPTON are similar in their conformation and paramagnetic behavior.

Figure 6: EPR spectrum of the γ-liradiated polyimide (KAPTON) in the presence of oxygen. Total absorbed dose, 100 kGy; dose-rate, 4.5 Gy s^{-1}.

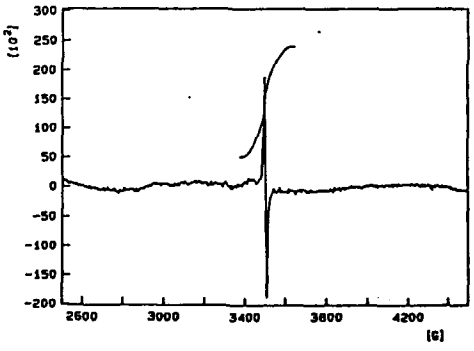


Figure 7: EPR spectrum of the heated polyimide (KAPTON) at 80°C (4 days) and relative humidity 100%.

[14] In Table 1, the character "+" indicates the presence of oxidational functional group.

Since in both cases, the EPR spectra were measured 20-25 days after the thermal oxidation treatment and irradiations, these free radicals remain as long-lived species. At this stage of this study and taking into account the shape and the region of absorption of 3200-3400 Gauss, these results suggest that these free radicals are of the carbon-centered R-C type. It should be mentioned, however, that one would expect the presence of the peroxy radicals from such a spectrum. By applying Equation 1, the number of spins per gram (number of free radicals per gram) for radiolytically produced free radicals (dose 100 kGy) was found to be 1.82×10^{19} ±10% unpaired spins per gram. The relatively high concentration of free radicals has the potential to initiate long-chain oxidation reactions.

Conclusion

The high amplitudes of the FTIR absorbance peaks at 1716, 1500-1600, 1660, 1673 cm^{-1} demonstrate the formation of high yields of oxidation products. In addition, the EPR results also indicate the high concentration of long-lived free radicals. The relatively high concentration of the oxidation products can be explained by the presence of trace amount of metals. Frequently, trace quantities of metals catalyze the oxidation process in polymeric electronic packaging materials. It is generally accepted that metals are catalytically active in the ionized state. The main function of metal ions consists in inducing the decomposition of hydroperoxides by redox reactions, thus generating free radicals. These free radicals react with oxygen to initiate more oxidation reactions.

As in the case of electronic packaging materials, a frequently encountered case pertains to an oxidative degradation process at metal/polymer interfaces. In this example, enhancement of the oxidation rate depends critically upon the rate of dissolution of metal ions, which, in turn, depends on the rate of diffusion of the ions into the polymer matrix[15]. Since GE-HDI board contains various metals (and/or metal oxide), it is very likely that the metals enhance the thermally-induced oxidation reactions. It may be expected that that these oxidation reactions occur in the metal/packaging material interfaces which will eventually lead to latent degradation and ultimately to delamination.

---

[15]Allara D. L. and White C. W., "Microscopic Mechanisms of Oxidatives Degradation and Its Inhibition at a Copper-Polyethylene Interface", J. Am. Soc. Div. Polym. Chem., Polym. Prepr. 18, (1977), pp 482-487.

Chip Surface Damage Induced by Internal Stress of Lead-On-Chip (LOC) Packages

Masazumi Amagai

New Package Development
Texas Instruments Japan, Hiji plant
4260 Takao, Kawasakiaza, Ohaza, Hiji, Hayami, Oita 879-15, Japan
Phone; 81-977-73-1729 Fax; 81-977-73-1582

1. ABSTRACT

In lead-on-chip (LOC) packaging technology, the lead fingers are attached directly to the surface of the chip using a double-side adhesive tape. This method of chip attachment naturally leads to concerns about stress on the polyimide coated chip surface. Device failure related to fracture in the passivation layers and the Al-Si metal has been observed in temperature cycle tests. To investigate the effect of material characterization on the surface damage, devices were fabricated with different types of molding compounds, tapes and polyimides. This paper describes the optimum material properties, the assembly process parameters, and the experimental and simulated results of the surface damage.

2. INTRODUCTION

The dramatic increase in the number of devices and functionality implemented on the latest ultra scale integration (ULSI) designs has resulted in large increases in chip size. On the contrary, package dimensions are continually being reduced to achieve higher component densities on circuit boards leading more rigorous requirements on the thermal and mechanical characteristics of the packaging technology. In parallel with shrinking package dimensions, the demand for increased device functionally leads to shrinking feature sizes, which are often more sensitive to package induced stress. As a result of these demands, the LOC package [1-5] (shown in fig.1) has begun to replace conventional package designs since it offers more margin from chip or die pad edge to the package outline. In the LOC packaging technology, the lead fingers are attached directly to the surface of polyimide deposited on the chip using a double-side adhesive tape. Polyimide films [6-7] are currently being used as stress buffers, to protect circuit elements from package induced stress damage. After the high temperature mounting process, double-side adhesive tape shrinks during cool down to room temperature leading to high internal stress. This stress can cause the chip surface damage during temperature cycling tests (-65 - 150 C) due to thermal expansion mismatch. This failure mechanism has been never reported in conventional packaging technology. This work investigates the role of material characterization and assembly process parameters on the surface damage during temperature cycling test.

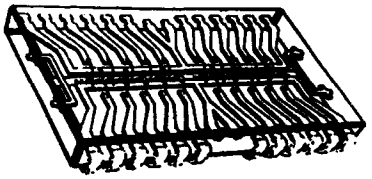


Fig.1 Small-outline-J-lead (SOJ) package with LOC design

3. CHIP SURFACE DAMAGE

In the LOC package, the lead frame is mounted to the polyimide coated chip surface by a double side adhesive tape. Dynamic random access memory (DRAM) packages with LOC structure exhibited the surface damage during temperature cycle test. An example of the observed surface damage is shown in fig.2. The polyimide, passivation, and metal line were fractured by mechanical stress induced by the thermal expansion mismatch between the different packaging materials. In most of the samples, the damage was localized at the lower corner of the tape edge, as illustrated in the cross-section in fig.3. It is believed that polyimide micro-cracking occurred primarily at the tape edge, subsequently propagating through the nitride/oxide and the metal lines during temperature cycling.

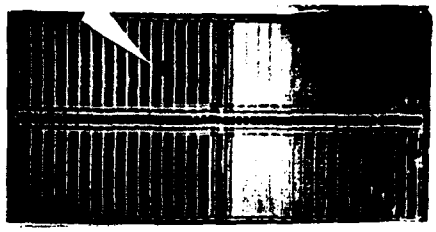


Fig.2 Chip surface damage

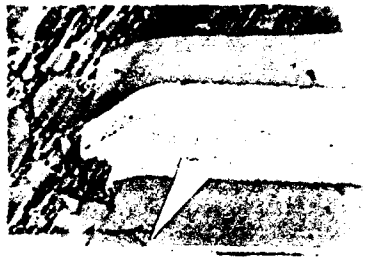


Fig.3 Cross-section

4. ASSEMBLY PROCESS PARAMETERS

4.1 Mounting impact

Dynamic stress simulations were performed to determine if the mounting impact could cause the surface damage. Finite elements model of SOJ devices were built using higher order (20 nodes) solid elements. A three dimensional solid element formation consisted of 20 nodes has been shown to produce a reasonably good approximation of the stress within chips even tough a mesh size is a little bit rough. Due to symmetry only quarter of the device was modeled. These models included a chip, lead frame, tape as shown in fig.4.The loading applied to the model was the mounting force of 4 kg/IC. The von-mises (Sig E) stress values for chip internal stress depended upon times are shown in fig.5.The von-mises stress is determined by equation (1).

$$\sqrt{\frac{1}{2}[(\sigma 1-\sigma 2)^2+(\sigma 2-\sigma 3)^2+(\sigma 3-\sigma 1)^2]} \quad (1)$$

Where $\sigma 1$, $\sigma 2$ and $\sigma 3$ are principle stress. The highest value of stress were 5.2 kg/mm2 at 0.8 E-6 seconds after the lead frame was attached to the chip surface. It is expected that this value could not cause the damage, since fracture strength above 10kg/mm2 was confirmed for the passivation, chip and polyimide, respectively.

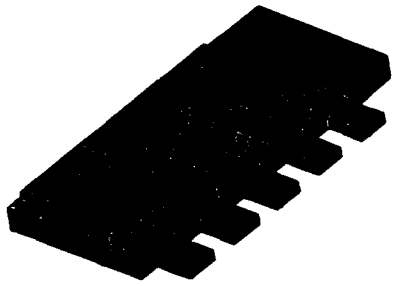


Fig.4 LOC mounting impact model

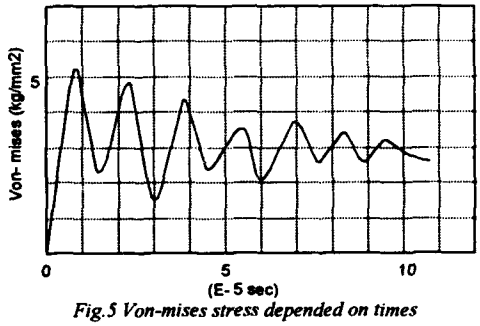


Fig.5 Von-mises stress depended on times

4.2 Bonding impact

In the analysis which follows we examine the dynamic response of finite element model, having a structure of lead frame mounted on the chip surface, a bonding capillary collided with the lead frame, and shock load of 0.2 kg. The model of this bonding impact is shown in fig.6. Due to symmetry, only quarter of the composite structure was analyzed. It consisted of three dimensional solid elements with 20 nodal points and three degrees of freedom. The von-mises stress values divided by times is shown in fig.7. It can be seen that the maximum stress (1.2kg/mm2) occurs under a stitch bonding placement at 0.8E-6 sec. It is believed that this stress could not attribute to the surface fracture.

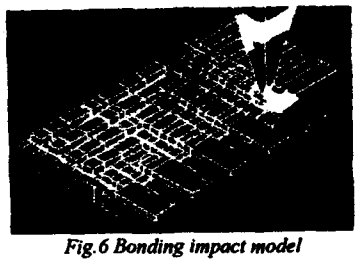


Fig.6 Bonding impact model

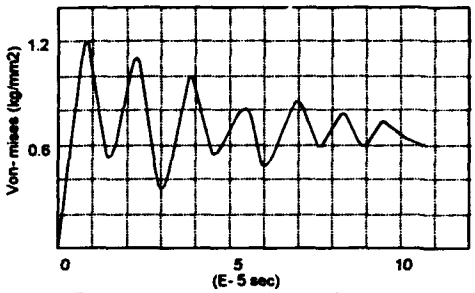


Fig.7 Von-mises stress depended on times

5. CHIP COATING POLYIMIDE

5.1 Chemical stripper selection

Two types of polyimides were used to evaluate chemicals as tape stripper, a photosensitive type, polyimide (A), and a non-photosensitive type, polyimide (B). Fig.8 illustrates the respective forerunners of polyimide (A) and (B) dissolving in a solvent prior to imide cyclic polycondensation by thermal dehydration. Half of the wafer samples were coated with polyimide (A) while the other half were coated with polyimide (B). Polyimide (A) coated wafers were subsequently divided into two process conditions as shown in table 1. Polyimide (A) was cured at 390 C in a nitrogen ambient and an air ambient, respectively, while polyimide (B) was cured at 350 C in the nitrogen ambient. All of the wafers were subsequently processed into chip samples, which were bonded by a double-sided adhesive tape with a lead frame. Four types of chemicals were utilized as tape stripper for removing the chip from the tape with the lead frame. The conditions for chemically stripping the tape from the chips are shown in table 2.

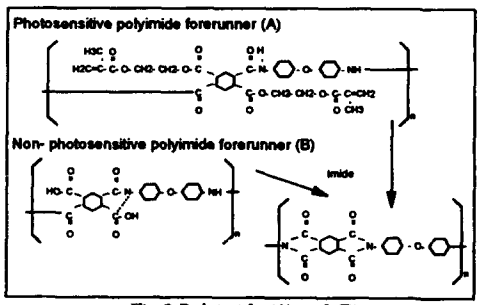


Fig.8 Polyimide (A) and (B)

The conditions were sulfuric acid (40 C 1 hour and 50 C 1 hour) followed by water cleaning, 105 (dimeth sulfoxide and monoethanol, alkali type) (80 C 1 hour) followed by acetone cleaning, EDA (ethylene diamine,

	Cure condition	
Polyimide	Cure temperature	Cure atmosphere
(A)	390 C	Nitogen gas
(A)	390 C	Air
(B)	350 C	Nitrogen gas

Table 1 polyimide cure conditions

	Stripper		
	Acid	Alkali	Non- Acid/Non- Alkai
Chemical	H2SO4	EDA *1 105 *2	104
Condition	40 C x 1 hour + 50 C x 1 hour	*1 40 C x 3 hours *2 80 C x 1 hour	80 C x 1 hour
Rinse	Water	Acetone	Acetone

Table 2 Tape stripper

alkali type) (40 C 3 hours) followed by acetone cleaning, and 104 (dimeth sulfoxide and N-methyl pyrolidinone, no acid and no alkali type) (80 C 1 hour) followed by acetone cleaning.

In order to avoid chemically induced polyimide cracking, four types of chemical tape stripper were evaluated. An example of the polyimide cracking induced by chemical damage and mechanical stress is shown in fig.9. The occurrence of polyimide cracking for each of the four different chemicals used in the experiments (as determined with SEM) is shown in table 3. Polyimide (A) coated chip cured in a nitrogen ambient exhibited no polyimide cracking for the 104 chemical while polyimide (A) coated chip cured in an air ambient was found to exhibit no polyimide cracking for the EDA and the 104 chemicals. For polyimide (B), the occurrence of polyimide cracking was observed for the sulfuric acid and the 105. From stress-strain characteristics shown in table 4, the tensile stress and strain break point were 15.7 kg/mm2, 44 %, 16.0 kg/mm2, 43%, and 18.5 kg/mm2, 57% for polyimide (A) (a nitrogen ambient and an air ambient) and polyimide (B), respectively. Polyimide (A) cured in the air ambient

Fig.9 Polyimide cracking

Polyimide	Cure condition	H2SO4	105	EDA	104
(A)	390 C, Nitrogen	Crack	Crack	Crack	No crack
(A)	390 C, Air	No crack	No crack	No crack	No crack
(B)	350, Nitrogen	No crack	No crack	Crack	Crack

Table 3 Polyimide cracking

Polyimide	Cure condition	Tensile stress	Strain break	Density	CTE(PPM)	
					< Tg	> Tg
(A)	Nitrogen, 390 C	15.7 kg/mm2	44 %	1.39	38	600
(A)	Air, 390 C	16.0 kg/mm2	43 %	1.42	37	3700
(B)	Nitrogen, 350 C	18.5 kg/mm2	57 %	1.40	40	800

Table 4 Polyimide properties

was found to exhibit superior chemical and mechanical resistance as compared with polyimide (A) cured in the nitrogen ambient, despite the fact that both conditions have similar bulk physical properties. Furthermore, the air ambient leads to high internal stress around the interface between the polyimide (A) and the tape as compared with the nitrogen ambient, since polyimide cured in the air ambient illustrates higher CTE above Tg than polyimide cured in the nitrogen ambient. It is believed that polyimide cracking is caused primarily by chemical damage. The density of polyimide (A) cured in the nitrogen and air ambient was 1.39 and 1.42, respectively. Difference in densities is attributed to a three dimensional structure of the polyimide molecule.[8] Based on the results of the experiments, the 104 chemical was used as tape stripper which eliminates the polyimide damage related to chemical attack.

5.2 Polyimide cracking

Half of various polyimide-coated chip samples mounted on the lead frame were stripped from the lead frame using the 104 chemical without causing chemically induced polyimide damage while the other half of the samples were decaped to observe whether or not chip surface damage occurs after encapsulation subsequent to temperature cycle tests. The polyimide mechanical property was determined from the tensile tests. Fig.10 illustrates polyimide mechanical energy which is an area under stress-strain curved based on the tensile mechanical test. The polyimide mechanical energy is typically determined by a yield point, elastic modulus and a broken point of polyimide during the tensile test. The mechanical energy is given by equation (2).[9-10]

$$\int_0^S \delta \, ds \qquad (2)$$

Where δ is stress, and S is strain. Fig.11 exhibits the relationship between polyimide cracking and the polyimide mechanical energy. Polyimides with a mechanical energy of, or exceeding , 575 kg/mm2 exhibited no polyimide cracking after the samples mounted on the lead frame were stripped from the lead frame using the 104 chemical. Fig.12 depicts the relationship between polyimide cracking and the chip surface damage after temperature cycle tests. If no polyimide cracking occurred, no surface damage was observed. This experiment has revealed that the temperature cycle performance of the LOC package is determined largely by the polyimide properties, specifically the polyimide's resistance to cracking. The results of a non-linear stress simulation model of the LOC package are shown in fig.13. A fine mesh was used at the lower corner of the tape edge to improve the resolution in this area. The simulation results suggested that a polyimide having more than 15.0 kg/mm2 of fracture strength and a 40 %, or greater, elongation, eliminated polyimide cracking and surface damage. These values of fracture strength and elongation corresponded to a polyimide mechanical energy of 575 kg/mm2. Simulation and experimental results have revealed that the degree of surface damage is determined by the polyimide mechanical energy, and that a mechanical energy of at least 575 kg/mm2 is required to eliminate polyimide cracking.

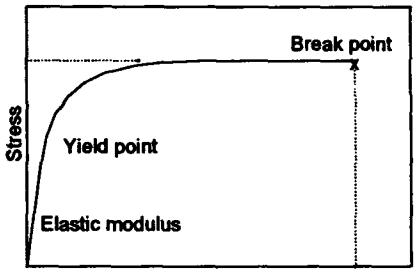


Fig.10 Polyimide stress-strain

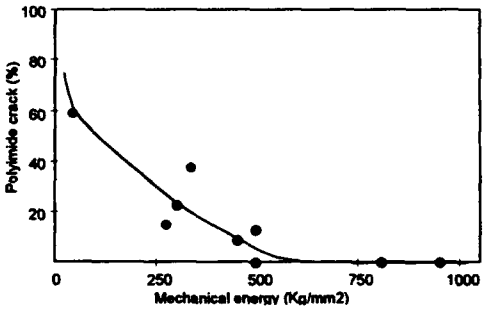


Fig.11 Mechanical energy and polyimide cracking

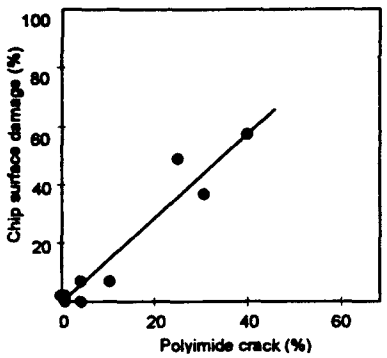


Fig.12 Chip surface damage and polyimide cracking

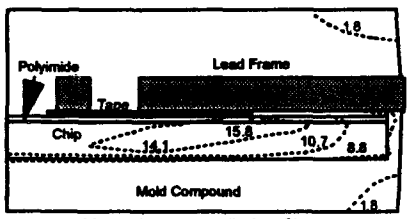


Fig.13 Package internal stress

6. LOC TAPE

Two types of LOC tapes were used to evaluate chip surface damage. Table 5 exhibits the respective properties of LOC tapes, CTE (coefficient of thermal expansion, E (elastic modulus), thickness and delta temperature ranging from tape attached temperature to -65 (° C). In order to accentuate the effect of LOC tapes, the polyimide (mechanical energy 260 kg/mm2) was used. All the samples were encapsulated in small outline J-lead (SOJ) packages. After encapsulation, all the packages underwent a temperature cycling test (-65 - 150 °C). The occurrence of surface damage was determined by Scanning electron microscopy (SEM).

Tape	CTE (E- 5 1/C)	E (Kg/mm2)	Thickness (um)	Delt temp (C)
A	3.0	302	100	465
A	3.0	302	175	465
B	1.0	900	100	465
B	1.0	900	100	265

Table 5 LOC tape properties

The values of chip strain, ε, is given by equation (3). Where E1 and E2 (20000 kg/mm2) are elastic modulus for the tape and chip, respectively, α1 and α2 (2.5 ppm 1/ °C) are respective CTE for the tape and chip, ΔT is delta temperature during the tape attached process cool down to -65° C,. Th1 and Th2 are thickness for the tape and chip.

$$\varepsilon = \frac{E_1(\alpha_1 - \alpha_2)\Delta T\, Th_1}{E_2\, Th_2 + E_1\, Th_1} \qquad (3)$$

Fig.14 illustrates the chip strain as a function of temperature cycles (Log N) for the surface damage. It was confirmed that the occurrence of surface damage related failures was linear from 2.4 to 3.5. The results of these characterization reveals that higher strain lead to lower cycles related failures. To eliminate the surface damage, the value of chip strain should be decreased down to lower one, using smaller values of CTE, thickness and delta temperature.

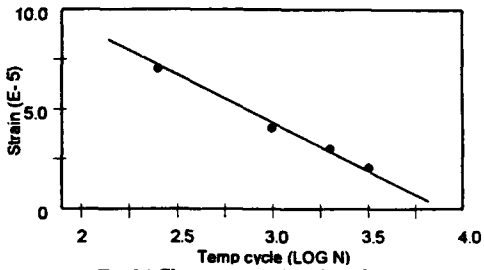


Fig.14 Chip strain and surface damage during temp cycle

7. MOLD COMPOUND

The impact of mold compound on the occurrence of surface damage was investigated with the experimental matrix illustrated in table 6. Three types of mold compound with several CTE were investigated. 260 kg/mm2 of mechanical energy for the polyimide was used. All samples were encapsulated in SOJ packages and subsequently underwent temperature cycles. After the samples were decaped, the surface damage was observed by SEM. The surface damage was typically observed to be caused by delamination between the tape side and mold compound. The occurrence of delamination was related to different thermal expansion between the tape and mold compound.

Mold compound	CTE (E- 5 1/C)	Elastic modulus (Kg/mm2)
A	0.8	2200
B	1.0	1800
C	1.2	1400

Table 6 Mold compound properties

The impact of varying properties of mold compound was investigated by the solution of following equation for the magnitude of chip surface strain and stress leading to polyimide change from elastic to plastic region. The strain ε1 for the tape and chip, ε2 for the tape and mold

compound, and ε3 for the chip and mold compound are obtained by equation (4), (5) and (6), respectively. The strain ε4 for the chip at the lower corner of tape edge is calculated by equation (7).

$$\varepsilon_1 = \frac{E_T(\alpha_T - \alpha_c)\Delta T\, Th_T\, W_T}{(E_c\, Th_c + E_T\, Th_T)\, W_c} \quad (4)$$

$$\varepsilon_2 = \frac{E_T(\alpha_M - \alpha_T)\Delta T\, Th_T\, W_T}{(E_M\, Th_M + E_T\, Th_T)\, W_T} \quad (5)$$

$$\varepsilon_3 = \frac{E_M(\alpha_M - \alpha_c)\Delta T\, Th_T\, W_M}{(E_M\, Th_T + E_c\, Th_c)\, W_c} \quad (6)$$

$$\varepsilon_4 = \varepsilon_1 + \varepsilon_2 - \varepsilon_3 \quad (7)$$

Where E_T (900 kg/mm2), E_c (20000 kg/mm2) and E_M are elastic modulus, α_T (10 ppm 1/°c), α_c (2.5 ppm 1/°c), and α_M are CTE, Th_T (0.1 mm), Th_c (0.23 mm) and Th_M (1.0 mm) are thickness, W_T (3.0 mm), W_c (5.0 mm) and W_M (2.0 mm) are width, and ΔT is temperature for the tape, chip and mold compound, respectively. The polyimide stress δ is determined by equation (8). The polyimide stress δ_y in Y direction is given by equation (9).

$$\sigma = \varepsilon_4 E_c\, Th_c / Th_p \quad (8)$$

$$\sigma_y = K_1 / \sqrt{2\pi x} \quad (9)$$

Where K1 is the polyimide stress intensity at the tip of delamination between the tape side and mold compound, x is the polyimide plastic region at the tip of delamination, and Th_p is the polyimide thickness. δy is similar to the yield point, δY, of polyimide. The length, r_p, of polyimide plastic region is almost equal to x. r_p is determined by equation (11).

$$\sigma_y = \sigma_Y,\ x = r_p \quad (10)$$

$$r_p = \sigma^2 a / 2\sigma_Y^2 \quad (11)$$

Fig.15 exhibits the length of plastic region of polyimide for the surface damage during temperature cycles (Log N). From the length of plastic region in conjunction with the surface damage related failures, it is evident that larger plastic region leads to lower cycles related failures.

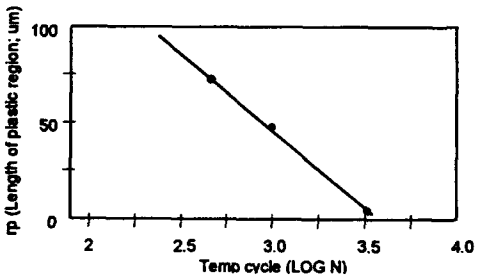


Fig.15 Length of plastic region and surface damage during temp cycle

8. CONCLUSION

The results of these experiments and simulation suggest that the chip surface damage and polyimide cracking in LOC packages is determined primarily by the polyimide mechanical energy based on stress-strain curve, properties of the LOC tape and mold compound. It is evident that polyimide micro-cracking occurred primarily at the tape edge caused by the tape shrinkage, followed by propagating through the nitride/oxide and the metal line during temperature cycling. The improved mechanical toughness of polyimide and the optimized properties of the tape and mold compound reduce the probability of the chip surface damage and polyimide cracking in the LOC package.

9. ACKNOWLEDGMENTS

The authors would like to thank Akira Karashima and Eiji Kawasaki in Texas Instruments Japan for their support and technical advice.

10. REFERENCES

[1] William C.Ward, "Volume Production of Unique Plastic Surface-Mount Modules for the IBM 80-ns 1-Mbit DRAM Chip by Area Wire Bond Techniques," Proceedings of the 38th IEEE ECC, 1989

[2] D.Hagen, et al., "Lead On Chip TSOP Assembly Process for Fast SRAM with Peripherally Located Bond Pads," Proceedings of the IEEE CHMT International Electronics Manufacturing Technology Symposium, September 1992, pp.39-47

[3] Mike Lamson, et al., "Lead-On-Chip Technology for High Performance Packaging," Proceedings of the 1993 IEEE ECTC, 1993, pp.1045-1050

[4] Masazumi Amagai, et al., "Development of a Tapeless Lead-On-Chip (LOC) Package," Proceedings of the 1994 IEEE ECTC, 1994, pp.506-512

[5] Masazumi Amagai, et al., "Cracking Failures in Lead-On-Chip Packages Induced by Chip Backside Contamination," Proceedings of the 1994 IEEE ECTC, 1994, pp.171-176

[6] Masazumi Amagai, et al., "Polyimide Surface Characteristics for Adhesion Strength at the Interface Between Polyimide and Mold resin," Proceedings of the 1993 IEEE IPFA, 1993, pp.6-10

[7] Masazumi Amagai, et al., "The Effect of Polyimide Surface Morphology and Chemistry on Package Cracking Induced by Interfacial Delamination," Proceedings of the 1944 IEEE IRPS, 1994, pp.101-107

[8] H.Ishida, K.kelley, Polymer 32, 1585-1588 (1991)

[9] Masazumi Amagai, "Computer Modeling for IC packages," Proceedings of the 1993 ANSYS Japan conference, 1993, pp.209-212

[10] Masazumi Amagai, "Chip Surface damage of internal stress of Lead-On-Chip Package," in the 1994 MRS symposium.

Survey on Electrostatic Susceptibility of Integrated Circuits

Horst Gieser*, Ingolf Ruge**

* Fraunhofer-Institute for Solid State Technology (IFT/TMS),
ATIS Applied Test for Integrated Circuits
Hansastr.27d, D-80686 Munich, Germany
Tel: ++49-(0)89-54759-020 Fax: ++49-(0)89-54759-100

** Technical University of Munich (TUM),
Chair for Integrated Circuits (LIS)
Arcisstr.21, D-80290 Munich, Germany
Tel: ++49-(0)89-231817-0 Fax: ++49-(0)89-231817-30

SUMMARY

From traditional assumptions and simplifications, different attitudes towards electrostatic discharges (ESD) in present microelectronics have been developed. The central question remains, which level of ESD-susceptibility and which level of costly external protection measures is needed to limit yield losses and potential reliability risks for integrated circuits. This paper addresses influences, links and current trends in protection design and ESD-test in the context of ESD-susceptibility of integrated circuits. One focus will be the Charged Device Model (CDM).

THE "SURVEY ON CRITICAL RELIABILITY ISSUES IN EUROPE"[1] ranked Electrostatic Discharge (ESD) among the top three problems for present and future generations of integrated circuits. For microelectronics, ESD means that differences in the electrostatic potential of two bodies result in fast to ultra-fast (10^{-7}... 10^{-10}s) current transients passing sensitive structures of integrated circuits and/or raising hazardous internal voltages. These "Real World" events may occur in various situations, simplified to the Human Body Model (HBM)[2], Machine Model (MM)[3] and Charged Device Model (CDM)[4,5]. If a certain threshold of stress is exceeded, ESD may cause different physical failure signatures. Technological and design measures help to protect and harden the sensitive structures. However, these measures may often collide with requirements for functional or reliability performance and a given chip size or circumference. The stress models HBM, MM and CDM were implemented in ESD-testers to quantify the susceptibilty by means of simulated ESD-events. From internationally standardized test methods[6] including the calibration of the tester we would expect tester independent correlated failure thresholds to decide between two products and to tailor external and internal protection efforts. However, even for HBM with its moderate double exponential discharge current and time constants, this is still not satisfactory[7,8]. However, it is improving by means of a more precise definition of the discharge characteristic[9]. For faster transients like the CDM, the interaction between device and tester becomes even more complex and the progress of the know-how is directly related to the most advanced metrology[10]. The fact that some types of failures are only reproduced by CDM-stress and that good HBM-susceptibility does not necessarily indicate sufficient CDM-results - and vice versa - calls for the application of at least two stress models[10,11]. One question to be answered remains, whether MM adds information to the HBM results, or addresses the same physical failures at a lower failure voltage.

This paper is intended as a brief survey of different factors, links and trends (Fig.1) that should be considered in the context of ESD-susceptibility.

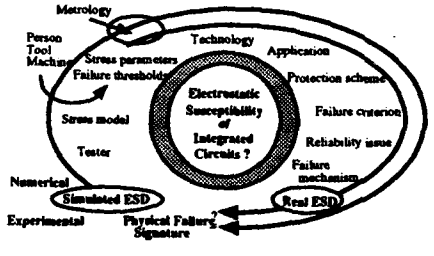


Fig. 1: Influences on the electrostatic susceptibility of integrated circuits to obtain an equivalent physical failure signature

Instead of its evolution for specific product groups or technologies, we will discuss principles and examples for the different issues of Fig.1 aiming at application oriented protection and circuit designs and better correlatable ESD-tests. Because of its increasing importance associated with automated handling and larger, thinner packages, one focus will be on ultra-fast transient CDM.

REAL WORLD ESD

Objects may be charged electrostatically by **contact electrification** or by **field induction**. The first means that two bodies of different work functions in intimate contact are seperated leaving excess electrons on the body with the higher work function and empty traps at the other. Friction is not mandatory, but increases the charge transfer. This effect depends too much on the actual, non-ideal surface status, to trust the triboelectric series for charge generation of materials[12]. The **field induced charging** of a body requires a surrounding electrostatic field and the process of contact and separation with another non-insulative body on a different electrostatic potential to allow a transient flow of mobile charge from one body to the other. In particular, one body may be either the "hot" pole of a voltage source or "earth". The first is commonly called "direct pre-charge", the latter "discharge". According to Poisson's Law, the overall distribution of mobile and immobile charges in a system results in the electrostatic field. However, only mobile charge, e.g. on the lead frame of a device or a person, contributes to the hazardous discharge current.

In "Real World", we may find numerous objects, such as persons, cables, machine parts, containers, bags, device packages and lead frames, which become contact (tribo-) charged even by dust particles. The differences of mobile and immobile charge are neutralized via various or no paths, which depend on the field strength and the environment, e.g. the humidity and ion contents of the air, and, of course, the external ESD-protective measures. The maximum amplitude of the current, which may be most hazardous to integrated structures, is reached at minimum resistance. Access to the **stress parameters pre-charge voltage** and **discharge current** may be rather difficult for "Real World" situations, and the reproducibility of experiments employing contact charging may be considered to be weak.

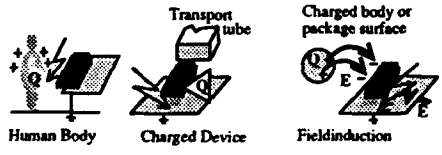


Fig. 2: Some "Real World" ESD-events

Nevertheless, the **analysis of the physical failure signatures** (FMA) of yield losses and customer returns together shows up to approximately 60% significant ESD and electrical overstress (EOS) failures[13,14]. Significant means, they are reproducible by simulated EOS and ESD, e.g., in an ESD-tester.

PHYSICAL FAILURE SIGNATURES

Devices failing electrical criteria, which will be discussed later, during parametric or functional tests are deprocessed to identify the physical failure signature. The localization of the physical failure is done by means of liquid crystal thermography, light emission microscopy or electron beam testing, sometimes in combination with laser cutting. Non of the methods is able to cover all types of defects. Depending on the type and level of ESD or electrical overstress (EOS), as well as the design[15], we may find **burned-out metal, burned-out pn-junctions** (Fig.3, left), **contact spikes** or **ruptured gate oxides** (Fig.3, right) by means of an optical or electron-optical inspection. Combinations of them cannot be excluded.

Fig. 3: Melt channel in burned-out junction (left) and ruptured gate oxide in input buffer (right)

The **reproduction of these physical failure signatures** is the conditio sine qua non of the **experimental simulation. Numerical simulation** may assist to tailor the experiment and to indicate power distribution and voltage clamping in the ESD-relevant high-current domain.

SIMULATED WORLD ESD

ESD-testers shall emulate real events under reproducible conditions and allow an easy and fast stressing of large numbers of devices and pins.

For this purpose, the various real events are reduced to few **representative "Stress Models": Human Body Model (HBM), Machine Model (MM) and Charged Device Model (CDM).** They are characterized by the discharge current over time for specific (ohmic) loads[3,6,9,16]. When an ideal lumped element model (Fig.4) is assumed, effective values of resistor R, inductor L and capacitor C (Tab.1) may be found that allow us to calculate the discharge current waveform. Tab.1 distinguishes the different models by means of the key stressors[5].

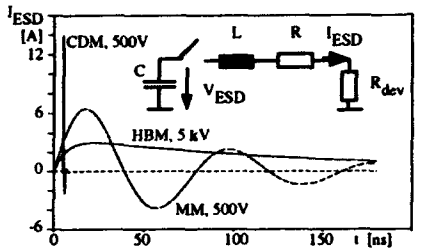


Fig. 4: Simplified RLC-circuit and resulting calculated discharge currents for typical 5kV HBM, 500V MM and 500V CDM discharge

Stress model	HBM	MM	CDM
C/pF	100	200	10
L/nH	10000	750	2.5
R/Ω	1500	10	10
Stress at load R_{dev}=10Ω @ V_{ESD}=500V			
-Oscillating	No	1)	1)
-Rise time/ns	≤10	<10	0.1
-1/e-Decay time/ns	150	<100	1
-Peak current/A	0.3	6.5	14
-Peak power/W	0.9	400	2000
-Energy, dissipated /µJ	0.08	13	0.63
-Energy, stored /µJ	12.5	25	1.25

Tab. 1: Standardized and typical parameter values for simplified RLC-circuits of HBM, MM, CDM (Fig.5)
1) Depends on L, R and breakdown voltage of protection device

We may learn from these strongly simplified circuits:

- The major part of the energy stored on the HBM-capacitor is dissipated in the 1500Ω-HBM-resistor and cannot heat the stressed structure (load);
- HBM-resistor and load resistor form a voltage divider. The resulting voltage may even stress the field oxide;
- When low additional non-linear circuit resistances (switch, protection device) may be assumed, the 1500Ω-resistor quasi-forces the HBM discharge current;
- Current levels of HBM are reached at significantly lower MM-voltages (here: V_{HBM}=22V_{MM}).
- The rise and decay times of HBM and MM are in the same domain. CDM-pulses have by two orders shorter rise and decay times;
- For equivalent voltage, the MM and CDM-currents are one order above the HBM-current. In our resistive structure, the peak power increases with the square of the current ($P=I^2*R$). Therefore, the CDM-peak power reaches up to the kW domain followed by the MM;
- The higher capacitance (200pF) of the MM stores approximately 20 times the energy ($W=0.5CV^2$) of the CDM-case. A large amount of this energy has to be dissipated in the resistive structure.

However, if we "implement" these simplified ideal stress models HBM, MM and CDM in a tester and stress devices, the actual discharge current results from more or less complex interactions between the tester that is not an ideal RLC-circuit and the device. Typical **tester parasitics** studied up to now are the **capacitance of the testboard** in the discharge path [7,8,17,18] and the **background capacitance** behind the device[10,17], the **capacitance and arc resistance in the discharge relay**[7] and the **non-zero-resistance and inductance in MM**[3]. In addition, it should not be overlooked that the interconnects in a real tester are **extended structures** and, thus, **very fast transients** may travel along these "quasi" transmission lines and be reflected at each impedance discontinuity[10,18]. The faster the transients become, e.g. CDM, the more important is this view.

It is obviously a non-trivial challenge to **match two testers of principally different designs** to be able to apply the **same quantity of stress** to an integrated structure on the basis of the **same pre-charge voltage**. Rise time dependent switching mechanisms (= dV/dt-triggering of SCRs and NMOS/Bipolar transistors [10,17]) in the device make the situation even worse.

For HBM, the high 1500Ω-source impedance implies reasonable correlation. However, the important testboard capacitance is shortened in the MIL-standard[6] characterization procedure and thus neglected. An additional characterization with a 500Ω–load resistor has been introduced by the ESD-Association's HBM-standard[9] that will probably be adopted by ANSI and MIL.

The bad correlation for MM that was originally intended in Japan as a **super-tough HBM** results from the absence of defined resistance, inductance or even the waveform[19]. The ESD-Association is working on an improved MM-standard, too. While in Japan MM may be dropped and replaced by HBM in near future, it seems to be mainly promoted by automobile industry were electrical overstress from low-impedant sources is common.

The ultra fast (<100ps) rising current transients make **CDM correlation** extremely difficult. Whether correlation can be achieved for CDM, depends on the design principles of the testers and their characterization.

The ESD-Association's CDM-draft standards distinguishes in first order by the type of discharge switch (ambient air=non contact "nc" or relay =contact "c*") and in second order by the fixturing (socketed s* or non-socketed ns)[16]. The commonly used classification "robotic" or "socketed" is traditional but of limited significance.

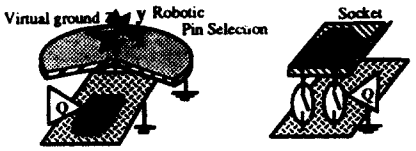


Fig. 5: Typical robotic (nc, ns) (left) und socketed (cs) CDM-tester (right) with charge stored on the device or the test board. Three further combinations of fixture and switch exist.

Comparative studies[20,21] demonstrate indeed a **fundamental lack of failure threshold voltage correlation** between the different design principles. The differences in failure voltage exceed the spread between devices tested on the same tester. Rather acceptable correlation is obtained in most cases for robotic (nc,ns) testers. In general, socketed testers and their individual test adapters dominate over the device related influences, in particular the device capacitance. This results in very reproducible discharge current waveforms in the whole stress voltage range. In (nc,ns) testers, the device should dominate the discharge (Fig.6), and the discharge current may be in-situ measured by means of a low-inductive, resistive current monitor[22]. However, robotic testers that establish the discharge via a closing air gap suffer from the spread of the discharge sparc for voltages above 1kV[20], if the current is not monitored.

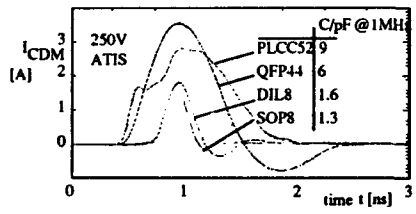


Fig. 6: Discharge current waveforms measured for different devices at 250V demonstrate the device influence in a (nc, ns) robotic tester. The "Dead Bug"-capacitance C @1MHz refers to the method explained in [10].

CDM-events and typical physical failure signatures gain importance with automatic handling and increasing device capacitances (TSOP, UTSOP)[16]. In our studies[10,11], they were reproduced by robotic as well as socketed testers. This work was and will be extended to further devices and testers[21]. As a rough rule, CDM withstand voltages **below 500V** indicate very sensitive devices that may fail during manufacturing or automatic handling. Up to now no CDM field failures have been seen for devices with withstand voltages exceeding 1500V and discharge currents of tens of Amperes in a robotic tester.

METROLOGY

The characterization of all ESD-testers and the in-situ study of stress parameters at a device demands a well-calibrated metrology for pre-charge voltage, capacitance and transient discharge current. Components of the metrology chain are the oscilloscope, the cables and attenuators, and last but not least the current probe. This may either be a current transducer for **all** systems employing a test socket or a low-inductive, coaxial resistor for non-contact CDM-testers[22]. Furtheron, we need calibrated pulse sources, test modules (capacitances) and load resistors to calibrate and verify the tester performance in terms of maximum current, rise time, decay time and ringing. The squares of the single component rise times sum up to approximately the square of the total system. The higher the system rise time is in comparison to the pulse rise time, the lower the measured peak current becomes. Therefore, the **lowest performing part limits the measured rise time and peak current performance of the total system! However, we have to distinguish the erroneously measured values from the actual device stress.** A step by step method for non-contact, non-socketed testers is suggested in Fig. 7 to 10 and further detailed in [10,11].

While for HBM and MM, a 350MHz (rise time <1ns) metrology chain should be sufficient, the spearhead of technology is reached capturing ultra-fast CDM-transients (rise time < 100ps). For CDM we should talk about rise times and real amplitudes and calibrate the total metrology step by step by means of adequate pulses, because of the possible non-linear characteristic of each system component. We should keep in mind that an error expressed by **only 1dB**, means an **error of 12%.**

Besides the weak correlation of CDM-tests, one **major acceptance problem** seems to be the investment and use of this most advanced metrology. Because we have to obey the laws of nature, the **final goal of research work** should be to establish - on a physical basis - a common agreement of the minimum, practicable effort required to identify weak designs by a correlatable method. However, this means to study first the complex interactions in detail before the limits should be loosened or tightened.

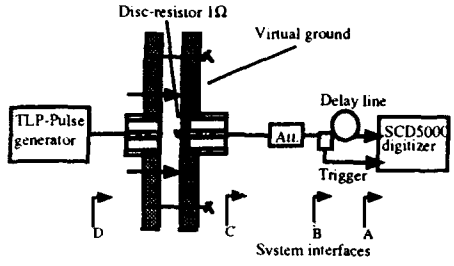


Fig. 7: Metrology chain for robotic CDM-tester with temporary used fully symmetrical 50Ω-adapter in front of the resitive current probe

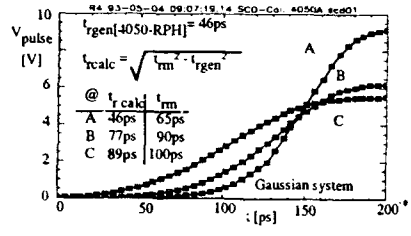


Fig. 8: Medium voltage characterization of the metrology at system interfaces A, B and C (fig. 7)

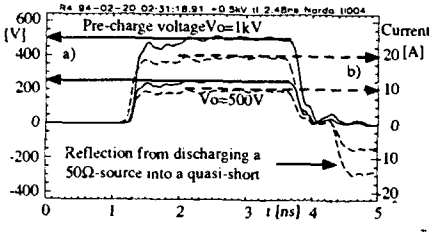


Fig. 9: TLP-pulse calibration of the resistive current probe at interface D of fig.8 in a face-to-face 50Ω-environment
*a) 50Ω-attenuator Vm=attenuation*Vel2*
b) 1.01Ω-current monitor

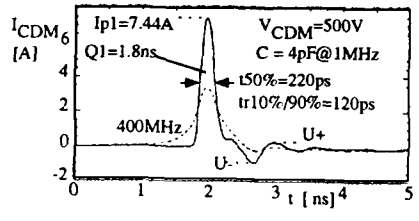


Fig. 10: Characterization of the performance of a CDM-tester by means of the discharge of a 4pF-test module employing a calibrated metrology chain. The dottet waveform estimates the pulse response of a 400MHz oscilloscope to this pulse. Pulse amplitudes up to 10A and risetimes below 100ps were measured under similar conditions.

ESD FAILURE MECHANISMS

An ESD FAILURE MECHANISM shall be considered as the ESD-induced process leading from a functional start condition to the physical failure signature. Inhomogeneities in current or field distribution become the seed for the physical failure signatures, as soon as certain stress thresholds are exceeded for a certain period. These inhomogenities result from process and topologic parameters, the individual layout of the (protection-) structure and the contacts between the different metallization layers.

In the first phase of the failure mechanism, local inhomogeneities increase locally the electrical field strength. The insulative properties of the reverse biased junction or the dielectric layer break down at these points first and current heats up the spot. The following mechanisms[12] may be distingished:

- **Power*time induced** failures mainly of the reverse biased **junction** according to the empirical Wunsch-Bell model
- **Contact spiking** (ETM electro thermo migration) which relates to the **energy** dissipated at a nearby junction or directly at the contact interface.
- **Electrical field strength*time** related dielectric breakdown

If the thermal carrier generation rate exceeds the doping level, thermal runaway occurs. At the external terminals we see a negative differential resistance. If *sufficent energy is dissipated in this volume, the* **second breakdown** with a melting of the junction follows. If aluminium without contact barrier is located close to this heated junction, at 525°C an eutectic Al-Si-spike may form and eventually cause a short to the substrate. However, while for pulses in the HBM and MM-domain the heat front travels about 3µm, it remains quasi-local in the adiabatic CDM-domain of few ns. In general, forward biased junctions are less critical, because, the heated volume is not confined to the depletion zone of the junction. The physical signature reaches from some small "notches" below the gate edge to significant melt channels.

Another power*time = energy related failure is the **fusing of metallization** in particular at topologic steps. However, there is no minimum breakdown threshold of the electrical field and a failure indicates that the metallization layout is weaker than the pn-structures in the path.

The **dielectric breakdown** is understood as the result of charge flowing through the oxide and partially being trapped there. If a critical amount of trapped charge is exceeded, the oxide breaks. However, this

time dependent dielectric breakdown theory seems to be disputable for the CDM-typical time domain of less than 1ns. We would like to suggest the following qualitative approach: A collective of electrons is trapped at the interface between silicon and oxide, preferably at points with thinner oxide (e.g. at the bird's beak). If the field strength becomes sufficently high, the traps are emptied at the same time. If they can build up an avalanche on their way through the oxide, the current increases exponentially. Additional energy, mainly stored on the gate capacitor, may melt a small conductive channel into the oxide. To obtain a permanent conduction, silicon from the gate or substrate must be molten in addition and driven by the electrical current through this channel. The conductive status of this filament may change in a wide range between open and short by means of consecutive electrical stress[23].

Systematic differences in the failure thresholds for the gate oxide breakdown of similar active structures on the same chip, however, with a different metallization layout could be an indication for **'antenna' effects**[24] in plasma processes.

We should generally keep in mind that the degree of inhomogeneity introduced by minor technological variations may have a significant influence on the failure threshold and, thus result in a band of failure thresholds on the same tester, which may even exceed the differences from tester to tester. However, **one goal of the protection scheme design is low sensitivity to such technological parameters.**

FAILURE CRITERIA AND THEIR THRESHOLDS

Electrical failure criteria are used to decide, whether the device is **fully operational, degraded** but still operational according to the parametric and functional specifications in the data sheet, or **faulty**. Typically, the thresholds may be set by the **data sheet** to **absolute values** or on application specific demands.

To test pins with **output** drivers they need first to be brought into a proper functional conditon (High-Z, High, or Low) before e.g. leakage or voltage levels can be checked.

The **power supply current** in the stand-by mode **Iddq** may increase or a series of test vectors are reqired to power the faulty circuit node first. However, V_{DD}/V_{SS}-leakage between guard rings not necessarily indicates a reliability problem, but may, e.g., limit the battery life time of mobile systems[25].

If thresholds that are more sensitive than the data sheet specifications are required, pure inputs without pull-ups or pull-downs may be tested for **absolute leakage thresholds**. Otherwise, a combination of absolute and **relative thresholds with respect to the virgin device** or a full parametric test has to be used.

The **reliability issue of low-level leakage or parametric degradation**, which may have different causes, is discussed very controversially and needs further detailed investigation. Possible explanations are:

- Charge trapping at the oxide interface
- Minor notches at the pn-junction or the gate edge
- Ruptured gate oxides

However, if **gate oxides are ruptured** - indicated by leakage currents in the order of 100nA@5.5V & FMA - and any additional electrical overstress may drive current through these conductive filaments, a possible **yield reduction** on system level or even a **reliability hazard** is indicated at least for inputs[11,23]. This minor level of leakage would currently not be considered as a problem for digital circuits, where leakage tolerances up to 10µA@7V may be found. In particular, CDM causes such degradations.
Leakage currents up to some µA often show a significant **"cold" and "warm" healing effect**[11,16,21]. Holes in the oxide remain, while trapped charges may detrap during this process and leave the device in a quasi-virgin state. Therefore, despite the practical problems, the devices should be tested as soon as possible after the ESD-stress.[16]

Low level leakage of protection or output transistors may even vary from stress pulse to stress pulse. Therefore, we may find several "passes" after "fails", before a catastrophic failure occurs in a step stress[26,27].

Virtual miscorrelation may result from different, however, not communicated definitions of the input voltage for leakage tests, the power and e.g. the chip eneable status. In general, devices can sustain higher stress if the input voltage is set to e.g. 0.6V instead of the maximum input voltage. However, setting the input voltage next to the breakdown voltage of the junction results in the highest sensitivity but may already pre-stress the protection element. Therefore, the standard[3,9,16] requires a test according to the full data sheet.

PROTECTION SCHEME

Protection elements shall provide a **fast turn-on, low-ohmic shunt path** for the discharge current to clamp dangerous voltages, before gate oxide or output transistor are damaged. The protection elements must also be able to **withstand sufficient energy** themselves. Low trigger and hold voltages support both. To reach these features, often different protection elements are combined into a protection structure. From the dis-

cusssion of the stress models, it is clear that HBM and MM are two-terminal (pin) events, while the "Real World"-CDM affects only one pin. The discharge current is on-chip conducted on the power rails. Therefore, the power rails are an important part of the total protection scheme. Their width should be maximized. Protection is also demanded between all power supply rails. We would like to call this system, including output transistors, a **"Protection scheme"**.

Over 200 patents (28 references in [12]) have been filed in this area, considering the following fundamental functional building blocks:

- **Protection elements intentionally active during EOS/ESD**
  (diode, zener-diode, thick-oxide transistor equivalent to a lateral npn, vertical npn, SCR)
- Capacitive suppression
- **Parasitic "protection" elements not intended to become active during ESD**
  (guard ring transistors and SCRs, output SCRs, internal paths)
- **Functional elements active during regular operation and EOS/ESD**
  (output transistor, pull-up, pull-down, decoupling resistor, power busses)

While highly efficient protection structure designs have been developed for inputs, which were traditionally most vulnerable to ESD due to their small dimensions and thin gate oxides, the **weak spots are now the outputs and the protection between the increasing number of power rails**[27,28]. Analog and ultra-fast pins have always been very vulnerable.

Internal parasitic "protection" elements or the metallization may fail at lower failures tresholds, because they are not designed to carry the amount of ESD or EOS current[28]. Alternatively, they may increase the hardness, especially for advanced CMOS and bipolar processes, if guard ring or output SCRs trigger under ESD-conditions. This effect may raise the failure threshold even if background capacitance is added (socketed CDM-testers)[10]. However, it is difficult if not impossible to predict, which of the possible structures triggers. Due to dV/dt-characteristics, the ESD-path may vary significantly from the less problematically accessible JEDEC-latch-up path. In advanced technologies, the lateral dimensions are shrunk and thus the probability to build unintended "protection" elements increases. A high susceptibility with respect to layout or process parameter jitter may thus be expected and we would like to recommend avoiding parasitic paths.

ESD-susceptibility and latch-up performance are in a close relation for bipolar and CMOS-technologies and SCR-structures. When the ESD-hardness is increased, the latch-up immunity drops.

The definition of withstand voltage goals for different stress models is obviously not a simple task and will be always disputable. However, it is needed to identify, or better avoid [illegible]. Many discussions with our industr[illegible] our own experience suggest that neither yield nor field returns are ESD-dominated for devices that withstand **2kV HBM, 300V MM, and 1kV non-contact, non-socketed CDM**. State-of-the-art ESD-precautions must be applied. However, this should not imply none of these devices could fail from ESD.

A **modular ESD-hardening approach**[29-32] helps to obtain the required ESD-performance by more systematic means. It employs an ESD-test chip that already accompanies process development. It starts with single protection elements and critical structures and ends-up with a full device-specific protection scheme.

Therefore, from first silicon in a new technology, **all possible basic ESD-protection elements, output transistors and potential parasitic structures** should be investigated for their high-current, turn-on and leakage characteristics (ESD and Latch-Up). Furtheron, the development of leakage should be evaluated in a reasonable stress current range from low-level up to maximum tolerable leakage or short. Step stress hardening effects have to be considered during this procedure. While the **high-current characteristic** is evaluated by means of transmission line pulsing[29], the **turn-on behaviour** for non-destructive repetitive pulses may be investigated by means of electro-optical sampling[33] and electron beam testing[34]. The tendency to current filamentation is indicated if the current does not linearily scale with the transistor width and finger count. The design flexibility of such structures should be increased in varying design parameters, ideally, in combination with the development of device and compact circuit simulation models[17]. At the present state-of-the-art, simulation allows to interpolate results to adapt the protection to specific needs. The remaining problem is the lateral current filamentation. This may call for larger safety margins in simulation.

The **second step** is to combine the elements to protection structures meeting the application specific parametric as well as ESD-demands and considering potential interferences between the single elements.

The **third step** is to arrange them at suitable positions and connect them via very low-ohmic power busses. A second test chip with simple functional blocks and typical power bus routing should be designed to double check the protection scheme.

This modular concept is valid to develop general as well as very specific design guidelines for **HBM, MM and CDM protection**. Well-protected devices withstand **CDM** peak currents of some tens of Amperes resulting from pre-charge voltages of 1.5kV and more, while less protected devices fail already at less than 2A and pre-charge voltages below 500V, which may be reached during manufacturing.

We are aware of earlier experience with bad transferability of ESD-single protection element performance to the ESD-susceptibility of functional devices. However, the recent findings of the system interactions and the advanced characterization techniques may again motivate this approach.

Great advances in the understanding and optimization of turn-on and current distribution **over the cross section** of the device may be expected from **transient electro-thermal device simulation**, if the parameters are extracted in the relevant current and time domains and lateral current filamentation by accident may be excluded. Considering output transistors, their snapback trigger and hold voltage may be optimized with respect to hot carrier degradation.[27,32]

Circuit simulation may assist to optimize the protection scheme. The turn-on of complex structures and the resulting power distribution as well as the system interactions between ESD-tester and device are studied [17,18]. However, significant effort in modeling and parameter extraction is still needed to reach the final goal of a universal design tool for application specific ESD-protection.

APPLICATION

The application may introduce special limitations to leakage current e.g. <1nA for analog pins. Reference pins may not tolerate significant series resistance. A demanded current and VLO defines the maximum output resistance. This is the series resistance of the output transistor and the optional decoupling resistance, which shall ensure the major current flow through an optional additional protection element[27]. Decoupling resistors are used in input protections as well. However, system considerations in particular speed may limit the RC-constants of the output as well as of the input.

TECHNOLOGY

The traditional myth that MOS-devices are much more sensitive to ESD than bipolar devices, is not attributable to a specific author and has lost its value completely for the small structures of advanced bipolar technologies.

Each technology provides different opportunities to realize protection elements. In most cases "parasitic" effects such as bipolar action of an NMOS-transistor or triggering of an SCR in a CMOS-technology have been empirically "cultivated" to obtain an optimum protection performance. When a constant energy to failure and a constant current are assumed, the high-current voltage drop across the element determines its withstand time.

Two major process innovations have aimed at improved hot electron reliability and speed, but have **increased ESD-problems significantly**, supporting lateral current filamentation. These are lightly doped drains (LDD) and silicided junctions.

For the **LDD-output and protection transistors**, the a 3...5µm spacing between drain contacts and gate edge has improved the lateral homogeneity of the current below the gate edge. Additional deeper implantations in the peripheral circuits have optimized the drain profile for ESD[27,32].

The **silicided junctions** of sub-micron technologies, where the RC-constants of metallization plays an important role, eliminate this additional drain resistance. The result are unpredictably low failure thresholds. An additional mask may be used to block the salicide in the output range[27].

A local compensation of the process step that is needed for the core of the circuit is an expensive solution. However, in combination with an EPI-substrate, protection elements with a vertical current flow may gain advantage over lateral devices.

The influence on future technologies has been discussed by D.Lin[36], S.Voldman[37]. In his extrapolation to channel length 0.1µm, junction depth 0.05µm and 4nm oxide thickness, Lin concludes that a factor of 10 shrinkage in feature size makes pn-junctions 5 times and gate oxides 10 times more sensitive to electrical stresses like ESD. Voldman investigated a double diode ESD-protection circuit for 1.2µm, 0.7µm, 0.5µm and 0.25µm CMOS-technology on EPI-substrate. Technologies below 1µm were fully salicided. The ESD-suceptibility dropped for his technology benchmarking ESD-protection circuit from 9kV to 3kV HBM. Both authors employed a combination of experiment and numerical simulation for their forecast.

A common view is that thinner gate oxides and smaller junction dimensions lead to an increased ESD-sensitivity.

STATISTICAL ASPECT

A statistical aspect of ESD-susceptibility associated with pin count and package size should be considered, too.

For present processors and application specific logic devices, the number of **multi purpose pins** including output structures increases to above 512. When the natural spread of the failure thresholds is considered, the tested **HBM-failure class** may drop for this specific device.

However, from a system point of view, the total number of **HBM-ESD-failures** per system assembly may drop due to the reduced number of components and manual handling steps.

Nevertheless, device capacitances increase with larger and thinner packages while the feature sizes are scaled. Together with an increased degree of automation, this may lead to an increased number and financial value of **CDM-degraded devices.**

CONCLUSIONS

- ESD damage is a result of complex interactions between the tester (environment) and the device, in particular, its protection scheme, leading to problems of correlation and a possible misinterpretation of the results.
- The process of **ESD-hardening and ESD-testing** should **not** be seen as the **"pig tail" of chip design**, because of its evidently close relation to all development steps. The earlier the different experts are involved in the process and the better the exchange of information is, the more reliable and the less expensive the final solution should be. One important point is the development of the common glossary.
- **Discharge current** should be regarded as the major stressor in **all** stress models HBM, MM and CDM. **Voltage thresholds** may become disputable, if the defined current is not forced by the impedance of the tester (e.g. $R_{HBM} = 1.5k\Omega$).
- **ESD-tests in, at least, two time and current domains** - fast transient, medium current **HBM and** ultra-fast transient, high-current **CDM** - on the basis of existing standards are urgently recommended **to identify all ESD-weak designs.**
- In particular for CDM, reliable and testable **failure criteria** are mandatory to distinguish between reliability relevant and non-relevant low level leakage.
- **Outputs, input-outputs and analog pins** tend to **limit** the ESD-susceptibility of modern devices.
- Major problems are **large spreads of failure thresholds associated with NMOS-thin oxide devices** resulting from current filamentation due to inhomogeneities. Silicided junctions reducing the drain resistance, as well as lightly doped drains (LDD) increase the tendency towards current filamentation.
- **In advanced technologies, protection elements** may reach lower clamping voltage levels that reduce power dissipation in the device and stress across the gate oxides. However, junctions and gate-oxides become more sensitive, too.
- **Advanced protection elements are fast enough to achieve sufficent CDM-protection.**
- A **modular ESD-hardening approach** starts with the systematic design and experimental evaluation (high-current, turn-on) of single protection and other peripheral elements in the relevant domain and ends with a device and application specific protection scheme. Step by step, guide lines are improved for further designs.
- State-of-the-art calibrated **device and circuit simulation** should support the experimental work to reduce the number of design variations and to study the interaction between device and tester for the different stress models supporting the standardization of ESD-testers.
- To achieve **correlation of ESD-tests,** we need **criteria for the characterization of ESD-testers with an optimum selectivity range** that is small enough to ensure correlation for different protection schemes, but sufficently wide to allow competitive test system approaches.
- **CDM-current metrology** is at the spearhead of the present technology. Therefore, CDM-tester characterization becomes a real challenge.
- The improvement of the **metrology** and the **modelling of the fast and ultra-fast breakdown mechanisms of junctions and dielectrics** are mandatory to avoid an ESD-limited yield for ULSI-devices and sub-systems.
- The **physics of ESD and semiconductors** is the same all over the world, therefore, we are convinced that further focused international cooperation and research is the only way to establish **international standards** serving the needs of customers as well as device and tester manufacturers.

ACKNOWLEDGEMENTS

We would like to thank the members of the ATIS-team FhG-IFT and at TUM in alphabetical order Peter Egger, Dr. Helmut Haudek, Rainer Kropf, Christian Mußhoff, Christian Russ, Peter Scheubert and Henry Wolf, as well as our partners Dr. Les Avery (David Sarnoff Research Center), Xaver Guggenmos (SIEMENS AG), Dr. Fred G. Kuper (PHILIPS SEMICONDUCTORS), Koen Verhaege (imec) and the members of the Sub-committee "Device Testing" of the ESD-Association for their excellent cooperation and the fruitful discussions. We further would like to thank Dr. Konrad Hieber, Dr. Günther Röska and Dr. Christian

Diekmann and their teams at FhG IFT for creating the pre-requisits for continuing ATIS contributions in the field of ESD.

REFERENCES

1. H.E. Maes (editor), "Survey of Present Status on Critical Reliability Issues within Europe", Final Report of the Technical Interest Group on IC-Reliability, 1992
2. National Bureau of Mines bulletin 520, 1962 Source: 12
3. EOS/ESD-Association, "EOS/ESD-S5.2: For ELECTROSTATIC DISCHARGE (ESD) SENSITIVITY TESTING - MACHINE MODEL (MM) - COMPONENT LEVEL", June, 20,1994
4. T.S.Speakman, "A Model for the Failure of Bipolar Silicon Integrated Circuits Subjected to Electrostatic Discharge", Proc. of the IRPS, 1974
5. P.R. Bossard, R.G.Chemelli, B.A.Unger, "ESD Damage from Triboelectrically Charged IC-Pins", Proc. of the 2nd. EOS/ESD-Symposium, EOS-2, 1980, pp.17-22
6. "MIL-Std883C Method 3015.7 notice 8", Military Standard for Test Methods and Procedures for Microelectronics: ESD Sensitivity Classification,March 22, 1989
7. L. van Roozendaal, et al., "Standard ESD Testing of Integrated Circuits", Proc. of the 12th EOS/ESD-Symposium, EOS-12, pp.119-130, 1990
8. K. Verhaege, et al. "Analysis of HBM ESD Testers and Specifications Using a 4th order Lumped Element Model", Proc. of the 15th EOS/ESD-Symposium, EOS-15, pp.129-139, 1993
9. ESD-Association -S5.1-1991, "Standard for ESD-SENSITIVITY TESTING- HBM -COMPONENT LEVEL", June, 6,1991, 200 Liberty Plaza, Rome, N.Y. 13440
10. H.Gieser, et al. "Influence of Tester Parasitics on "Charged Device Model" Failure Thresholds", Proc. of the 16th EOS/ESD-Symposium, EOS-16, 1994
11. H.Gieser, et al., "A CDM-only Reproducible Field Degradation and its Reliability Aspect", Proc. of the 6th ESREF-Conf., pp.371-378,1993 and selected to be published in a special issue of QREL, 1994
12. O.McAteer, "Electrostatic discharge control", McGraw-Hill Publ., New York, 1989
13. Weber, Private Communication, 1991
14. R.Merrill, et al., "ESD Design Methodology", Proc. of the 15th EOS/ESD-Symposium, EOS-15, pp.233-237, 1993
15. I.Morgan, "Identifying EOS/ESD Failure Signatures", Tagungsband 3. ESD-Forum (ISBN 3-9802845-2-2), Grainau, 1993
16. ESD-Association, "EOS/ESD-DS5.3: For ELECTROSTATIC DISCHARGE (ESD) SENSITIVITY TESTING - CHARGED DEVICE MODEL (CDM)- Component Level"
17. C. Russ, et al..,"Compact Electro-Thermal Simulation of ESD-Protection Elements", Proc. of the 4th ESREF, 1993, pp.395-400
18. C.Russ, et al. ,"ESD-Protection Elements During HBM-Stress Tests- Further Numerical and Experimental Results", Proc. of the 5th ESREF, 1994
19. EIAJ-Standard IC-121-1988: "Testing Method 20: Electrostatic Destructive Test, Electrostatic Discharges, ESD"
20. H.Gieser, et al., "CDM-Tester und ihre Standardisierung", Tagungsband 3. ESD-Forum (ISBN 3-9802845-2-2), Grainau, 1993
21. K. Verhaege, et al.,"Influence of tester, test method and device type on CDM ESD testing", Proc. of the 16th EOS/ESD-Symposium, EOS-16, 1994
22. R.G. Renninger, et al.," A Field Induced Charged Device Model Simulator", Proc. of the 11th.EOS/ESD-Symposium, EOS-11, 1989, pp.59-71
23. J.C.Reiner,"Investigation of Latent Gate Oxide Damage Caused by CDM-ESD", Proc. of the 5th ESREF, 1994
24. S.U.Kim,"ESD Induced Gate Oxide Damage during Wafer Fabrication Process", Proc. of the 14th.EOS/ESD-Symposium, EOS-14, 1992, pp.99-105
25. L.R. Avery, Private Communication
26. F.G.Kuper, et al."Suppression of Soft Failures in a Submicron CMOS Process", Proc. of the 15th EOS/ESD-Symposium, EOS-15, pp.117-122, 1993
27. X.Guggenmos,"ESD-Optimierung von hochintegrierten CMOS-Schaltungen - Design- und/oder Prozeßmaßnahmen", Tagungsband 3. ESD-Forum (ISBN 3-9802845-2-2), Grainau, 1993
28. C.C.Johnson,et al. "Two unusual HBM ESD Failure Mechanisms on a Mature CMOS Process",Proc. of the 15th EOS/ESD-Symposium, EOS-15, pp.225-231, 1993
29. T.Maloney, et al. "Transmission Line Pulsing Techniques for Circuit Modeling of ESD-Phenomena",Proc. of the 7th.EOS/ESD-Symposium, EOS-7, 1985, pp.49-54
30. A.Amerasekera, et al."An Investigation of BiCMOS ESD Protection Circuit Elements and Applications in Submicron Technologies", Proc. of the 15th EOS/ESD-Symposium, EOS-15, pp.265-276,1993
31. D.Krakauer, et al.,"ESD-Protection in a 3.3V Sub-Micron Silicided CMOS Technology", Proc. of the 15th EOS/ESD-Symposium, EOS-15, pp.250-257,1993
32. C.Diaz, et al., "Studies of Susceptibility in 0.6µm nMOS ESD I/O Protection Structures", Proc. of the 15th. EOS/ESD-Symposium, EOS-15, 1993, pp.83-91
33. K. de Kort, et al., "The transient behaviour of an ESD input protection", Proc. of the EOBT 1993, Zürich, pp.7.15-7.18
34. R.Kropf, et al., "Zeitaufgelöste Untersuchung des Snapback-Verhaltens eines ESD-Schutztransistors", Tagungsband des 3.ESD-Forum, Grainau, 1993
35. C.Diaz, et al., "Studies of Susceptibility in 0.6µm nMOS ESD I/O Protection Structures", Proc. of the 15th. EOS/ESD-Symposium, EOS-15, 1993, pp.83-91
36. D.Lin,"ESD Sensitivity and VLSI Technology Trends: Thermal Breakdown and Dielectric Breakdown", Proc. of the 15th. EOS/ESD-Symposium, EOS-15, 1993, pp.73-82
37. S.Voldman, et al." Scaling, Optimization and Design Considerations of Electrostatic Discharge Protection Circuits in CMOS Technology", Proc. of the 15th. EOS/ESD-Symposium, EOS-15, 1993, pp.251-260

ESD PROTECTION ELEMENTS DURING HBM STRESS TESTS — FURTHER NUMERICAL AND EXPERIMENTAL RESULTS

Christian Russ
Chair for Integrated Circuits, Technical University of Munich (TUM)
Arcisstrasse 21, D-80290 München, Germany
Phone: + 49 89 231817 52, Fax: + 49 89 231817 50

Horst Gieser
Fraunhofer-Institute for Solid State Technology (IFT)
Hansastrasse 27D, D-80686 München, Germany
Phone: + 49 89 54759 020, Fax: + 49 89 54759 100

Koen Verhaege
IMEC
Kapeldreef 75, B-3001 Leuven, Belgium
Tel: + 32 16 281 299, Fax: + 32 16 281 501

ABSTRACT

Correlation problems for HBM-ESD testing result from the complex interaction between device and tester. The *HBM* stress of *different* well-characterized testers [1] is applied to protection elements. By means of circuit simulations and in-situ measurements, snapback and second breakdown during HBM are investigated. For fast transient events, a new transmission line approach of the tester improves the correlation between experiment and simulation.

1. Introduction

It is not sufficient to know the HBM failure thresholds of certain protection devices to meet the demands for good ESD performance of a current technology. The major demands are fast turn-on behaviour, effective voltage clamping, current carrying and high ESD hardness. However, we also need to know if e.g. a protection structure with a high HBM ESD threshold provides a good turn-on behaviour as well. Even for the HBM time domain, "conservative" with respect to CDM, only the knowledge of the transient behaviour of a protection device may confirm the ideas of voltage clamping capability, switching behaviour and failure models.

It is the objective of this paper to study the behaviour of protection devices during HBM stress and the interaction between HBM tester and ESD-protection devices.

First we consider and characterize the HBM tester as an isolated system. Then compact models for protection elements are presented. A correlation study is performed to point out the influence of tester parasitics [3], [4] on the HBM thresholds of real devices. For a deeper insight into HBM testers, we present a setup for in-situ current and voltage measurements and its calibration. This setup is used to investigate the transient behaviour of protection diodes and transistors during HBM stress. The results are confirmed by numerical circuit simulations.

2. Advanced characterization of HBM-ESD testers

2.1 The 4th order lumped element model

A lumped element model as an equivalent for HBM ESD testers (fig. 1) was presented by L. van Roozendaal et al. [3]. This 4th order lumped element model (LEM) was solved analytically and an analytical fitting method was developed by K. Verhaege et al. [4] to extract the element values from measured current pulse data (short and 500Ω load). Deviations between the *nominal* precharge voltage and the actual peak current, from which the 4th order LEM extraction method re-calculates an *effective* precharge voltage, are considered by a correction factor k.
Recommendations to improve tester specifications suppressing the main parasitic - the test board capacitance - were also given in [4].

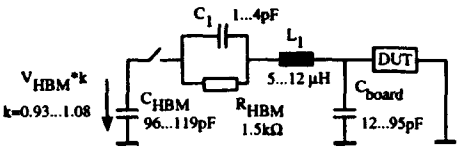


Fig. 1: The HBM-ESD tester equivalent containing the HBM elements and the main tester parasitics [3], [4].

The stability and the convergence behaviour of the 4th order LEM method were proved with a study of reduced observation periods of the measured pulses. Figure 2 shows the relative deviations of the element values related to their final values for different periods. It reveals clearly that it is sufficient to consider an observation period of the pulse of not more than 200ns.
The method is now used as a tool for detailed HBM tester characterization.

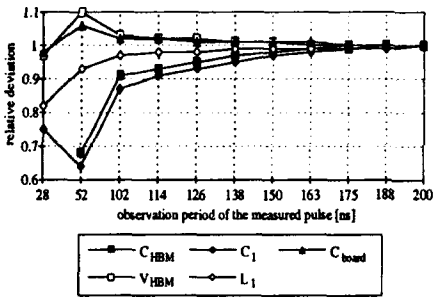


Fig. 2: Convergence behaviour of the different lumped elements. Sufficient accuracy is achieved for an observation period of about 200ns.

2.2 Influence of the test board capacitance

The amount of the test board capacitance, one of the most important parasitic tester elements related to snapback devices, was studied at one certain tester (refering to T3 in [4]) for different pin combinations of a DIL48 socket. The results of the dynamic characterization method using the 4th order LEM were compared to a method, where the test board capacitance was determined from the static charge by means of a Keithley Model 617 system (fig. 3).
The test board capacitance is split in two parts, A and B. Part A results from the motherboard of the tester, part B from the circuitry behind the relevant relay S of the switching matrix. However, since S cannot be closed regularly during the static measurement, part B can only be addressed via the common ground, which means that pin A and B must not be changed.

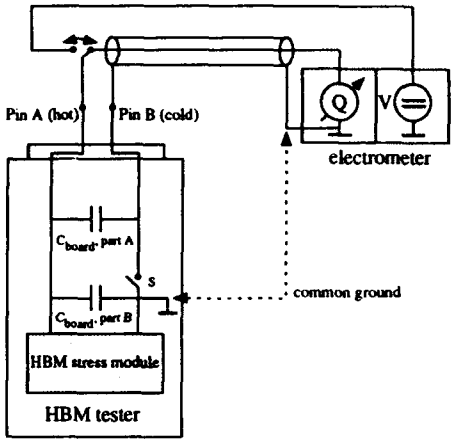


Fig. 3: Measurement setup for the determination of the static test board capacitance.

The measurement was performed by, first, charging up the test board to a known voltage (V), second, discharging it into the calibrated, internal capacitor of the electrometer to measure the total charge stored on the board (Q). Then the capacitance could be determined easily given the relation C = Q/V. The results are depicted in figure 4. It reveals clearly that the test board capacitance has to be considered as dynamic and effective for HBM stress. This effective capacitance is only valid in the HBM time domain but thus very valuable. The capacitances gained by the static method were distinctly higher. However, the correlation among characteristic pin combinations due to different on-board wiring was confirmed.

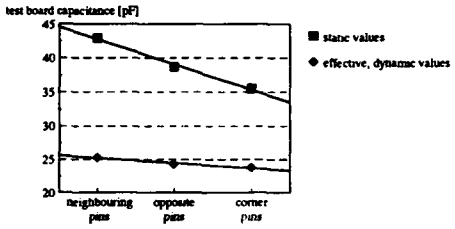


Fig. 4: Test board capacitances gained by the 4th order LEM method [4] for the HBM-ESD relevant time domain in comparison to a static method.

Depending on the type of tester, its particular test board and its capacitances, a detailed investigation of characteristic pin combinations should be taken into consideration. However, for the tester data of figure 4, all distinctions in the capacitances extracted by the 4th order LEM method are within a well-acceptable spread.

2.3 Tester characterization for different precharge voltages

For a final characterization, the dependence was measured between the *nominal* and the *effective* precharge voltage recalculated by the actual pulse (fig. 5). For the 1kV standard calibration pulse, the factor k was determined by 0.96 and also considered in the numerical simulations. As depicted in figure 5, there are only some slight deviations in the region above 1.5kV. The *actual* DC-precharge voltage of the HBM capacitor is almost slightly higher than the *nominal* precharge voltage. Therefore, losses in the tester due to electromagnetic radiation in the discharge path are considered. By means of the difference between the *actual* and the *effective* precharge voltage, these losses were determined up to 7%.

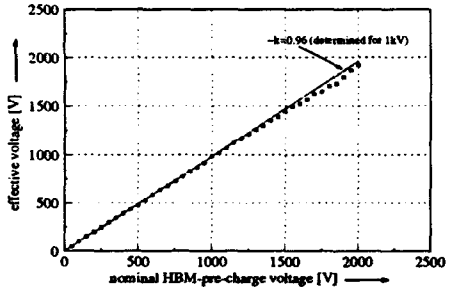


Fig. 5: Correlation between nominal HBM precharge voltage and voltage extracted by HBM pulse measurements.

3. Improved compact models for single protection elements

Compact electro-thermal models were developed for single protection diodes and snapback nMOSFETs [5], [6] to use within the circuit simulator SABER[1].
The non-linear high current IV-characteristics of the protection elements were gained by rectangular pulses with an automated transmission line pulser [7] setup (TLP) at quasi-static conditions.
Efforts were made for improved numerical stability in the calculation of the highly non-linear IV-characteristics. The usage of the models with regard to parameter input and output of internal model quantities was improved. Now the models also contain the ability for scaling some relevant device geometry parameters aiming at the developement of a tool for design optimization.
The models contain a behavioural description of the failure due to second breakdown.
Both the diode and the transistor model were verified successfully [5] with HBM experiments for the a stable industrial technology (1.0µm CMOS) considered in this paper, too.

3.1 Protection diode

The high current IV-characteristics of reverse biased protection diodes differ strongly from the static breakdown voltages. We achieved a good correlation with the TLP measurements (fig. 6) by describing the behaviour of the diode in the avalanche conducting regime ($V_d > V_b$, V_b breakdown voltage) with an empirical approach [6]. It is assumed that the slope of the IV-characteristic dV_d/dI_d is a linear function of the voltage given by

$$\frac{dV_d}{dI_d} = R_1 + \Delta R \cdot \left(V_d - V_b\right)$$

with

$$\Delta R = \frac{R_2 - R_1}{V_{t2} - V_b} \quad \text{for} \quad V_d \geq V_b \ .$$

Thus the current through the diode can be described by

$$I_d = \frac{\ln\left(\dfrac{R_1}{R_1 - \Delta R\left(V_d - V_b\right)}\right)}{\Delta R} \ .$$

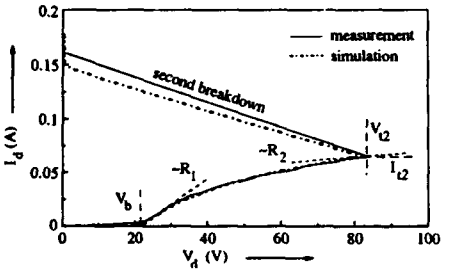


Fig. 6: IV-characteristic and modeling of a typical ESD protection diode.

[1] TM by Analogy Inc.

3.2 Protection transistor

The transistor model describes the static and dynamic behaviour of parasitic bipolar transistors utilized as ESD protection elements. The IV-characteristics were studied by means of TLP measurements (fig. 7). Thus it contains both turn-on mechanisms, the snapback phenomenon and the dV/dt-triggering effect which was verified experimentally by fast transient measurements with an electron beam prober by R. Kropf et al. [8] and by means of electro-optical sampling techniques by J.R.M. Luchies et al. [9].

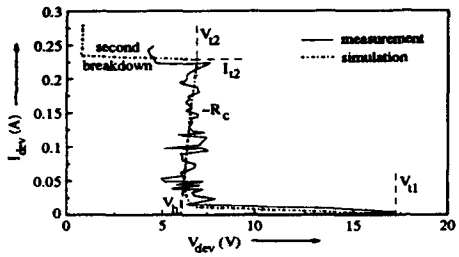


Fig. 7: IV-characteristic and modeling of a typical ESD protection nMOSFET.

4. Comparison of HBM-ESD testers: influence of the test board capacitance

A set of single ESD protection p^+pn^+ diodes was used to demonstrate the considerable influence of the test board capacitance on HBM failure thresholds.

In general, this influence may appear in very different ways:

1. There are integrated circuits which do not show any influence at all [4]. The results of all testers correlate. This may occur, e.g. for relatively low resistive diodes with a good voltage clamping behaviour. The damping effect of the pulse via the test board capacitance is neglectible.
2. The *higher* the test board capacitance values are, the *lower* the HBM thresholds of snapback transistors are due to additional stress caused by the sudden discharge of that capacitance at snapback [4].
3. The *higher* the test board capacitance values are, the *higher* the HBM thresholds are due to the damping of the HBM pulses. Sensitive devices are e.g. relatively high resistive diodes enabling the damping effect of the test board capacitance.
4. The deformation and damping of the pulse by the test board capacitance could address another failure mechanism. However, this speculation still has to be proved.

Four devices of each diode type were stressed in three different HBM testers with a wide variety of test board capacitances. The applied failure criterion was 1µA@5V. As was found by previous HBM stress tests, the spreads of the HBM failure thresholds of each device type were little. This ensured that already four devices gave significant results.

Figure 8 depicts the experimental HBM failure thresholds and compares them to the HBM results simulated numerically. The numerical simulations taking into account the entire test system [10] were performed with all relevant tester parasitics determined by [4]. The simulations confirm this significant dependence of the HBM threshold on the test board capacitance, although the simulated results were lower than the measured HBM thresholds.

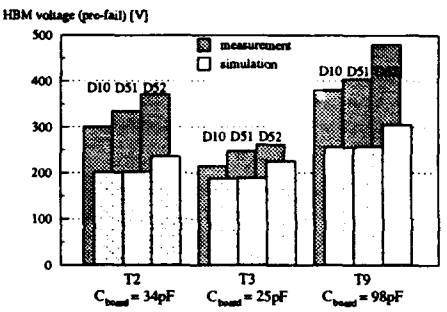


Fig. 8: HBM results for a set of protection diodes. (The tester numbers T2, T3, T9 refer to [4])

We attributed the lower simulated thresholds to the selection of the failure criterion in the numerical simulation. Failure parameters (**power to failure**) could be taken only from 100ns transmission line pulses. Since, for the setup used, it was impossible to determine the relevant parameters from a power profile [11], [12], [13], it did not seem efficient to apply the original Wunsch-Bell-model [14] for a fit of the failure parameters.

Therefore, we decided on the following approach: We now compare the energy dissipated in the device up to 100ns during an HBM pulse with that dissipated during the 100ns transmission line prefail pulse. This gives us a rough estimation, where the relevant failure parameter might be somewhere inbetween energy and power.

Promising results are depicted in figure 9. The trend towards the dependence on the test board capacitance is confirmed here very well. The HBM thresholds

achieved by the new failure criterion point clearly toward the experimental results. Moreover the relative differences between the failure thresholds of different devices tested on the same tester become clearer.

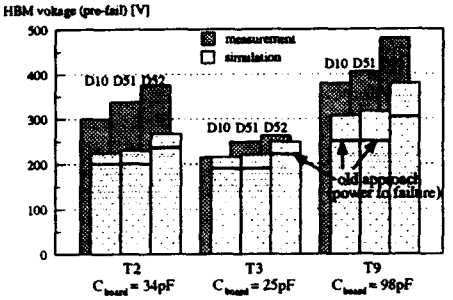


Fig. 9: Simulation using a new failure criterion with the objective to give a zero order approximation with respect to the Wunsch-Bell-model [14] (the simulated thresholds of figure 8 are shown well).

Future work will be done with an extended setup of the transmission line pulser which enables us to gain power profiles by a set of different pulse lengths. A waveform conversion algorithm, which was presented by D. Pierce in [12] yielding a better comparability, will be implemented in the compact models, too.

5. Interaction between HBM tester and ESD protection devices

5.1 Measurement setup

As shown in the previous chapter, the HBM failure thresholds of the protection devices investigated on different HBM testers were dominated by test board capacitance, one of the main tester parasitics. For a more detailed insight into the transient behaviour of the complex interaction between HBM tester and ESD protection devices, an approach was made to measure both the current and the voltage during HBM stress pulses simultaneously [15], [16].

As is well-known, particular in-situ voltage measurements may suffer from electro-magnetic stray fields coupled into the ground loop and a voltage drop on the shield of the cable to the scope [17]. They result in a low signal-to-noise ratio. Therefore, we developed a setup minimizing the parasitics (fig. 10).

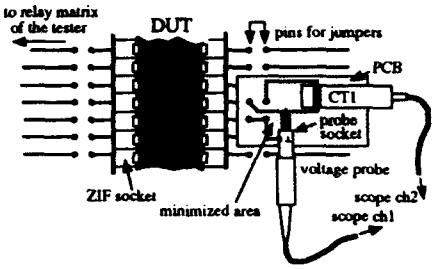


Fig. 10: Setup for the in-situ measurement of HBM stress.

The current transients could be measured easily by a Tektronix CT1 current probe inserted close to the device in the discharge path. An additional inductance could be neglected due to the very small wire loop. The high resistance voltage probe (HP 10440A, 100:1, 10MΩ/2.5pF) was located close to the device, too. The ground lead inductance of the probe was minimized using a special probe socket instead of the probe ground lead.

The setup is calibrated with current and voltage waveforms with respect to a 500Ω load. The same low inductance 500Ω CADDOCK type MG precision resistor as for the 4th order LEM method [4] was used. Current and voltage transients were recorded by the 2*500 MHz HP54111 digitizing oscilloscope (fig. 11).

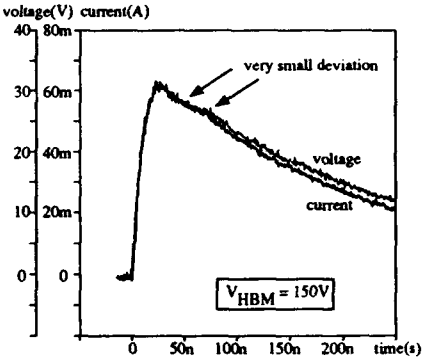


Fig. 11: Pulse forms of a 500 ohm low inductance resistor for characterizing the setup.

From the "current vs. time" and the "voltage vs. time" data, a plot "current vs. voltage" is derived and depicted in figure 12. It shows good accordance to the ideal 500-ohm slope. However, the currents (fig. 11) tend to droop due to a measurement artefact of the CT1 acting as a transducer. This leads to the current decreasing faster than the voltage.

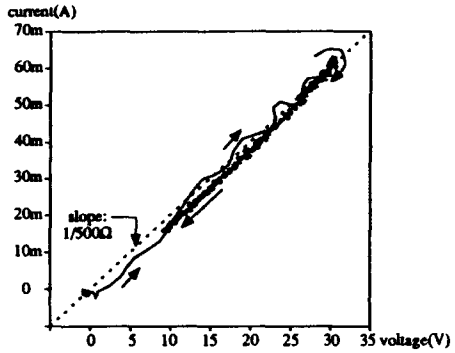


Fig. 12: IV-plot from the waveforms of figure 11 demonstrating the accuracy of the setup.

The pulses in the HBM tester used for these investigations are tuned by a resistance and an inductance in parallel at each pin in order to match the standards [1], [2] (fig. 13). These pulse forming elements in the ground path were shortened via the probe ground lead, the probe shield and the scope ground causing some very small deviations with respect to a reference pulse (fig. 11).

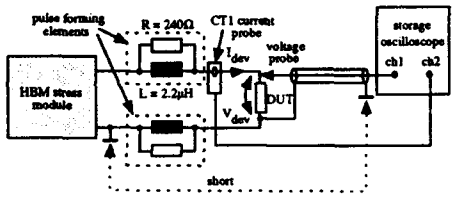


Fig. 13: Circuitry of the motherboard of the HBM-ESD tester.

The characterization of the setup demonstrates that not only qualitative but also quantitative statements can be obtained. For low ohmic resistors of 10Ω and less (and thus also for the on-resistances of protection devices) the signal to noise ratio was acceptable. Therefore, a voltage resolution of better than 1V was achieved.
The development of a differential signal probe will be considered in future work with respect to the investigation of the rising edges of HBM voltage transients.

5.2 Snapback devices under HBM stress

Non-destructive HBM pulses were applied to a snapback nMOSFET (W/L: 122μm/1.0μm) and recorded by the setup described in chapter 5.1. The high current IV-behaviour by means of TLP measurements of this device was presented in figure 7, chapter 3.1. The voltage and current transients of a 200V HBM pulse applied to this device are shown in fig. 14. Snapback events occur at the beginning and as well at the end of the pulse.
The device was triggered within a few nanoseconds after the pulse started. The triggering time was determined by the RC-time constant of the tester and the input capacitance in conjunction with the turn-on characteristics of the device. Immediately, the voltage dropped from breakdown voltage v_{t1} (plus overshoot due to ringing) to the hold voltage v_h [5], [8]. This snapback caused a discharge of the test board capacitance [4], which occured with a fast current during the rising edge of the HBM pulse. Depending on the difference between breakdown and hold voltage and the amount of test board capacitance, the sudden discharge could cause additional stress for the device.

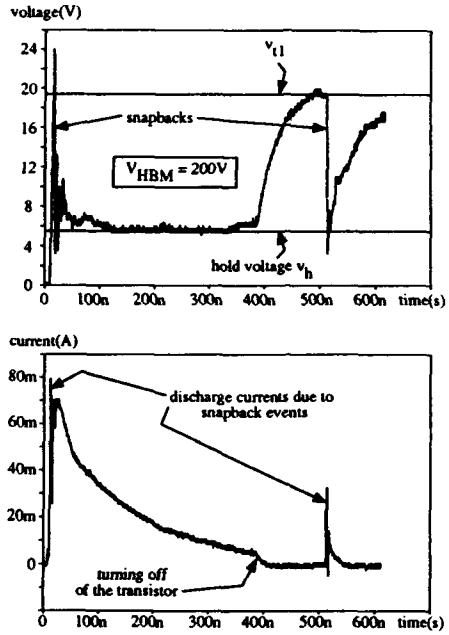


Fig. 14: Transients of an nMOSFET during HBM stress: snapback events also occur at the end of the pulse.

For different HBM precharge voltages, different currents were forced in the device (fig. 15). Fast current peaks occurred all during the HBM pulses. However, the lower the HBM pulses were, the more crucial the discharge peaks were. For HBM precharge voltages equal or less than 200V, the snapback current peaks obviously exceeded the proper HBM current maximum. This may cause a severe problem if the test board capacitance is large.

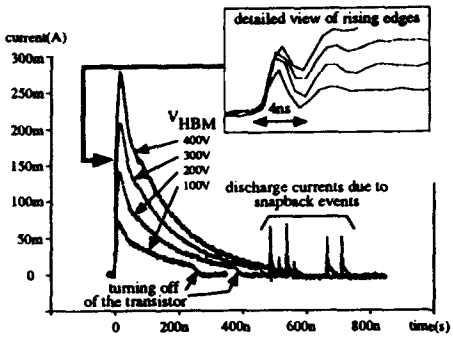


Fig. 15: Current transients of an nMOSFET during different HBM stress levels.

At the end of the HBM pulse, when the current decreased below a certain hold current, the device switched off (figs. 14, 15). It changed from bipolar mode to blocking state. The remaining weak current of the HBM pulse slowly charged up the test board capacitance until the triggering voltage for snapback was reached again (fig. 14). The shape of the charging characteristic was dominated by R_{HBM} and C_{HBM}.
As a consequence, not only the hold voltage v_h, but voltages up to the breakdown voltage v_{t1} would be applied to the device. Relatively high and fast current peaks appear several times due to the low resistive discharge of the test board capacitance. The consequences this may have on the device before it returns to its thermal equilibrium are unkown.

5.3 In-situ failure detection during HBM stress

The measurement setup used for nMOSFETs is then applied to devices for destructive HBM stress. In figure 16 the voltage and current transients for an ESD protection diode are depicted.
The curves for a precharge voltage of 200V represent the case of "moderate" stress and for 260V the pre-fail case. When 270V was applied to the device the occurrence of second breakdown could be detected in-situ!
As predicted by former circuit simulations [4] - [6], a powerful discharge current supplied by the test board capacitance could be observed. This additional stress might even enlarge the electrical or physical failure signature.
Ringing in the transients appeared only for the second breakdown case due to the fast changes in the current.
The failure threshold of this diode was compared to the thresholds determined previously for the same type of device. First, the devices showed only a very little spread in their HBM thresholds in general. Second, the HBM threshold of the device taken for the in-situ HBM failure detection was exactly within this spread. Therefore, it can be concluded that the measurement setup did not influence the HBM threshold.

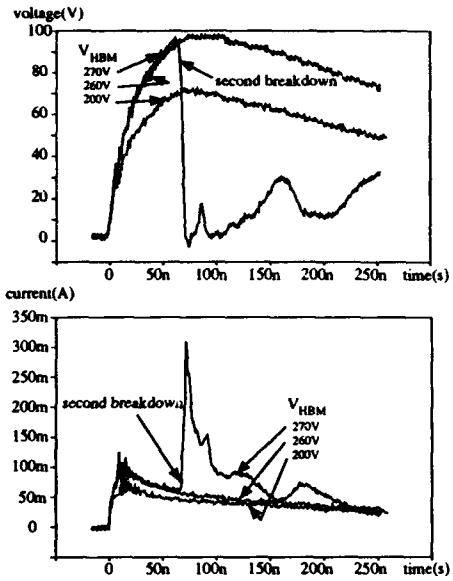


Fig. 16: In-situ failure detection of an ESD protection diode under HBM stress: second breakdown causes a voltage drop and a discharge of the test board capacitance.

6. Approach for an HBM-ESD tester model related to fast transient events

6.1 Transmission line approach

For circuit simulations of the entire test system, the 4th order lumped element tester model (fig. 1) was combined with both the snapback transistor model (fig. 7) and the diode model (fig. 6) [10].
Since in the 4th order LEM the test board capacitance is treated as a concentrated element, it may not be sufficient for describing fast transient interactions between tester and device. Fast transient events may occur when e.g. a protection transistor goes into snapback or the voltage across a protection device suddenly drops due to second breakdown.
For this time domain, we developed a tester model with a transmission line (TL) approach. It uses distributed elements for the test board capacitance and partly for the serial inductance. The estimation of the

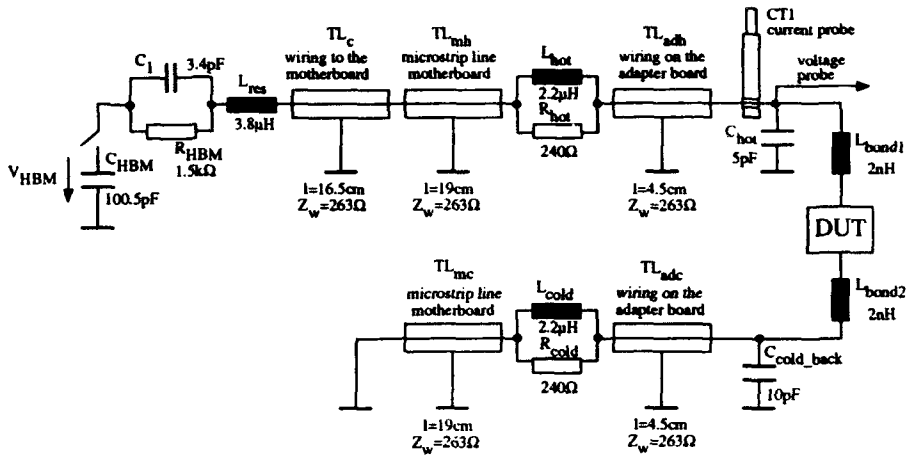


Fig. 17: Transmission line approach of an HBM tester

line impedance and its capacitance and inductance per length was performed by the assumption that the impedance is constant over the entire signal path. The wiring in a real HBM tester, of course, is a complex three-dimensional line structure. The impedance was determined by a microstrip geometry which is also a good estimation for the geometry "wire over ground plane" if the distance to ground is much wider than the wire size. In our case we gained an impedance of 263Ω from a microstrip line width of 1mm at an effective distance of approximately 10mm from ground [18]. The capacitance per length was calculated by the measured static test board capacitance for the case of corner pins (see fig. 4) in order to exclude line coupling. For the line length of the hot discharge path we obtained $C' = 35.5pF/0.39m = 91pF/m$. Then the inductance per length was given by $L' = Z_w^2 \cdot C' = 6.3\mu H/m$. Since the transmission lines covered an entire length of 0.605m, we had $6.3\mu H/m \cdot 0.605m = 3.8\mu H$. With the 4th order LEM method, the complete inductance of the discharge path was determined at 12µH. The pulse forming inductances were present as concentrated elements of 2.2µH, each. Therefore, a residual inductance of $12\mu H - 2 \cdot 2.2\mu H - 3.8\mu = 3.8\mu H$ was considered by L_{res}. C_{hot} depicted the input capacitance of the voltage probe and the parasitic capacitance of the CT1. C_{cold_back} is the background capacitance at the cold side of the device, which represents the conducting parts of the IC, the package and the socket. It was found that the bonding inductances play only a minor part for the TL approach. They were the only current limiting elements for the LEM model in the discharge path between test board capacitance and device. Each transmission line was considered in the circuit simulation by a set of 100 differential line elements. Thus the total computation time with a nonlinear DUT did not exceed 10 minutes.

6.2 Application of the lumped element model and the transmisson line approach to a snapback transistor

In fig. 18 different tester models, the LEM as well as the transmission line (TL) approach, are compared to the HBM pulse measurements gained from the nMOSFET.

For both, measurement and simulation, the device was triggered to the bipolar mode and clamped the voltage at its hold value v_h. Ringing and the discharge of the test board capacitance due to the snapback were reproduced very well by the TL approach. The LEM could not model the discharge correctly since it considered the test board as a discrete element.

After the current was below a certain hold current the transistor switched off. The remaining current of the HBM pulse started to charge up the test board once again. The simulated charging characteristic of the test board capacitance (LEM) was very similar to the measurement. The time constant of the charging is mainly determined by $\tau = R_{HBM} \cdot C_{board}$. The time constant gained by the 1/e-value from the measurement equals 38.2ns, which is in very good accordance with the value for the simulation $\tau = 1.5k\Omega \cdot 25pF = 37.5ns$. For this time domain, the amount of the concentrated test board capacitance considered as an effective value is confirmed correctly. Here, the TL approach showed ringing due to signal reflection at the mismatched line ends. However, it reproduced the same behaviour in general.

After the breakdown voltage was reached again, the device went into snapback. However, it could not exceed its holding current and changed to blocking state. Thus, charging started once again. With the discharge current at about 500ns, a good correlation

betwe the TL approach and the measurement could be demonstrate

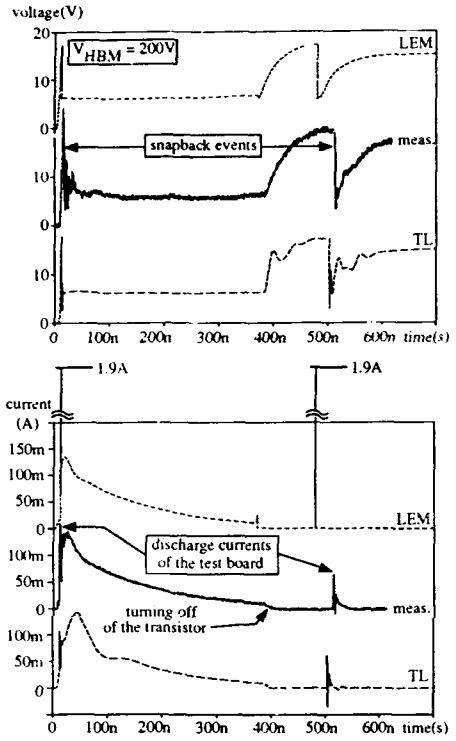


*Fig. 18: The **nMOSFET** during HBM stress: Numerical simulations by means of the 4th order LEM and the transmission line (TL) approach in comparison to the measurement of fig. 14.*

Concerning the application of both tester models, we can concluded that the lumped element model (LEM) is better suitable in the HBM time domain. On the other hand, fast switching effects of less than some nanoseconds are treated properly by the transmission line approach.

6.3 Application of the lumped element model and the transmisson line approach to a a protection diode

Figure 19 depicted clearly that the TL approach is able to reproduce the waveforms of a protection diode at the moment of a sudden voltage drop caused by second breakdown. The current transients were only provided satisfactorily with a transmission line discharged into the device. The charge was stored mainly in the transmission line TL_{adh} between the device and the inductance L_{hot}. Fast discharges of the other transmission lines were blocked by this inductance. L_{hot} and L_{cold} were responsible for the low frequency ringing of the voltage after the second breakdown.

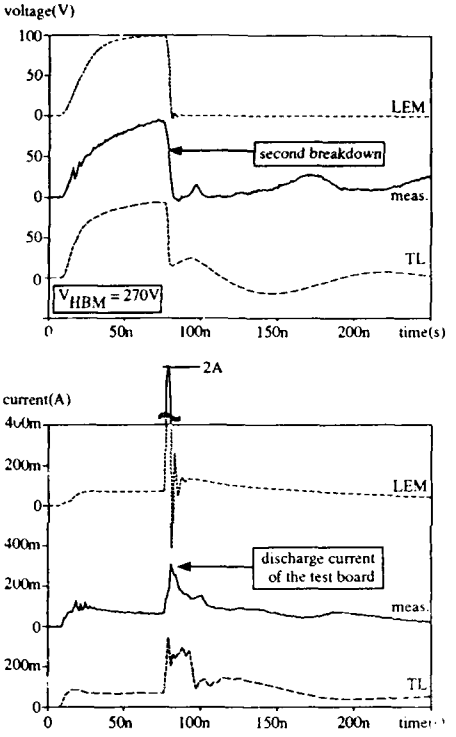


*Fig. 19: The **protection diode** during HBM stress: Numerical simulations by means of the 4th order LEM and the transmission line (TL) approach in comparison to the measurement of fig. 16.*

7. Conclusions

The 4th order lumped element tester model was proved according to the EOS/ESD HBM standard [1]. Furthermore, it provided information for additional tester characterization. Interactions between tester and device were reproduced correctly in the "conservative" HBM time domain. However, the transients of fast switching events were confirmed only by the transmission line approach.

The in-situ observation of snapback and failure in HBM testers reveals progress in failure modeling and new aspects for the interaction between device and tester.

The comparison of 3 different HBM testers revealed a clear dependence for the HBM thresholds of the devices stressed in this study on the parasitic test board capacitance. The higher the value of the capacitance is, the more the HBM pulse is damped and thus the tester pretends a higher ESD hardness. However, depending on the actual protection device and/or structure, the influence of the test board capacitance may appear in very different ways.

Further knowledge on the transmission line characteristic of the test board capacitance should be gained.

8. Acknowledgements

Heinrich Wolf (TU Munich) is highly acknowledged for his valuable support in measurement and simulation and Peter Scheubert (TU Munich) for fruitful discussions concerning transmission line theory. The authors would like to thank Tilo Brodbeck and Xaver Guggenmos (Siemens AG, Munich, Germany) for providing an HBM tester, test structures and their helpful discussions.

9. References

[1] "EOS/ESD-S5.1-1993", in Standard for ESD Sensitivity Testing - HBM - Component Level, June 9th 1993 Standard

[2] "MIL STD 883C Notice 8, Method 3015.7", in Military Standard for Test Methods and Procedures for Microelectronics: Electrostatic Discharge Sensitivity Classification, March 22nd 1989

[3] L. v. Roozendaal, A. Amerasekera, "Standard ESD Testing of Integrated Circuits", Proc. of EOS/ESD Symposium 1990, pp. 119-130

[4] K. Verhaege, P.J. Roussel, G. Groeseneken, H.E. Maes, H. Gieser, C. Russ, P. Egger, X. Guggenmos and F.G. Kuper, "Influence of Test Board Capacitance of HBM ESD Testers and Impact on HBM ESD Tester Standard Specification, Using a 4th Order Lumped Element Model", Proc. of the EOS/ESD Symposium 1993, pp. 129-138

[5] C. Russ, H. Gieser, P. Egger, S. Irl, "Compact Electro-Thermal Simulation of ESD-Protection Elements", Proc. of the ESREF Symposium, Bordeaux, 1993, pp. 395-400, to be published in the Quality and Reliability Engineering International, J. Wiley & Sons

[6] C. Russ, P. Egger, "Task Summary on Electro-Thermal Circuit Simulation", ESPRIT project 5005, ESD Protection for Submicron Technologies, June 1993

[7] T.J. Maloney and N. Khurana, "Transmission Line Pulsing Techniques for Circuit Modeling of ESD Phenomena", Proc. of the ESD/ESD Symposium 1985, pp. 49-54

[8] R. Kropf, C. Russ, R. Kolbinger, H. Gieser, S. Irl, "Zeitaufgelöste Untersuchung des Snapbackverhaltens eines ESD-Schutztransistors", Tagungsband 3. ESD-Forum, Grainau, VP Verlags GmbH Herrenberg, 1993, pp. 19-26

[9] J.R.M. Luchies, C.G.C.M de Kort, J.F. Verweij, "Bipolar Transient Turn-On of an ESD Protection Circuit", Proc. of the ProRisc IEEE Benelux Workshop on Circuits, Systems and Signal Processing (CSSP), 1994, pp. 151-155

[10] C. Russ, H. Gieser, P. Egger, R. Kropf, K. Verhaege, "Kompakte Bauelementemodelle von ESD-Schutzstrukturen und deren Anwendung", Tagungsband 3. ESD-Forum, Grainau, VP Verlags GmbH Herrenberg, 1993, pp. 131-140

[11] D.G. Pierce and D.L. Durgin, "An Overview of EOS Effects on Semiconductor Devices", Proc. of EOS/ESD Symposium 1981, pp. 120-131

[12] D.G. Pierce, W. Shiley, B.D. Mulcahy, K.E. Wagner and M. Wunder, "Electrical Overstress Testing of a 256K UVEPROM to Rectangular and Double Exponential Pulses", Proc. of EOS/ESD Symposium 1988, pp. 137-146

[13] C. Diaz, C. Duvvury, S. Kang and L. Wagner, "Electrical Overstress (EOS) Power Profiles: A Guideline to Qualify EOS Hardness of Semiconductor Devices", Proc. of EOS/ESD Symposium 1992, pp. 88-94

[14] D.C. Wunsch, R.R. Bell, "Determination of Threshold Failure Levels of Semiconductor Diodes due to Pulse Voltage", IEEE Trans. Nuc. Si., NS-15, 1968, pp. 244-259

[15] D. Krakauer, "ESD Protection in a 3.3V Sub-Micron Silicided CMOS Technology", Proc. of EOS/ESD Symposium 1992, pp. 250-257

[16] C. Buj, C. Leroux, J.-P. Chante, "Validation of the Medici Electrothermal Module by ESD Failure Study of a PIN Diode", Proc. of the ESREF Symposium, Bordeaux, 1993, pp. 427-432

[17] D. Smith, "Applying High Frequency Techniques to Measuring ESD Phenomena and its Effects", Proc. of EOS/ESD Symposium 1989, pp. 114-119

[18] R.K. Hoffmann, "Integrierte Mikrowellenschaltungen", Springer-Verlag, Berlin, 1983

A Physical Model for the Creation of Latent Gate Oxide Defects by Very Fast Electrostatic Discharges

Joachim C. Reiner

Professur für Zuverlässigkeitstechnik, Eidgenössische Technische Hochschule Zürich,
ETH-Zentrum, CH-8092 Zürich, Tel ++41 1 632'41'96, Fax ++41 1 251'21'72

1. ABSTRACT

Many MOS devices in polysilicon gate technology show a characteristic kind of gate oxide damage after exposure to an electrostatic discharge (ESD) according to the Charged-Device-Model (CDM), which can not be reproduced by ESD according to the Human-Body-Model. In this paper, a possible physical explanation is proposed for the creation of this kind of damage. According to this, during the CDM-ESD-event the gate oxide is destroyed and an amorphous region in the silicon is created. This can explain the observed recovery behavior of the leakage current of damaged devices. The sudden increase of the leakage current of damaged devices following an electrical overstress could consequently be due to thermally induced recrystallisation of the amorphous region.

2. INTRODUCTION

Recently, several publications [1,2,3,4] have documented that Electrostatic Discharge according to the Charged-Device-Model (CDM-ESD), specified in [5], can cause a particular type of damage when applied to MOS-devices. In all cases a gate oxide is damaged. In the case of an input pin of an nMOS-device, a study has shown that the gate oxide of the input transistor behind the protection structure is damaged [1,2]. Whereas the discharge amplitude necessary to create a damage depends strongly on the type of device, the damage itself is always of the same type. The amount of electrostatically stored energy that is necessary to create this damage is much smaller than that in the case of Human-Body-Model (HBM) - ESD (specified in [6]). Compared to HBM-ESD, CDM-ESD is much faster. The rise time and the duration of this event are shorter by more than one order of magnitude. This probably explains why many devices are more sensitive to CDM-ESD. CDM-ESD corresponds to the situation where a device has been charged electrostatically, usually by triboelectricity, and discharges to a metallic ground. This situation can occur for example during automatic handling.

The electrical properties of the type of damage created by CDM-ESD are unstable in two respects: 1. The leakage current caused by this damage is reduced by 50 % in about two weeks at 25°C, and 2. The electrical conductivity of this damage switches irreversibly from the initial low leakage state to a high leakage state when the device is subjected to a small electrical overstress. This behavior can cause a reliability problem, when devices with this type of damage are used in systems, because small electrical overstress can increase the leakage current to an extent which is likely to cause functional failures. Both of these instabilities can be explained with the model proposed in this paper.

3. DESCRIPTION OF THE DAMAGE

In a typical input pin, a leakage current of the order of 1 µA at 5 V can be measured after the device has been damaged by CDM-ESD (Figure 1).

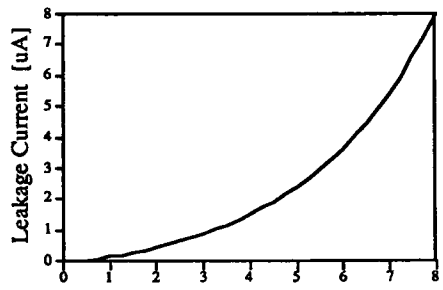


Fig. 1: Typical leakage current measured at an input pin of a damaged device.

The high sensitivity of this type of damage to electrical overstress [1,2] is illustrated by Figure 2 (a), which shows the leakage current characteristic of damaged pins when a voltage ramp is applied. All damaged pins "break down" between 10 and 14 V. The same tape of behavior could be reproduced by damaging the gate oxide of single MOSFET test transistors, shown in Figure 2 (b).

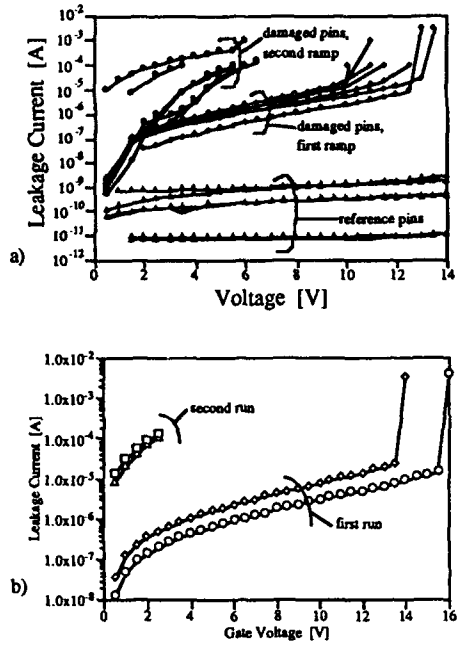


Fig. 2: Gate leakage current characteristic of a number of gate oxide damages in a commercial VLSI MOS-device (a), data taken from [2]) and in two single test structure MOSFETs (b). Each object is measured twice by applying the same voltage stair (first run, second run). The "breakdown", observed during the first run, transforms the latent state into the permanent state.

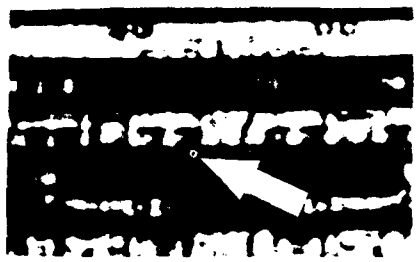


Fig. 3: Picture of the light emission (arrow) of a damage site taken by Emission Microscopy.

The only way of localizing this damage without destructive action to the device takes advantage of a faint emission of light accompanying the leakage current flow (Figure 3). Wet chemical removal layer by layer down to the polysilicon gate reveals no visible damage even by inspection with a SEM. After removing the gate with HF, the damage is located in the form of a little etch hole in the substrate (Figure 4). Because no staining etch was used, the size of this hole can be assumed to be the size of the physical damage. It has a diameter of about 300 nm, about ten times as large as the thickness of the gate oxide.

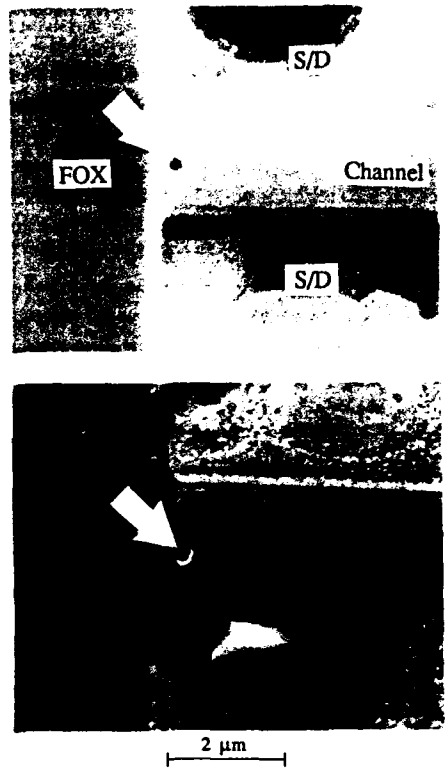


Fig. 4: Damaged input transistors after HF lift off of all layers. The etch holes in the gate region mark the damaged region (arrows).

The appearence of such an etch hole without using a dedicated staining etch implies that the material which was in the place of that hole had a completely different etching behavior compared to the surrounding material. Chemically it must be mainly silicon, but it must have had a highly damaged physical structure.

It can be assumed that this material had been molten for a short time due to an electrical discharge after breakdown of the oxide caused by a transient overvoltage at the gate. This transient overvoltage is caused by the CDM-ESD event. It lasts at most a few nanoseconds, the same as the discharge current flow of the CDM-ESD-event itself,

which is shown in Figure 5.

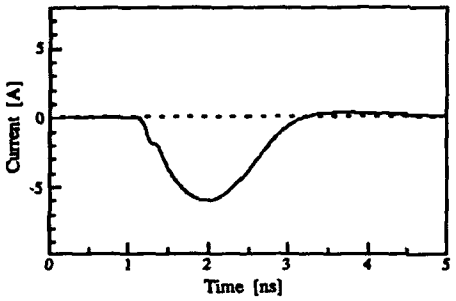


Fig. 5: Discharge current transient flow through the stressed pin. (Same type of device as shown in Figure 3 and 4.) The device was charged to -340 V relative to the laboratory ground.

4. MELT BALL MODEL

The model proposed here is called "Melt Ball Model" (MBM). It models the flow of the heat which is set free as a consequence of the discharge of the gate capacitor following the electrical breakdown of the gate oxide. The discharge is a local event, dissipating heat in a very short time. It is modeled here as a source of heat δ-shaped in space and time $Q \cdot \delta(\vec{x},t)$ at the location of the breakdown $\vec{x}$, thus simplifying the problem to spherical symmetry. The flow of heat is described by the heat conduction equation

$$\frac{\partial T}{\partial t} = D \cdot \Delta T \ . \qquad (1)$$

According to [7], for this simplified situation it has the solution

$$T = \frac{Q}{\rho \cdot c} \cdot \frac{a^3}{\pi^{3/2}} \cdot e^{-a^2 r^2}; \quad a = \frac{1}{\sqrt{4Dt}} \ . \qquad (2)$$

r is the distance from the position of the heat source. As an approximation, the material parameters are assumed to be those of silicon at room temperature, and to be temperature independent. The thin oxide layer is ignored, so is the heat of fusion. Therefore, $\rho = 2.3$ g/cm^3 is the density, $c = 0.7$ J/g/K the specific heat and $D = 0.9$ cm^2/s the temperature diffusion coefficient, for silicon at 300 K [8]. For the very small amount of energy $Q = 5 \cdot 10^{-11}$ J set free as heat, this solution is shown in Figure 6 for different times t.

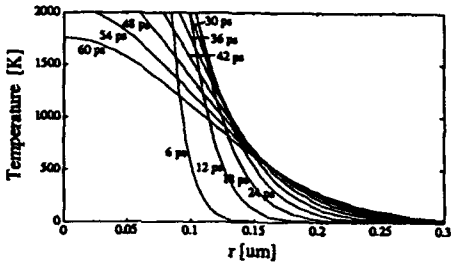


Fig. 6: Solutions of the heat conduction equation for different times t = 6,12,18,...60 ps (Eq. 2), for $Q = 5 \cdot 10^{-11}$ J. Temperature here means the temperature increase on top of the ambient temperature.

From Figure 6 it can be concluded that a region with about 120 nm radius would melt for a short time. The heat of fusion, which has been neglected, will reduce the radius by approximately 10%.

The conclusion can be drawn that a fairly small amount of energy is capable of creating a melt damage, if only it is dissipated into heat locally and fast enough. The above model is an approximation in this sense. It is valid, when the time of heat dissipation τ and the size of the heat source r_o are small enough.

5. COMPARISON WITH EXPERIMENTAL OBSERVATION

A typical situation is that of an input pin, where CDM-ESD causes damage in the input transistor connected to this pin. The experimental situation of an input signal path to the input transistor containing pieces of polysilicon lines as well as aluminium lines can be simplified to the picture shown in Figure 7.

From the occurrence of the damage after CDM-ESD it is clear that the CDM-ESD event created a voltage between the gate of the input transistor and the substrate that has been high enough to create a gate oxide breakdown. The following discharge through the breakdown site causes local heating. According to the above heat flow analysis, a limited amount of heat can induce melting only if it is set free in a small volume and in a short time. This means that for the creation of this damage only those energy reservoirs are relevant which can be accessed very quickly (< 25 ps).

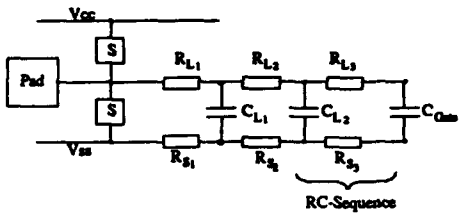


Fig. 7: Electrical schematic of the signal path from the pad to the damaged input transistor. S = ESD protection structures, R_{Li} = polysilicon line resistances, C_{Li} = metal line capacitances, R_{Si} = substrate resistances, C_{Gate} = gate capacitance.

For the situation depicted in Figure 7 the following consideration is made to find out which energy reservoirs are relevant. First the available amount of energy is estimated. An important parameter is the breakdown field $E_{ox,BD}$, which determines the amount of energy that can be stored electrostatically in the capacities. From an investigation [9] it can be estimated to about 15 MV/cm for oxides of 20 nm thickness. This yields a voltage of ≈ 30 V. The electrostatically stored energy of the gate capacitance is

$$W = \tfrac{1}{2}CV^2 = \tfrac{1}{2}\varepsilon_o\varepsilon_r A_{gate} d_{ox} E_{ox,BD}^2 \,. \qquad (3)$$

With d_{ox} = 20 nm, and A_{ox} = 75 µm², the gate oxide region of this case, this yields

$$W \approx 5{\cdot}10^{-11}\ \text{J}.$$

The capacitor C_L next to the transistor (≈ 400 x 4 µm² aluminium over ≈ 1 µm oxide giving $C_L \approx 5{\cdot}10^{-14}$ F) stores about

$$W = \tfrac{1}{2}CV^2 \approx 2.2{\cdot}10^{-11}\text{J}. \qquad (4)$$

If $Q \approx 5{\cdot}10^{-11}$ J would be transformed into heat, according to Figure 6 the melt ball would reach its maximum radius of $r_{max} \approx 120$ nm after $t_{rmax} \approx 25$ ps.

The RC time constant of the RC sequence just beside the input transistor is estimated as follows:

$R_L \approx 2.5$ x 50 µm² poly of R ≈ 50Ω/□ (50Ω per square) yields $R_L \approx 1$kΩ, so that

$$\tau = RC \approx 50\ \text{ps}. \qquad (5)$$

This shows that even the first RC sequence is not likely to be relevant any more for the creation of the melt ball. According to Figure 6 an amount of dissipated heat of $Q \approx 5{\cdot}10^{-11}$ J should melt a region of ≈ 240 nm in diameter. This agrees well with the observed ≈ 300 nm as shown in Figure 4. Due to overetching the size of the etch hole could well be larger than the melt ball. But even a difference of about 50 % would be acceptable for such a simple model.

From this observation the conclusion is allowed that a discharge must have taken place (Figure 4), and that an energy of ≈ 10^{-11} - 10^{-10} J stored in the gate capacity must have been dissipated in the form of heat within a diameter of ≈ 200 nm and a time shorter than 50 ps.

6. STRUCTURE OF THE DAMAGE

The question arises: Is there still an adequate oxide left? The leakage currents observed from damaged devices decreases with time, even at room temperature [1,2]. The observed leakage current in the range of 1 µA is much smaller than the leakage currents typically found after gate oxide breakdown. Is the oxide in a state of degradation, so that it will recover its isolation capability? Both the damage found after deprocessing and SEM inspection (Figure 4) and the "Melt Ball Model" imply a geometry of the damage as scetched in Figure 8. It shows that the diameter of the created melt region is expected to be much larger than the oxide thickness.

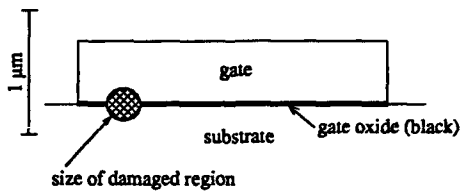


Fig. 8: Geometry of the damaged region, to scale.

Here are two considerations according to which the oxide is probably destroyed. (1) The first consideration follows from the described model, according to which the center of the molten volume should reach temperatures far above the melting point of the oxide, which is about 1600°C. A film of liquid oxide would be surrounded by liquid silicon. It would be hard to believe that the oxide could withstand this as an intact layer. (2) The second consideration compares the observed leakage current to the current conduction in a "good" oxide. A leakage current of ≈ 1 µA flowing through a region with a diameter of 300 nm corresponds to a current density of ≈ 1000 A/cm². At the applied voltage of 5 V, an oxide of 20 nm thickness would conduct an unmeasurably small leakage

current. On the other side, a tunneling current density of ≈ 1000 A/cm² would be expected to occur at fields >> 15 MV/cm according the high field approximation of Fowler and Nordheim [10]. Even a "good" oxide would break down immediately at this electrical field.
From this considerations it is clear that the oxide must have been destroyed at least in a part of the damaged region. It can be expected to be punctured at the damage site.

The second question is: What does the rest of the damaged area look like? Here the dynamics of the resolidification has to be considered. According to Figure 6, the melting and cooling of the damage region takes place in less than 1 ns, if the heat is set free in a short time. Heat can be set free during the whole CDM-ESD discharge event. The shape of the CDM-ESD discharge current, shown in Figure 5, had a half width of less than 2 ns. This is in agreement with other investigations which show durations of CDM-discharges of less than 5 ns [12,13]. So the value of 5 ns can be taken as an upper limit for the time constant of the decay of the dissipation of thermal power in the damage area. Together with the melt ball diameter, a lower limit for the speed of the resolidification front can be estimated:

$$\frac{dr(t)}{dt} \approx \frac{r_{max}}{\tau} = \frac{150\ \text{nm}}{5\ \text{ns}} = 30\ \frac{\text{m}}{\text{s}}. \qquad (6)$$

Even this conservative estimate yields a value which is larger than the amorphisation limit of 15 m/s by a factor of two, at which silicon solidifies in amorphous form [11]. Therefore, amorphisation of the damage region is likely to occur. It can be speculated that the damage region looks like the sketch in Figure 9.

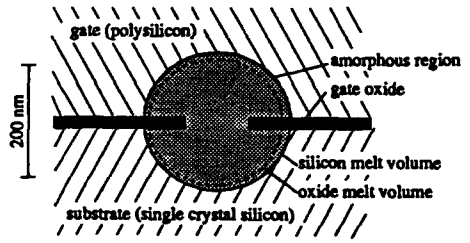


Fig. 9: Sketch of the silicon melt volume and the oxide melt volume (exceeding melt temperature) for Q ≈ 5.10⁻¹¹ J.

7. SENSITIVITY TO ELECTRICAL OVERSTRESS

Finally, a possible explanation is given for the observed high sensitivity of CDM-damaged devices against electrical overstress. Two experiments providing small electrical overstress to the devices have been described, both leading to an increase in leakage current of about two orders of magnitude [1,2]. Neither one of them does change undamaged devices. In the first case, the leakage current increase was induced by ESD stress according to the Human-Body-Model (HBM) [6]. Here the increase occurred after a stress voltage of less than 50 V. For comparison: "Good" devices stand more than 1000 V of HBM-ESD stress without measurable change. In the second case, the leakage current increase occurred when a voltage ramp was applied, as shown in Figure 2. In both of the described cases the effect was perfectly reproducible. Both cases imply that electrical stress is applied to the damage region. Under the assumption of a metastable, probably amorphous damage region, it is possible that the leakage current increase is caused by a transformation of a part of the damage region into polycrystalline silicon, forming a filament with increased conductivity. The electrical conductivity of polysilicon is usually several orders of magnitude larger than that of amorphous silicon [8,14], especially when considering that dopants from the highly doped polysilicon gate can penetrate into the region of interest. As the HBM-discharge decays with a time constant of 150 ns [6], the speed of the resolidification front can be estimated to

$$\frac{dr(t)}{dt} \approx \frac{r_{max}}{\tau} = \frac{150\ \text{nm}}{150\ \text{ns}} = 1\ \frac{\text{m}}{\text{s}}, \qquad (7)$$

which is clearly smaller than the amorphisation limit of 15 m/s [11], so the remolten material has time to turn into a polycrystalline state. In the case of the voltage ramp measurement, the decay time constants are in the millisecond range and thus even slower.
Therefore the increase in conductivity of the CDM-ESD damage caused by the described types of (weak) overstress can be interpreted as a change of the material in the damage region to a structure with higher crystallinity and conductivity.

8. DISCUSSION

The observed type of gate oxide damage leads to a leakage current of about 1 μA at 5 V, which is very low. Failure analysis showed damages under the gate with a diameter of about 300 nm. A model is proposed for the creation of this damage that yields a diameter which is within less than a factor of two of the experimental observation.

A possible explanation for the electrical behavior of the damage could be that the material in the damaged region turned amorphous, a state which can transform to a structure with higher electrical conductivity under the influence of electrical stress. TEM investigations of such damages are in progress.

9. CONCLUSIONS

The properties of the type of gate oxide damage in MOS devices caused by Electrostatic Discharge according to the Charged-Device-Model have been described. In the view of their high sensitivity to electrical overstress, they can be called latent. A model for the creation of this type of damage is proposed.

Although the observed leakage currents are very low, it can be concluded that the gate oxide is completely destroyed. This conclusion is consistent with the proposed model and the observed properties of this type of damage.

10. ACKNOWLEDGEMENTS

I would like to thank Horst Gieser, Technical University of Munich, for the measurement of the CDM-ESD discharge current, and Faselec AG, Zürich, for providing the MOSFET test transistors.

11. REFERENCES

1. H.A.Gieser, P.Egger, M.R.Herrmann, J.C.Reiner, A.Birolini: "A CDM-only Reproducible Field Degradation and its Reliability Aspects", (Proceedings of the) ESREF 1993, p. 93

2. J.C.Reiner, M.R.Herrmann, H.A.Gieser, P.Egger: "Untersuchung der Zuverlässigkeit einer CDM-reproduzierbaren Feldvorschädigung", Tagungsband 3. ESD-Forum 1993/1994, p. 170

3. W.H.Tan: "Minimizing ESD Hazards in IC Test Handlers and Automatic Test Equipment", EOS/ESD Symposium Proceedings 1993, p.57

4. J.Colvin: "The Identification and Analysis of Latent ESD Damage an CMOS Input Gates", EOS/ESD Symposium Proceedings 1993, p.109

5. EOS/ESD-Association, "EOS/ESD-DS5.3-1993 Draft Standard for ELECTROSTATIC DISCHARGE (ESD) SENSITIVITY TESTING – CHARGED DEVICE MODEL (CDM)-Component Level", ISBN 1-878303-43-0

6. "MIL-STD 883C Method 3015.7 Notice 8" in Military Standard for Test Methods and Procedures for Microelectronics: ESD Sensitivity Classification, March 22, 1989

7. Gilbert Strang, "Introduction to Applied Mathematics", Wellesley-Cambridge Press, 1986, Chapter 6.4

8. "VLSI Technology", Ed. S.M.Sze, McGraw-Hill, 1988

9. Y.Fong, C.Hu: "The Effects of High Electric Field Transients on Thin Gate Oxide MOSFETs", EOS/ESD Symposium Proceedings 1987, p.252

10. M.Lenzlinger, E.Snow.: "Fowler-Nordheim tunneling into thermally grown SiO2", Journal of Applied Physics 40(1), Jan 1969 p.278

11. M.O.Thompson et al.: "Melting Temperature and Explosive Crystallization of Amorpous Silicon during Pulsed Laser Irradiation", Phys. Rev. Lett. 52(26) 1984, p.2360

12. R.G.Renninger, "A Field Induced Charged Device Model Simulator", 1989 EOS/ESD Symposium Proceedings, p.59

13. K.Lundmark, "ESD Sensitivity; Experimental Verification of Three Different Models", (Proceedings of the) ESREF 1993, p. 379

14. P.Thomas, J.C.Flachet: "Conductivity Measurements on UHV Deposited Amorphous Silicon", Journal de Physique C4 (10) 42, 1981, p.151

COMPARISON OF TRANSIENT SUPPRESSORS DEVICES

E. Mino, J. E. Vila

TELEFONICA INVESTIGACION Y DESARROLLO

Tel. 34 1 3374428 - Fax. 34 1 3374212

Emilio Vargas, 6

28043 MADRID (SPAIN)

INTRODUCTION

This work presents the results of the strength to surge voltage tests carried out on different discharge protection devices, with application in telecom equipment.

The objetives of this study were to determine the discharge resistance of these components against a standard surge voltage, to compare the results with the manufacturer specification (usually for a different waveform) and to perform measurements of the transients voltages transmitted, that can affect the circuit to protect, the current absorbed by the transient suppressor and the leakage current.

DESIGN OF TEST

Series of 10 positive and negative voltage pulses were applied with a 1/1000μs waveform, with an amplitude (in open circuit) from 500 V to 6000 V, in steps of 500 V, until the failure of the component. The time between pulses was 1 min. The discharge generator was a Haefely P6T with an output resistance of 15Ω + 25Ω.

The voltage pulse 1/1000 μs is the standard requirement for surge protection in telecommunication equipment in Spain (the raise time of this pulse is 1 μs and the fall time to half peak voltage is 1000μs). In shortcircuit the discharge generator gives an approximate current waveform of 2/500μs.

The leakage current was measured with a Semiconductor Parameter Analizer HP 4145B.

TEST COMPONENTS

Tests were carried out on crowbar devices (trisils and gas arresters) and clamping devices (transils and varistors) with a breakdown voltage about 200 V for all devices. All of them have similar size.

ANALYSIS OF TEST RESULTS

Tabla 1 summarizes the strength discharge test showing the maximum peak voltage and current before the component failure, and the data sheet specification.

Tabla 1

Strength discharge test

	VARISTOR	GAS ARRESTER	TRISIL	TRANSIL
DATA SHEET	15A (1)	10.000A (2)	90A (3)	2.2A (3)
TEST DATA	86A (4)	>140A (5)	>140A (5)	<6A
	(4.500V)	(>6.000V)	(>6.000V)	(<500V)

- (1) Rectangular current waveform of 2 ms.
- (2) Exponential current waveform of 8/20 μs.
- (3) Exponential current waveform of 10/1000 μs.
- (4) Exponential current waveform with times dependents of the voltage amplitude.
- (5) Exponential current waveform of 2/500 μs.

The obtained results were the following:

- The trisil and the varistor show more strength than specified, although the vendor specified tests define waveforms with higher energy.
- The transil failed at the first discharge at 500 V. It's a device with application in low voltage protection. The reason is that the peak power dissipation (Vpp x Ipp) is a constant for all transils with the same silicon size, then for high voltage components the current available is very small.
- The trisil and specially the gas arrester show the highest resistance to discharges and they didn't fail during the test.

Figura 1, Figura 2 and **Figura 3** show the current and the voltage waveforms obtained when a discharge 1/1000 μs and 1500 V was applied to different devices.

The varistor peak voltage rose to 320 V, a 56% more than the data sheet breakdown voltage at 1 mA (205 V). The voltage duration is 3.7 ms, and the current duration is 1.8 ms this corresponds with the time when the voltage is higher than 200 V.

The ohmic value at maximum peak voltage varied from 50 Ω to 4 Ω, depending on the transient amplitude (500 V to 4500 V).

The gas arrester peak voltage rose to 650 V, a 317% more than the data sheet dc spark-over voltage (230 V). The pulse voltage duration was 400 ns.

The ohmic value presented by the gas arrester, when fired, varied between 1568 mΩ at 500 V and 102 mΩ at 6000 V.

The trisil peak voltage rose to 160 V, that is smaller than the data sheet breakdown voltage (200 V), because the intrinsic trisil capacity helps to fire it; for this reason the circuit to protect will not suffer perturbation. The pulse voltage duration is 2 μs.

The ohmic value presented by the trisil, when fired, varied between 342 mΩ at 500 V and 74 mΩ at 6000 V.

Tabla 2 summarizes the characteristics of the pulses transmitted by the transient suppressors devices.

Tabla 2
Voltage pulse characteristics at 1.500 V

	DURATION	AMPLITUDE
VARISTOR	3.7 ms	320 V
GAS ARRESTER	400 ns	650 V
TRISIL	2 μs	160 V

Tabla 3, Tabla 4 and **Tabla 5** shows the measurements at differents peak pulse voltages on the protection devices.

Tabla 3
Measurements on the varistor

V_{GEN} (V)	V_{VAR} (V)	I_{VAR} (A)	R_{VAR} Ω	P_{VAR} (W)	
500	290	5.8	49.65	1683	
1000	312	15	20.8	4680	
1500	320	29.2	10.9	9344	
2000	332	37.6	8.82	12483	
2500	342	44	7.72	15048	
3000	350	56	6.25	19600	
3500	357	64.6	5.50	23133	
4000	366	76	4.81	27816	
4500	372	86	4.32	31992	
5000	376	98	3.83	36848	(1)
5500					
6000					

- V_{GEN} Peak pulse voltage applied by the discharge generator in open circuit
- V_{VAR} On-state peak pulse voltage.
- I_{VAR} On-state peak pulse current.
- R_{VAR} Varistor fired resistance = V_{VAR} / I_{VAR}
- P_{VAR} Peak pulse current power dissipation in the varistor = $V_{VAR} * I_{VAR}$

(1) The varistor exploded after 8 pulses at 5.000V.

Tabla 4

Measurements on the gas arrester

V_{GEN} (V)	V_{OFF} (V)	t_{OFF} (μs)	I_{ON} (A)	V_{ON} (V)	R_{ON} mΩ	P_{ON} (W)
500	500	1.5	12.5	19.6	1568	245
1000	600	0.9	25	15	800	500
1500	650	0.9	37.5	15	400	562
2000	534	0.9	50	20	400	1000
2500	573	0.9	62.5	15	240	937
3000	620	0.9	75	14.4	192	1080
3500	650	0.9	87.5	16	182	1400
4000	683	0.9	100	16	160	1600
4500	737	0.9	112.5	14.6	130	1642
5000	758	0.9	125	15.6	125	1950
5500	771	0.9	137.5	15.2	110	2145
6000	847	0.9	150	15.4	102	2310

- V_{GEN} Peak pulse voltage applied by the discharge generator in open circuit.
- V_{OFF} Stand-off peak pulse voltage.
- t_{OFF} Stand-off voltage pulse duration.
- I_{ON} On-state peak pulse current.
- V_{ON} On-state peak pulse voltage.
- R_{ON} On-state resistance = V_{ON} / I_{ON}
- P_{ON} On-state peak power dissipation = V_{ON} * I_{ON}

Tabla 5

Measurements on the trisil

V_{GEN} (V)	V_{OFF} (V)	t_{OFF} (μs)	I_{ON} (A)	V_{ON} (V)	R_{ON} mΩ	P_{ON} (W)
500	200	2.5	10.5	3.6	342	37.8
1000	160	1.6	20	3.8	190	76
1500	158	1.6	32	4.6	143	147.2
2000	178	1.6	44	5.2	118	228.8
2500	154	1.5	52	5.8	111	301.6
3000	148	1.5	62	6.4	103	396.8
3500	142	1.5	75	7	93	525
4000	141	1.7	76	7.6	97	592
4500	144	1.9	82	8.4	102	688.8
5000	139	2.2	104	9	86	936
5500	142	2.2	120	9.6	80	1152
6000	139	2.2	140	10.4	74	1456

- V_{GEN} Peak pulse voltage applied by the discharge generator in open circuit.
- V_{OFF} Stand-off peak pulse voltage.
- t_{OFF} Stand-off voltage pulse duration.
- I_{ON} On-state peak pulse current.
- V_{ON} On-state peak pulse voltage.
- R_{ON} On-state resistance = V_{ON} / I_{ON}
- P_{ON} On-state peak power dissipation = V_{ON} * I_{ON}

Tabla 6 shows the leakage currents measured for the test samples at 200V, and the maximun specified by the manufacturer data sheet.

Tabla 6

Leakage current at 200 V

	VARISTOR	GAS ARRESTER	TRISIL
TEST DATA	25 μA	2.5 pA	100 nA
DATA SHEET	1mA	no spec.	1 mA

The results evidence the low leakage current for the gas arrester and the trisil, the high varistor leakage current can limit its application in circuits with low leakage requirements.

Figura 1 Varistor voltage (V) and current (I)

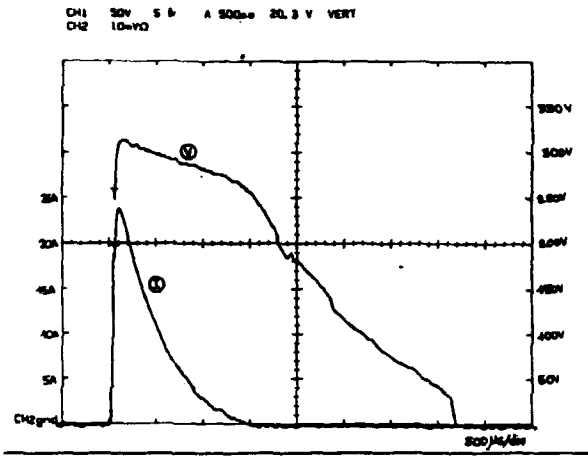


Figura 2 Gas arrester voltage (V) and current (I)

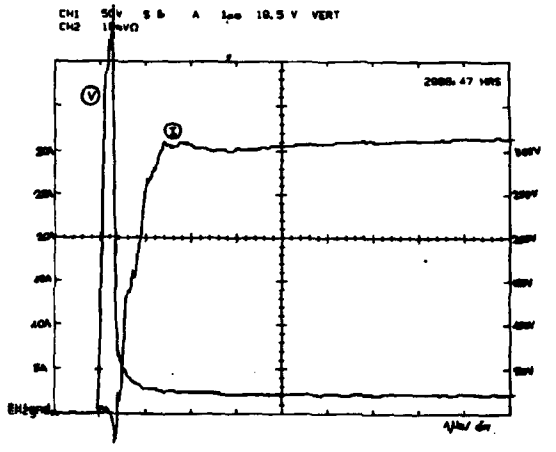
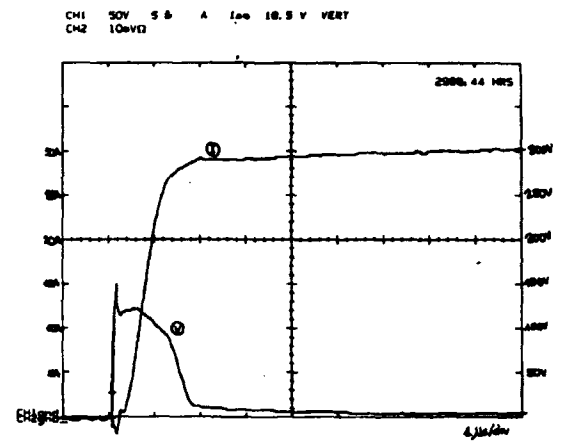


Figura 3 Trisil voltage (V) and current (I)



CONCLUSIONS

As a result of the test we can come to the following conclusions:

1. The more resistant devices are, in this order: gas arrester, trisil, varistor and transil. The high voltage transils show very low resistance to surge voltage, they are not suitable to be used in telecom applications.

2. The gas arrester and the varistor transmit voltage peaks dangerous to the circuit to protect. The gas arrester peak has a short duration (~0.4 μs) and very high voltage (~650 V), the varistor peak has a large duration (~1.8 ms) and high voltage (~320 V). The trisil doesn't transmit perturbations.

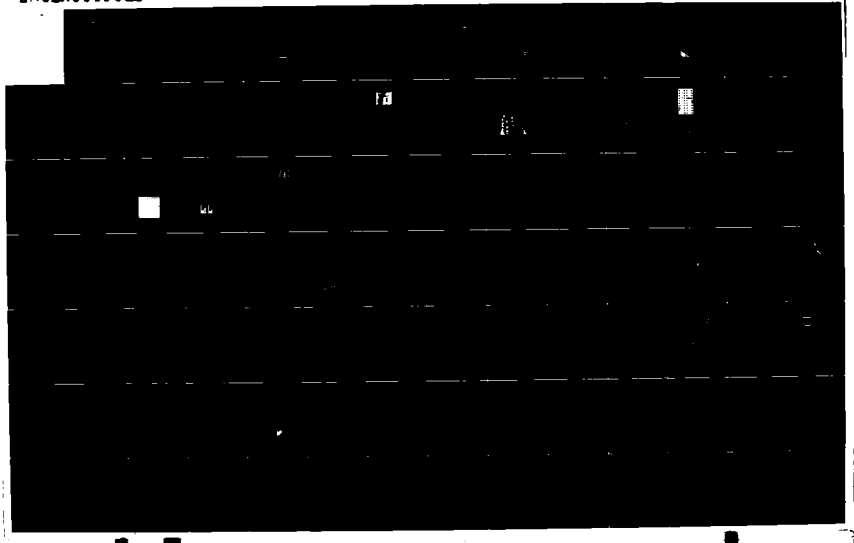
3. The leakage current must be taken into account in the designs using varistor (25μA at 200 V), the trisil has a lower value (100 nA) and the gas arrester has a neglectible value (2.5 pA).

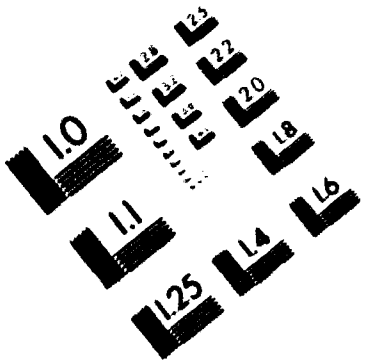
AD-A286 614 EUROPEAN SYMPOSIUM ON RELIABILITY OF ELECTRON DEVICES FAILURE PHYSICS AND..(U) GAREL ASSOCIATES NORTHAMPTON (UNITED KINGDOM) G M BRYDON 7 OCT 94 XB-ONREUR

6/7

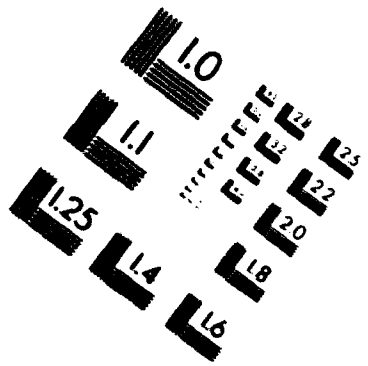
UNCLASSIFIED

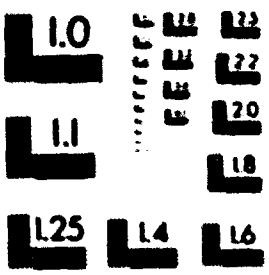
NL

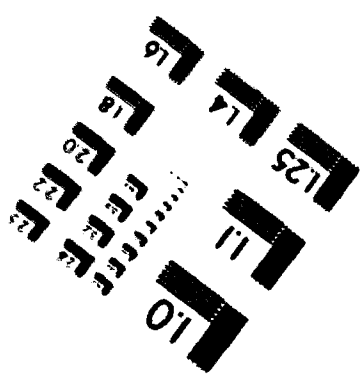


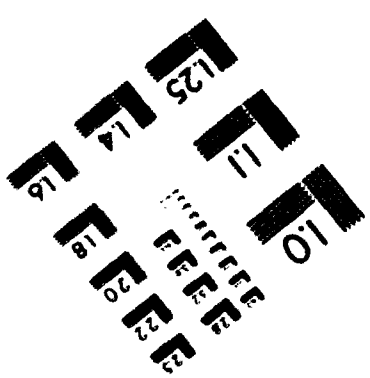


Association for Information and Image Management









MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

COMPONENT ROBUSTNESS: THE KEY TO REDUCING FIELD FAILURES

Finn Jensen
Finn Jensen Reliability Consultancy ApS, Pile Alle 11, DK-2840 Holte, Denmark

Abstract

This paper examines the role of the load-strength approach as applied to electronic components. It is argued that if field failures are dominated by extrinsic, random events, the components will display a constant hazard rate that depends on two parameters. One parameter is the rate of occurrence of the failure causing load, i.e. a purely event related parameter. The other parameter depends on the amplitude of the extrinsic load and on the strength of the component towards such loading. Component strength is defined in terms of a new parameter called component robustness. It is argued that field reliability and component robustness are closely related. This opens avenues for assessing component reliability by testing for strength or robustness as an early qualification procedure.

1 INTRODUCTION

The controversy regarding the question of whether electronic components exhibit a constant hazard rate or not in their useful life period is still unsettled. The closely connected question of the validity or even desirability of the standard reliability prediction methods is likewise undecided. This paper does not pretend to settle the controversies, but it sets out to establish some useful ground rules concerning component reliability and the most likely causes of field failures. The paper introduces the concept of component robustness and suggests that real life failures are closely connected to this parameter. Increasing component robustness will increase field reliability.

There is nothing sophisticated about the approach, but it may help in resolving some of the problems surrounding reliability prediction. Most importantly, it can help in developing and establishing new approaches to component qualification and testing with special relevance for component reliability performance in the field. The models are only first order models, and no precision or completeness is intended at this stage.

2 INTRINSIC AND EXTRINSIC FAILURES

In order to understand the robustness concept it is necessary first to establish that there are basically two underlying causes of component failures, intrinsic and extrinsic.

Intrinsic component failures are those connected with the component

- design
- materials
- processing
- assembly
- packaging

and provoked under circumstances within the component's design specification. The inclusion of the design or data sheet specification is particularly important. As a simple example you might consider a ceramic multilayer capacitor specified to operate within 64 volts under certain temperature and humidity conditions. If a failure should occur whilst the capacitor is operating within rated conditions, and if it can be ascertained that the capacitor during handling and circuit board assembly has been treated within the manufacturer's specifications, such a failure will be termed intrinsic. It is a problem of component design or manufacture.

If, on the other hand, the component failure is related to

- electrical overstress (EOS)
- electrostatic discharge (ESD)
- rough handling at higher levels of assembly
- overload (mechanical, thermal, chemical)
- misapplication (wrong component for the job, - a circuit design problem)

it will be termed an *extrinsic failure*.

A small point: ESD can be either an intrinsic or an extrinsic problem. If the failure occurs due to problems at the component manufacturer's plant, it is

an intrinsic ESD problem. If the ESD problem is caused by handling conditions beyond specified limits at higher levels of assembly (e.g. printed circuit board assembly), it is an extrinsic problem. In fact, it is quite valid to state in general that all non-intrinsic failures are extrinsic.

Most failures in the components long-term wearout period will be intrinsic, if the component is used properly in the circuit design. The failures that occur during early life, the infant mortality failures, can be either intrinsic or extrinsic. If component quality assurance is effective, possibly including a burn-in procedure, although not necessarily so, the intrinsic infant mortality population should be non-existant or at least very small. However, during assembly manufacture, handling and testing, components may be latently damaged, for example by ESD or mechanical overloading. Whilst the components may perform correctly during circuit testing or equipment check out, the induced damage can nevertheless be such that the component fails prematurely under field use conditions. It is then an extrinsic infant mortality failure.

This leaves us now with the failures in the useful life period. Are they intrinsic or extrinsic? They can be either or, but a number of recent case studies seem to indicate that *the majority of the recorded useful life period failures are extrinsic*. If this indeed is the case, there is all the reason to concentrate on improving the ruggedness of the components to cope with the real causes of field failures, the extrinsic problems, whatever they might be.

The following discussion concerns only the useful life period of electronic components. It is assumed that most of the failures are extrinsic and in essence caused by overloading of the component, i.e. the load is greater than the component strength at that particular point in time. This does not mean to say that all useful life failures necessarily are extrinsic, but the contribution from the intrinsic component reliability is considered negligible.

3 THE LOAD-STRENGTH CONCEPT

The load-strength or stress-strength concept has gained considerable interest over recent years, also when dealing with electronic component reliability. The basic tenet is that component failures are caused by a load exceeding component strength. Component strength is in fact *defined* as the value of the load which causes the component to fail. Whilst several different load types in principle can be taken into account, the following explanation considers only a single load type at a time, such as voltage, temperature, mechanical tension, etc.

An example of a strength distribution of a thin film capacitor with respect to electric field is shown in Fig 1. The mass population is the one of interest, as any weak samples will be assumed to have been eliminated before the components are taken into use, or they will have failed as infant mortality failures.

Fig 2 shows a smoothed strength distribution and a narrow load distribution, which represents the expected loading conditions under field use. Also shown is a single value denoting the data sheet maximum for this particular component. The important lower tail of the strength distribution (for example the 3-s or even the 6-s lower limit) defines what we will term the destruction level or the design capability. If loads are less than the design capability, no failures will occur. As long as the strength can be assumed not to degrade over time (a reasonable assumption for electronic components, when we are looking at the first few years of useful life), failures will not occur in the situation illustrated by Fig 2, and the component exhibits 100% reliability or a hazard rate of zero, i.e. $\lambda=0$.

Let us now proceed with the following two assumptions:

1. The strength does not degrade over time (as mentioned above).
2. The field conditions generate freak loads in a random manner.

A *freak load* is defined as one that is greater than what can be expected within reason under the field use circumstances of the equipment under consideration. The situation is described by Fig.3. The freak loads are considered *event related*, and their time of occurrence is unpredictable, random. They will not normally exceed the destruction level of the component, but extreme events such as for example strokes of lightning or careless testing during maintenance can easily generate internal voltages that destroy one or more components.

It is not difficult to compute the probability of component failure, when the components are subjected to *freak* loads of a known amplitude distribution. Let us assume that this has been done, and let us denote the probability of a component failing upon a single freak load application as p_f. Let us further more assume that the random occurrences of freak loads (freak events) can be described by a Poisson distribution with the rate of occurrence denoted ρ

If the assumptions above are fulfilled, it may be shown that the components reliability becomes exponential (Ref 1):

$$R(t) = \exp[-p_f \rho]\, t$$

In other words, the component exhibits a constant hazard rate:

$$\lambda = p_f \rho$$

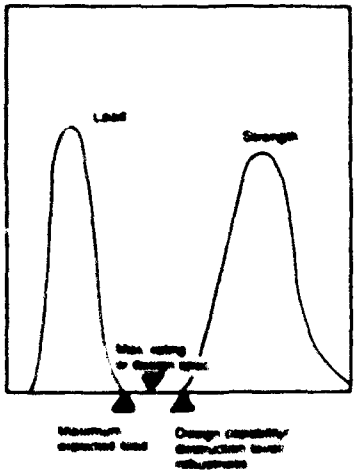


Fig. 2 Load strength distributions and important parameters

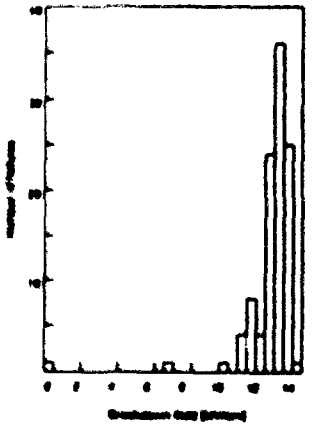


Fig. 1 Example of a strength distribution (Ref. 1).

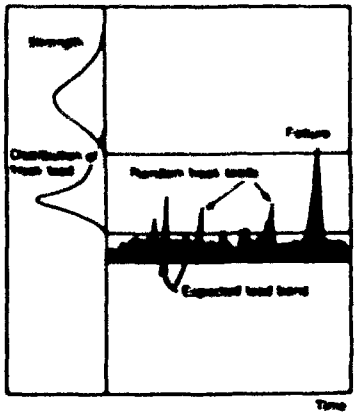


Fig. 3 Illustration of component strength and freak loads.

The interesting part is that the constant hazard rate is not a pure component *attribute*, but depends on the very existance of freak loading events. If there are no freak loads, the hazard rate is zero. If, however, freak events do occur, then the hazard rate is made up of two parts, i.e. the pure event related part, i.e. the *rate of occurrence*, ρ, of freak loads, and the probability p_F, which is a parameter that depends on the freak load distribution (amplitudes) *and* the components strength distribution, especially of course the lower tail. The hazard rate is a useful *reliability parameter*, when it is used intelligently. It is sometime useful to think of the hazard rate as a parameter that expresses our inability, or limitation, to measure the real causes of component reliability in the field, i.e. the strength distributions and the random, freak overloads.

4. COMPONENT ROBUSTNESS

Let us now define component robustness more stringently:

Component robustness is defined with respect to a particular load parameter or set of load parameters. It is the lower 0.1%-limit of the corresponding strength distribution for a component that comes off a production process that is in statistical control, compared to an industry accepted standard for that particular technology and packaging process.

Robustness is thus a set of non-dimensional values expressed as ratios or percentages. A robustness figure of 100% is an acceptable value, but the higher the value the better, the more robust the component. Robustness addresses a multitude of environmental and operating conditions. Industry standards do not exist at the time of writing, although several working groups, for example within the CECC, are discussing the application of robustness and its inclusion in future standards and recommendations. In the definition above the lower 0.1% tail was used, but any other low percentile can be equally acceptable.

At this time we will use the robustness concept mainly to help us understand better, qualitatively and quantitatively, the reliability performance of components in field use.

To further describe the envisaged development of component robustness it is useful to look at the concept in the context also of existing standards for component *qualification testing*, such as MIL-STD-883, MIL-STD-750, IEC-68, CECC, or others. A qualification test has the intention of establishing, in a general sense, a component's fitness for use. It will normally be performed on an early sample of production line components, the type qualification. Some of the tests will also be performed on an on-going sample basis to demonstrate continued process capability. The tests are devised to verify that a component will perform as expected in its final environment, i.e. with the end customer. The documents are very detailed. We will confine ourselves to examining parts of MIL-STD-883 (Ref.2) "Test Methods and Procedures for Microelectronics" to illustrate how these documents can be expected to form a useful foundation for tests leading to the evaluation of robustness.

In principle, MIL-STD-883 applies only to microcircuits that comply with specific U.S. MIL standards. The document has, however, found widespread use also in the commercial sector. Most quality and reliability handbooks from component vendors refer to tests according to MIL-STD-883. The introductory part of the standard introduces the scope and purpose of the document. Although lengthy, the first paragraph includes a useful set of statements relating to the general goals. The following text is excerpted from that section:

1.1 Purpose. The standard establishes uniform methods, controls, and procedures for investigating, testing, classifying and certifying microelectronic devices suitable for use within military and aerospace electronic systems including basic environmental tests to determine resistance to deleterious effects of natural elements and conditions surrounding military and space operations, physical and electrical tests, design, package and material constraints, general marking requirements, workmanship and training procedures, and such other controls and constraints as have been deemed necessary to ensure a uniform level of quality and reliability suitable to the intended applications of those devices.

The statements are in line with the approach to extrinsic component reliability being taken in this paper, and the tests to " determine resistance to deleterious effects of natural elements and conditions surrounding operations" is precisely what component robustness is about.

The tests of the standard are broadly divided into three groups: one group deals with environmental tests,

another with mechanical tests, and a third with electrical tests. For each test the document carefully states the purpose, the apparatus required, the test procedure, including the actual test condition and duration as well as failure criteria, and concludes with a summary. The test conditions, i.e. the loads, are typically "harsh", but are set to emulate the conditions which the components within reason might encounter during handling, transportation, and use. The key words are *within reason*. The loading conditions subject the components to the maximum (within *reason*) conditions they might encounter during circuit board assembly, higher levels of assembly, field use, and all the handling and transportation conditions connected herewith.

A good impression of the type of testing written into these standards can be obtained by looking at the example given below. The example is the lead pull test of MIL-STD-883. It is excerpted from Method 2006.4, Lead Integrity, Test Condition A - Tension.

1. Purpose. This test is designed to check the capabilities of the device leads, welds, and seals to withstand a straight pull.

2. Apparatus. The tension test requires suitable clamps and fixtures for securing the device and attaching the specified weight without lead restriction. Equivalent lateral pull test equipment may be used.

3. Procedure. A tension of 0.227 kg (8 ounces), unless otherwise specified, shall be applied, without shock, to each lead or terminal to be tested in a direction parallel to the axis of the lead or terminal and maintained for 30 seconds minimum. The tension shall be applied as close to the end of the lead (terminal) as practicable.

3.1 Failure criteria. When examined using 10x magnification after removal of the stress, any evidence of breakage, loosening, or relative motion between the lead (terminal) and the device body shall be considered a failure. When a seal test in accordance with method 1014 is conducted as a post test measurement following the lead integrity test(s), monotonic cracks shall not be cause for rejection of devices which pass the seal test.

4. Summary. The following details shall be specified in the applicable procurement document:

a. Weight to be attached to lead, if other than 0.227 kg (8 ounces) (see 3).
b. Length of time weight is to be attached, if other than 30 seconds (see 3).

The tests are essentially attribute or go/no-go tests conducted using loading conditions that are *within reason*. They accept or reject the component being tested, but they are not designed to generate a distribution of component strengths. The standard tests thus give no indication of any margin towards loads greater than those specified, that is "*unreasonable*" loading conditions. They cannot therefore be used to evaluate robustness figures. It is, however, easy to imagine an extension of many of the standard qualification tests, such as the above, to allow for robustness evaluation. This would require testing (step-stress testing) beyond the standard load in order to generate the strength distribution towards the load being investigated. The robustness figure can then be extracted from this strength distribution, if there is an industry agreed standard with which to compare the strength values. Even without an industry standard the *usefulness of establishing a "safety margin"* towards excessive or unreasonable loads should be evident.

The question naturally arises, which loading conditions to consider especially important, if the robustness figures are to relate to extrinsic component reliability. This means that we on the one hand need to look at the conditions that components may encounter during handling and assembly at the equipment manufacturer's plant, and on the other hand look at the loading conditions recognized as potential killers in the field. Some of the adverse loads are fairly easily dealt with. ESD susceptibility, for example, is already dealt with in qualification testing, and it is just a matter of extending the test to find the strength distribution. But how do you test for robustness against soap, contamination and cigarette ash? Mouse nests and shark attacks?

The situation is not quite as hopeless as it might seem at first hand. What needs to be recognized is that many of the bizarre sources of fail are reported from the field produce damaging effects at the circuit board or equipment level, and that the component failures are secondary effects. The presence of soap, cigarette ash, or any other form of contamination will often introduce leakage currents that can bring about malfunctioning of the circuit, possibly with current surges or excess voltages on component input pins. If this is the situation, we need to test robustness against, say, pin voltages. A variation on MIL-STD-883 Method 3008.1 could be a starting point. Mouse nests may incur the same types of failure as

contamination, or they may block a ventilation path thus increasing device operating temperatures. Robustness towards ambient temperature would be an appropriate test. If the test is performed on plastic encapsulated devices, one failure mode is thermal damage to the encapsulation material.

The ideas surrounding robustness measures are still very much in a state of development, and very little testing has been reported in the open literature. Tests for evaluating forces occurring during pick-and-place operations on surface mount devices have for example been recommended, but suitable tests are not yet in place (CECC 00 802, 1992). It is well known that the dimensions of components especially in surface mount technologies can have a direct impact on the component's susceptibility to damage in the manufacturing environment. Masuraman, et.al., 1990 (Ref. 3), for example, report that larger and thicker ceramic capacitors are less robust towards cracking than the smaller ones. A production count of cracked capacitors on an SMT board showed that by far the largest number of failing capacitors were among the 1812 types, whilst only a few of the 1210 types, and none with smaller dimensions, failed. The "strength" of such capacitors in a robustness sense could be the ppm levels obtained in individual production runs.

The strength distribution already shown in Fig 1 indicates the kind of result that is necessary for robustness evaluation. In conclusion to this section we show a strength distribution towards ESD published by Harris (Ref 4), see Fig. 4. The test was performed using the human body model (HBM). Without actually stating a robustness value, Harris reports that the ESD strength is adequate based on limits specified in MIL-M-38510. The robustness is therefore at least 100%.

Irrespective of the components's robustness, we must accept that defects can, and most likely will, be induced during the handling and manufacturing procedures at all higher levels of assembly, i.e. from the printed circuit board to the final system level. Also, the handling, transportation, installation, operation and maintenance procedures from there onwards will on occasion introduce defects in the components. The defects will lower the strength of the

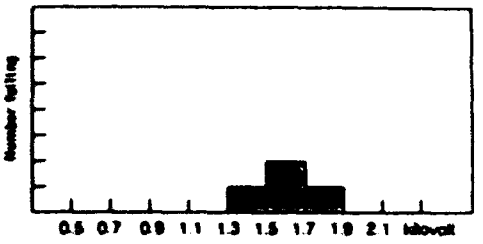


Fig. 4 ESD breakdown data from a component supplier (Ref 3)

components inflicted, see Figure 2, and reliability will inevitably suffer. It is vital to understand, however, that the damage introduced by the external circumstances, i.e. the size of the defects, will be smaller the more robust the component is. The number of early life and useful life failures is thus smaller, the more robust the component.

5. THE COMPONENT MANUFACTURER'S ROLE

It is sometimes stated that a component hazard rate is not a component attribute in line with capacitance, resistance, amplification factor, etc. On the other hand, you might also hear people say "Yes, that is all very well, but if you look at the field results, you must admit that some components exhibit much higher hazard rates than others. Integrated circuits for example nearly always come out with relatively high hazard rates, whereas simple resistors or capacitors seemingly never fail". The implication being that the failure probability or hazard rate, to a degree, is connected to the component technology and complexity.

If we look at the situation in the light of the load-strength and robustness concept, we will see that the apparent incompatibility of the two viewpoints can be resolved.

Statement 1: Component hazard rate is not a component attribute.

Comment: This is only partly true, as may be seen by the expression for component hazard rate $\lambda = p_f \cdot \rho$. However, it is also apparent that in a certain loading environment, where fresh loads must be accounted for, component robustness influences p_f and therefore also λ. What *is* a component attribute, is robustness.

Statement 2: Integrated circuits have higher hazard rates than less complex components.

Comment: This is mostly true. In the framework of the robustness idea it is not surprising that small geometry, fine line circuits of high complexity and comparatively large package sizes are more vulnerable to external loading events than the simpler and often smaller passive circuits and transistors. The margins between standard loading conditions in the field and the destruction level will be different for different technologies, see Fig 5. However, there may well be exceptions, so the statement cannot be taken as a firm rule.

The component manufacturer has, naturally, a significant role to play as far as component field reliability in the useful life period is concerned. In simple terms, the manufacturer that comes out "best" is the one that can supply a component with a greater robustness figure (or figures) than his competitors.

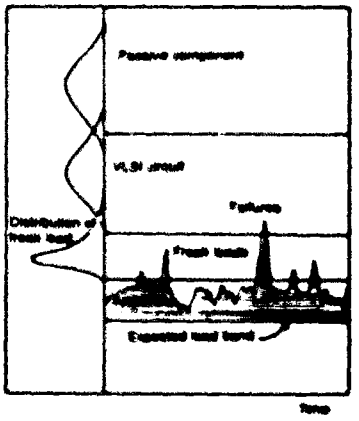


Fig. 5. Illustration of components with different robustness figures and therefore different field reliabilities.

The component manufacturer will still with justification be able to claim that components used outside rated or specified limits constitute a misuse and cannot be covered by any guarantee. The user knows and accepts this. But the user will nevertheless in his own best interest choose the most robust of components, if all other attributes including size and cost are comparable from one vendor to the next. Field reliability will benefit without any doubt.

6. THE EQUIPMENT MANUFACTURER'S ROLE

Irrespective of the details of component robustness, the equipment manufacturer has (and has always had) the opportunity of making a particular equipment or system design less vulnerable to the hazards of the environment and rough handling by incorporating various design measures. The equipment designer is in effect "shielding" his design against the damaging events of nature and human interference. The whole area of environmental testing at the equipment level deals with this aspect.

Designers can protect their equipment against lightning by incorporating transorbers at the circuit board level. They can protect the equipment from dust by encapsulating vulnerable areas in dust-proof shielding. The damaging effects of vibration can be eliminated by using vibration dampers, etc, etc.

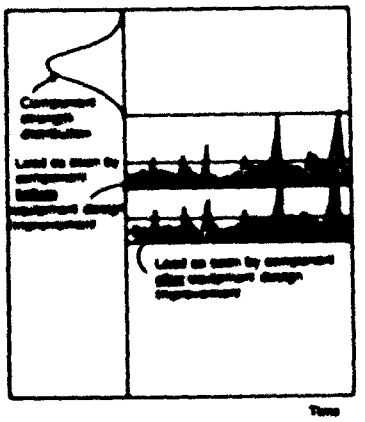


Fig. 6. Equipment designers can reduce the impact of fresh events on the components.

The end result of these various design measures is that freak loads are quenched, thus appearing less harmful at the component level. The peak loads have *far* less chance of getting to the components. Fig 6 illustrates the concept, which results in a smaller p_f and thus a lower field hazard rate.

7 CONCLUSIONS

This paper suggest a *new*, non-exotic approach to assessing the reliability of electronic components in their *useful life period under real life* use conditions. The approach relies on the load-strength concept and introduces the idea of component robustness as an important parameter. *Field failures caused* by external events can be reduced by increasing component robustness.

There are *no industry* standards in *place* at the present time that would allow an immediate evaluation of component robustness. The required tests are however closely *related* to the already existing standard tests used for component qualification. It would seem in the interests of industry and the environmental standards committees both to further develop the concepts in order to provide guidelines and standard tests that in the future can give a realistic appraisal of component robustness and ultimately of field reliability.

References

1. Jensen, F., "Electronic Component Reliability" Book to be published by John Wiley & Sons, Ltd. Chichester, 1995.
2. MIL-STD-883, "Test Methods and Procedures for Microelectronics" Version 883D, U.S. DOD, 1991.
3. Mourmans, R., Hunte, A., Allaert, K., and Stevens, L., "Ceramic Multilayer Capacitors for Surface Mount Applications: A Burning Need for Alternatives" 7th International Conference on Reliability and Maintainability, 432-436, Brest, 1990.
4. Harris CMOS Digital Data Book, Harris Corporation, 1986.

CRITICAL AREA ANALYSIS FOR DESIGN BASED YIELD IMPROVEMENT OF VLSI CIRCUITS

Doris Schmitt-Landsiedel*, Doris Keitel-Schulz**, Jitendra Khare***, Susanne Griep*, and Wojciech Maly***

*Corporate Research and Development
**Semiconductor Division
Siemens AG
81730 München, Germany

*** Dept. of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh PA 15225, USA

1 ABSTRACT

Yield improvements can be achieved by both contamination control (manufacturing) and defect sensitivity decrease (design). In this paper, the need for critical area analysis is demonstrated for design based yield prediction and improvement. Experimental results for a typical CMOS process are provided.

2 INTRODUCTION

In modern IC manufacturing with its increasing competition, short time-to-profit is of prime importance. This requires, amongst other things, that the cost of manufacturing per working product be as low as possible. This cost is determined by die area and yield. Through tightening of the structure sizes the area decreases, but the yield may decrease as well. Thus, for cost efficient product planning the yield of a product in dependance of structure size must be known as precisely as possible.

Increase of the production yield can be achieved by improved process and contamination control. This implies high efforts in installation of better equipment, in-situ monitoring techniques, failure analysis, etc. - all of which increase the cost of manufacturing per die. Therefore, any additional means of increasing yield with less impact on cost have to be exploited.

Another possibility of improving product yield is decreasing the defect sensitivity of the layout by design modifications. This can be achieved through cooperation of different areas of process and design development. As shown in Fig. 1, from computer-aided failure analysis and defect simulation we can derive process and equipment modifications /1/. For layout related analysis, the methodology of critical area calculation /2/ can be used. First, from the analysis of a monitor chip, the defect density distribution of a process line is obtained. Then the layout of product chips is analyzed in order to estimate the yield, to identify critical structures in the design and to optimize the layout for better yield.

In this paper, we report the results of an industrial experiment, which was carried out to demonstrate the sensitivity of manufacturing yield to IC layout. In section 3, the correlation of various structural design features to yield is presented. In section 4 we give a brief overview of the defect and yield models used and the methodology of critical area determination. Section 5 reports the industrial experiment, with results in the determination of

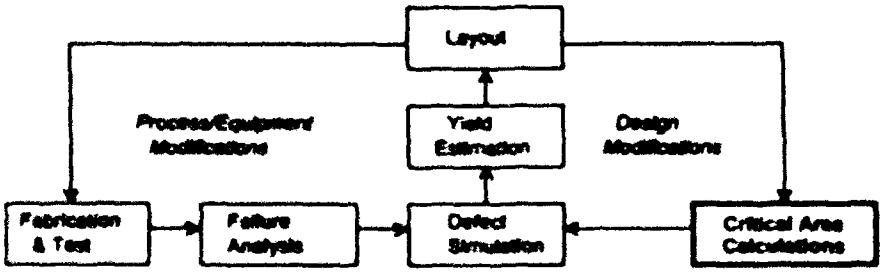


Fig. 1 Cooperating areas for design based yield improvement

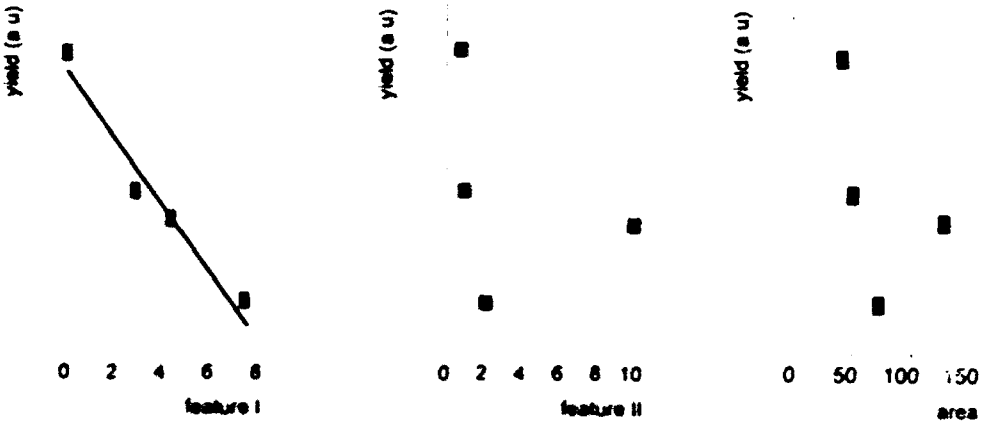


Fig 2 Dependence of yield on three design features for different products. Only feature I shows a correlation with yield

defect distribution parameters and in the analysis of different layouts, demonstrating the sensitivity of the critical area analysis. Finally, section 6 discusses the conclusions that can be drawn from the experimental results.

3 CORRELATION OF DESIGN FEATURES AND YIELD

The influence of various design features on the yield was investigated. Different products fabricated in the same process, which were designed with full custom and/or semicustom methodology, have been examined. Fig. 2 shows the yield of a number of products in dependence of three structural design features. Feature I is the total length of lines with minimum feature size. A high correlation of this quantity with yield is observed. Feature II corresponds to the total number of transistors, a quantity that is strongly dependent on the design methodology. It does not show a strong correlation with yield. The same is true for Feature 3, the total chip area.

This illustrates that area related yield models, as e.g. point defect models, are not suitable. Such models regard defects as dimensionless "points" and calculate IC yield in terms of only the total chip area and the total defect density. Instead, the topology of the different layers of the IC as well as the shape of the defects have to be taken into account. Hence, for accurate yield prediction and enhancement, a realistic model and a method for parameter derivation must be available.

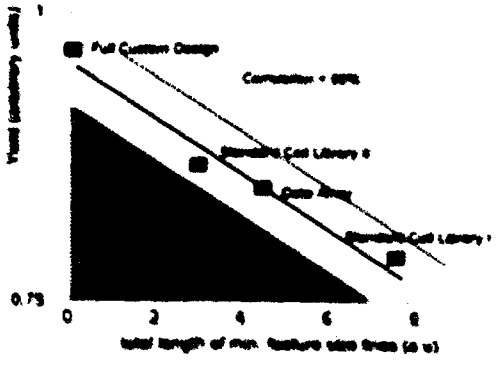


Fig. 3 Dependence of yield on total length of minimum feature size lines (feature I of Fig. 2). Designs with yield losses due to electrical properties are in the shaded area.

Observe that the investigations shown in Fig. 2 can also be used to identify losses caused by parametric failures, once the dominant correlation, like the one for feature I in Fig. 2, has been found. In a stable process we can assume that the main reason for yield loss are spot defects. The yield is then close to the straight line in Fig. 3. If it is much beyond this line and lies in the shaded area of Fig. 3, then the reason for yield loss must be of parametric (electrical) nature. These assumptions were confirmed by practical examples of chips from 10k gates to 80k gates in different CMOS processes. First investigations on bipolar processes also show promising results.

Failbitmap	Faults	Defect types
• BM1: One column stuck at 0	• Bit-Gnd short	• Extra M2 (1µm≤R<3µm) • Missing M1/2 oxide
	• Bitbar-VDD short	• Missing M1/M2 oxide
• BM2: One column stuck at x	• Bit-Bitbar short	• Extra M2 (1µm≤R<3µm) • Missing M1/2 oxide
	• Bit-Bitbar-GND short	• Extra M2 (3µm≤R<5µm)
• BM3: Two columns stuck at 0 / 1	• Bit1-Bitbar2-GND short	• Extra M2 (3µm≤R<5µm)

Table 1 Part of catalog for bitmap - fault - defect mapping

4 MODELS

From the requirements given above, it is obvious that the defect characteristics and critical area functions have to be determined separately for each layer of the chip. The defects are described as disks with radius R of missing or additional material in any conducting, insulating or semiconducting layer of the IC /3/. Extra material may cause a short in a conducting layer or an open in an insulating layer, respectively. The inverse holds for missing material. For each type i of defects, the total density D_{0i} as well as the size distribution $f_i(R)$ is introduced. $f_i(R)$ is given in the form /4/

$$f_i(R) = \begin{cases} \dfrac{2(p_i - 1)R}{(p_i + 1)X_{oi}^2} & 0 < R \le X_{oi} \\ \dfrac{2(p_i - 1)X_{oi}^{p_i - 1}}{(p_i + 1)R^{p_i}} & X_{oi} \le R \end{cases} \qquad (1)$$

X_{0i} is small compared to the minimum feature size. So actually the lower part of the definition of $f_i(R)$ describes the defect density distribution. The two parameters for $f_i(R)$ have to be extracted from experiment.

Based on the defect model, a yield model is used /5/, which takes into account the layout topology

$$Y = \exp\left\{ - \sum_{\text{all layers } i} D_{0i} \int_0^{\infty} f_i(R) \cdot A_i^c(R) dR \right\} \qquad (2)$$

where $A_i^c(R)$ is the *critical area function* for defects of type i. The critical area characterizes the sensitivity of the IC layout for defects of type i. For a defect of radius R, the critical area is defined as that fraction of the chip area on which the center of the defect must fall to create an electrical fault in the IC. The critical area functions can be obtained from the layout of the IC with a suitable algorithm /2/.

Thus for a given fabrication facility (i.e. given D_{0i} and $f_i(R)$ functions) the defect related yield loss of a product completely depends on the critical area functions of the layout.

5. EXPERIMENT

5.1 Determination of defect density distribution

To demonstrate how the parameters of the defect distribution functions can be determined, an industrial experiment was performed. An array of 16k SRAMs was used as a test vehicle /6/. Compared to single layer test structures, advantages of the SRAM are full process topology and efficient

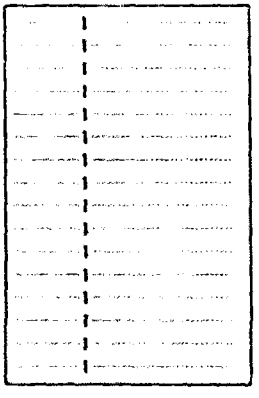


Fig. 4 Part of a bitmap of type BM1, column stuck-at-0

use of area. The SRAM has a topology and failure rate similar to complex logic circuits, but is much easier to test. The faults of the SRAM are characterized by their bitmaps. Defect simulation experiments with this kind of SRAMs for speedup of failure analysis have been described before /1/. A part of a catalog for bitmap - fault - defect mapping is shown in Table 1. Fig. 4 shows part of a bitmap of type BM1, for a column stuck at 0. It can be seen that we do not obtain a one-to-one mapping: BM1 can be caused by 2 different faults, and a fault can be caused by 2 or more types of defects. On the other hand, one type of defect can cause several types of faults and bitmaps. However, experiment showed that more than 80% of the observed bitmaps could be mapped onto one type of defect /5/.

Similar to equation (2), the yield equation for a unique type of bitmap can be calculated:

$$Y^b = \exp\left\{-\sum_f \sum_i D_{0i} \int_0^\infty f_i(R) \bullet A_i^c(R) dR\right\} \tag{3}$$

where f denotes all fault types f causing bitmap b, and i denotes all defect types causing fault f. Now the yield loss due to a unique bitmap can be calculated and compared with measurement results,

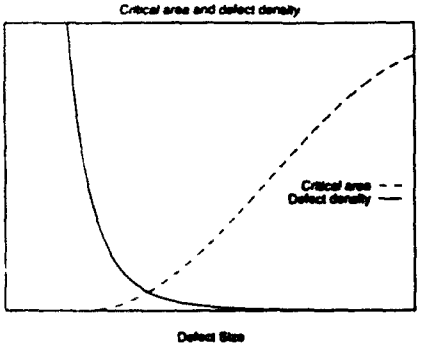


Fig. 5 Defect density distribution for metal2 defects and critical area function (dashed line) summed for all faults caused by metal2 defects

and finally the defect density distribution $f_i(R)$ can be determined /4/. Fig. 5 gives the defect density distribution obtained for metal 2 defects and the critical area function summed for all faults caused by metal 2 defects.

The important region is for defect diameters shortly above the minimum feature size, where the product of defect density and critical area is non-zero. It is interesting to note that for the metal 1 layer, a different total defect density as well as a different form of the defect size distribution were found.

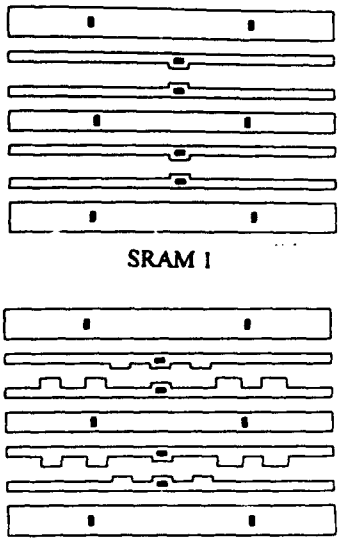


Fig. 6 Part of the layout of metal 2, showing the differences of the two SRAMs

5.2 Comparison of two different layouts

In a further experiment, two electrically identical SRAMs were analysed. The basic cell for each SRAM had an identical area and the same design up to the metal 1 layer. In the second metal layer containing the bitlines of the memory cells, SRAM2 had a larger amount of minimum distance line sections, see Fig. 6. This difference in layout between the two SRAMs meant that the critical area functions for both types of SRAMs were identical up to the metal 1 and viahole layer, but were different for the metal 2 layer. The main difference between the two cells was the layout of the bit and bitbar lines. As a result of this, there was a marked difference in the critical area functions for metal 2 shorts for 2 sets of faults (Fig. 7): (a) shorts between bit and bitbar lines and (b) shorts between bit bitbar and gnd lines. As can be seen in the figure, SRAM2 is more sensitive to these two fault types than SRAM1.

The critical areas for the other fault classes in metal 2 were also different for the two SRAMs. However, the difference was mainly at the larger defect radii and was small enough to be ignored.

As a consequence, the yield of SRAM2 with respect to bitline shorts and bitline gnd shorts (Bitmaps BMA of table 2) was lower, whereas the yield of all

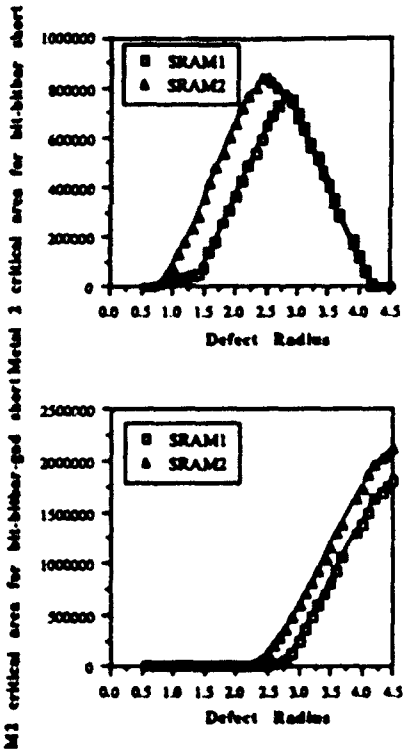


Fig. 7 Metal 2 critical area functions for (a) bit bitbar shorts and (b) bit bitbar gnd shorts for the two SRAMs (defect radius in μm and critical area in μm²)

other fail-bitmaps with metal 2 shorts (BMB) was equal for both kinds of SRAMs. Table 2 gives the measurement and simulation results for the relative yield of the SRAMs.

From this result it is evident that the defect model and critical area concept presented her are sensitive tools for the evaluation and tuning of yield related layout properties.

6. CONCLUSION

The concept of critical area calculation together with precise defect and yield modeling was applied in an industrial experiment. The parameters necessary for yield prediction could be determined for the metal and contact layers. To confirm the dependence of yield on layout topology, the change of yield for a variation in the memory cell layout of an SRAM was simulated and measured with good agreement. Thus a methodology is available for accurate yield prediction and layout analysis. This method can be applied to optimize technology and design for yield.

Yield (SRAM2) / Yield (SRAM1)	measured	calculated
Bitmaps BMA	0.65 ± 0.10	0.5
Bitmaps BMB	1.04 ± 0.12	1

Table 2: Relative yield for two electrically identical SRAMs with different layouts, measured and calculated results and confidence intervals (1 σ). BMA are the failbitmaps caused by metal 2 bitline shorts and bitline gnd shorts, having a different critical area function for the two SRAMs. BMB are all other metal 2 related failbitmaps.

References:

1. S. Griep, et al., "Speedup of failure analysis using defect simulation", Proceedings Europ. Symp. Reliability and Failure Analysis, pp. 577-582, Bordeaux, France, Oct. 1993

2. D. M. H Walker, "Critical area analysis", Proc. of the 1992 International Conference on Wafer Scale Integration, pp. 281-290, Jan. 1992

3. J. Khare, D. Feltham, W. Maly, "Accurate estimation of defect-related yield loss in reconfigurable VLSI circuits", IEEE Journal of Solid-State Circuits, vol. 28, no.2, pp. 146-156, Feb.1993

4. J. Khare, et al., "SRAM-based extraction of defect characteristics", Proc. Int Conf. on Microelectronic Test Structures, San Diego, April 1994

5. W. Maly, J. Deszczka, "Yield estimation model for VLSI artwork evaluations", Electron Letters, vol. 19, no. 6, pp. 226-227, March 1983

6. D. Schmitt-Landsiedel, et al., "Use of a CMOS static memory array as a technology test vehicle", Quality and Reliability Eng. Int., vol. 8, pp. 219-223, 1992

INFLUENCE OF GATE AND TUNNEL OXIDE DEFECTIVITY ON EEPROM YIELD AND RELIABILITY

E.Ciantar (1 2), P.Boivin (1), M.Burle (1) and J.Oualid (2)

(1) SGS-THOMSON Microelectronics, 13790 Rousset, France
(2) Laboratoire des Matériaux et composants Semiconducteurs, ENSPM, Domaine Universitaire de St-Jérome, 13397 Marseille Cedex-20, France

1 INTRODUCTION

Microelectronic manufacturers must produce devices with very good yield. These devices must also have a mean time to failure higher than ten years. At the present time, this objective is more and more difficult to be obtained as the integration density increases continuously, and so the dimensions of transistors (the gate length, the gate oxide thickness,...) are reduced. So the use in production line, of a quick and reliable test capable of predicting yield or reliability problems is very important.

For Electrically Erasable Programmable Read Only Memories (EEPROM), an eventual yield (or reliability) degradation is often attributed to the oxide quality. In this study, this oxide quality is defined by the defectivity determined from the charge injected before breakdown (Q_{bd}) distribution obtained on test MOS capacitors. The Q_{bd} value can be deduced either by the constant current stress or by the exponential ramp current stress methods. Due to its lower consumed time, the latter seems to be well adapted to manufacturers.

In this study, we describe the method which is used to determine the defectivity. Then, we prove that standard probe functional or reliability yields can be correlated to oxide defectivity defined for different Q_{bd} values.

2 EXPERIMENT

The Exponential Ramp Current Stress method (ERCS) [1] is used to determine oxide defectivity. It has several advantages on the other methods:

_ It is a time-saving accelerated test (typically less than 20 seconds per studied MOS capacitors).

_ It explores a large Q_{bd} range (more than 6 orders of magnitude).

_ It is of a particular interest in the identification of early fails. By the other methods used to determine oxide quality, these early failures are often not detected.

The initial current density is 10^{-3}A/cm² and the staircase ramp rate used is 0.1 decade per second. The gate voltage reading is performed just before the end of each step. The breakdown criterion is satisfied if the voltage abruptly shuts down ($V_i < 0.8V_{i-1}$). So, the Q_{bd} measurement is affected by a systematic overestimation, due to the last step time.

The results are plotted in a Gumbel plot which represents the variation of Ln(-Ln(1-P)) versus

$Log(Q_{bd})$, where P is the cumulative probability for capacitors to fail with an injected charge less than Q_{bd}. The Q_{bd} distribution presents generally two parts:

_ The breakdowns observed for low Q_{bd} values are said extrinsic. They characterize the oxide defectivity defined by [2]:

$$D(Q'_{bd}) = -\frac{Ln(1 - P(Q'_{bd}))}{A} \quad (1)$$

where $P(Q'_{bd})$ is the cumulative probability to find capacitors which fail with a Q_{bd} value lower than Q'_{bd}, A is the area of test capacitors.

_ The breakdowns obtained for the highest values of Q_{bd} are called intrinsic. We define the intrinsic Q_{bd} by $P(Q_{bd})=0.9$.

Evidently, it is required that the gate and tunnel oxides used in EEPROM must have the lowest defectivity and the highest intrinsic Q_{bd} value.

3 RESULTS

3.1 Functional yield

We have systematically compared the standard probe functional yield (Y_f) determined after 200 successive write and erase cycles and the gate oxide defectivity $D(Q'_{bd})$, on 27 lots. The defectivity is measured on test MOS capacitors ($A=10^{-3}cm^{-2}$, tox=33nm) processed with the same steps than EEPROM. We have tested five capacitors per wafer and six wafers were randomly chosen per lot. We have verified that the Gumbel plot obtained with these 30 capacitors are practically identical to the one obtained with 250 capacitors.

Figure 1 represents for each lot the distribution of $Log(Q_{bd})$ represented by a box plot representation. The bottom and the top of boxes represent 25% and 75% of the distribution respectively. The bottom and the top of whiskers give 10% and 90% of the distribution respectively. On figure 1, we can observe that some lots present an anomalous distribution related to problems identified or not during the process.

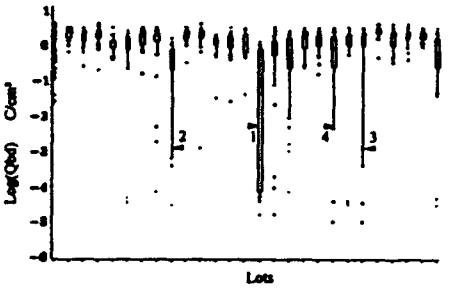


Figure 1: Box plot $Log(Q_{bd})$ of each lot

We have distinguished four lots labelled 1 2 3 and 4 which present a singular distribution compared to the other lots. We have plotted, in figure 2, the Gumbel plots of these four lots.

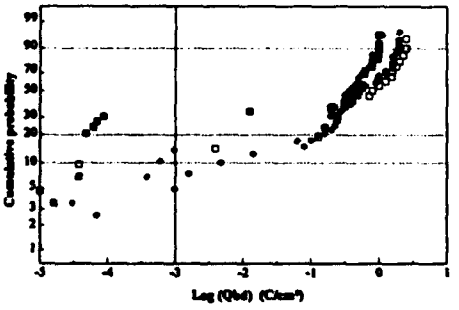


Figure 2: Gumbel plots of the four selected lots : 1 (fill square), 2 (diamond), 3 (star), 4 (open square)

In order to correlate $D(Q'_{bd})$ with the standard probe functional yield, we must chose a Q'_{bd} value as

low as possible with the condition that the number of capacitors which fail with Q_{bd} lower than Q'_{bd} is significative. For the used process, the area and the number of studied capacitors, we have chosen $Q'_{bd}=10^{-3}C/cm^2$. We compare, in table 1, the intrinsic Q_{bd}, and the defectivity with the standard probe functional yield (Y_s), for the four selected lots

Lots	1	2	3	4
Q_{bd} (C/cm^2) @ P=0.90	1	1	2.9	3.5
$D(10^{-3}C/cm^2)$ ($/cm^2$)	370	150	50	100
Yield (%)	55	87	92	89

Table 1: Intrinsic breakdown, defectivity and yield of the four selected lots

It appears that the yield is correlated to extrinsic defectivity and not to intrinsic Q_{bd}. A second order polynomial expression gives a good fit of the empirical relation between Y and $D(10^{-3}C/cm^2)$. This relation is representative of the design of the memory cells and of the used process. No correlation was observed on some other lots. For these, we have verified that the eventual yield degradation is associated to identified problems independent of the gate oxide quality (like particles, contaminations, parametric problems, ...). It is possible to write for all lots $Y \leq a D^2 + b D + c$ (with $a = -2.8\ 10^{-4}cm^{-4}$, $b = 5.6\ 10^{-3}cm^{-2}$, $c = 93$).

These results indicate that the standard probe functional yield is reduced when the gate oxide defectivity, defined for $Q'_{bd}=10^{-3}C/cm^2$, is higher than $50cm^{-2}$.

3.2 Reliability yield

The same methodology was applied to qualify some processing steps, and particularly the polysilicon deposition. Two lots labelled A and B of EEPROM devices were identically processed except for the polysilicon gate deposition carried out in two steps. A first polysilicon film was deposited (thickness of 50nm) in a LPCVD reactor with a very slight flow of nitrogen for lot B and without gaz flow for lot A. Then the wafers were etched in a HF solution (1%) during 30 seconds. The final gate thickness (450nm) has been obtained by a second polysilicon deposition with the same conditions for the two lots.

We compare, in table 2, for both lots, the standard probe functional yield Y_s of devices determined as previously, the reliability yield Y_R obtained after a write operation followed by 10^4 readings, the intrinsic charge injected before breakdown Q_{bd} (P=0.9) and the defectivity D measured on test MOS capacitors with gate oxide (G_capacitors: t_{ox}=33nm, $A=10^{-3}cm^2$) or tunnel oxide (T_capacitors: t_{ox}=9nm, $A=2.4\ 10^{-4}cm^2$). The Gumbels plots for the two types of capacitors and for each lots are given in figure 3.

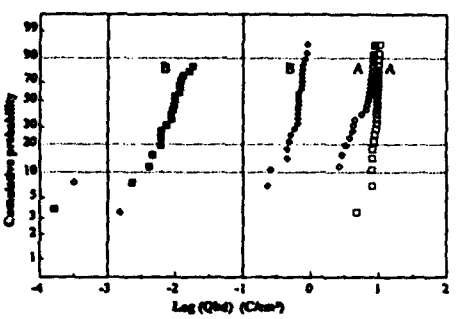


Figure 3 : Gumbel plots of gate (diamond) and tunnel (square) of lot A (open symbols) and lot B (fill symbols)

We can observe different points:

_ For the lot A, the intrinsic charge injected before breakdown is slightly higher for T_capacitors than for G_capacitors. It is well known that the tunnel oxide (t_{ox}<10nm) better resist to F.N. electron injections than the gate oxide (20nm<t_{ox}<40nm). This observation can also be justified by the larger area of G_capacitors. Both conjugated reasons could also explain why the breakdown distribution of T_capacitors are less dispersed than the one of G_capacitors.

_ The T_capacitors of lot B present two distinct breakdown distributions. The distribution which corresponds to lower charges injected before breakdown (between 10^{-3}C/cm² and 10^{-2}C/cm²) presents a smaller dispersion than the usual extrinsic distributions observed in figure 2. It can be assumed that the breakdown for these Q_{bd} values is initiated by a well-defined type of defect which occurs with a great probability for the lot B (more than 70% of the whole distribution). The second type of distribution corresponds to intrinsic breakdowns which happen for a charge injected before breakdown only slightly smaller for the lot B than for the lot A.

_ For the G_capacitors of the lot B, we have essentially observed only one breakdown distribution which probably corresponds to the same type of defects presumed to be also present in T_capacitors. The detrimental effect of these defects is more pronounced in T_capacitors due to their thinner oxide thickness. The singular distribution obtained for the T_capacitors is not observed for the G_capacitors because their area is larger.

_ No significative difference was observed on the gate oxide defectivity defined for Q'_{bd}=10^{-3}C/cm² on both lots. This observation can be correlated with the fact that the standard probe functional yields Y_s of both lots are identical.

However, the reliability yields Y_R of both lots are fully different. The memories of lot B are nearly all non functional after the reliability test when the memories of lot A are almost not pertubed. This difference of behaviour could be related to the values of the tunnel oxide defectivity corresponding to a higher value of Q'_{bd} than to the one adopted for the gate oxides. The value Q'_{bd}=0.1C/cm² proposed by P.Cappelletti [3] for the definition of the tunnel oxide defectivity, which is reported in table 2, seems to be well adapted to quantify the differences observed on the Q_{bd} distribution between the two lots. The defectivity D(10^{-1}C/cm²) is unchanged for the gate oxide when it is quite different for the tunnel oxide of both lots in relation with the difference observed on reliability yield Y_R.

Lots		A	B
Gate oxide	Q_{bd} (C/cm²) @ P=0.90	8	0.8
t_{ox}=33nm A=10^{-3}cm²	D(10^{-3} C/cm²) (/cm²)	80	30
Tunnel oxide	Q_{bd} (C/cm²) @ P=0.90	9.3	8.3
t_{ox}=9nm A=2.4 10^{-4}cm²	D(10^{-1} C/cm²) (/cm²)	<125	7900
relative probe functionnal yield		100	100
relative yield after R.T (%)		99	10

Table 2 : Intrinsic breakdown, defectivity for gate and tunnel oxide, relative standard probe functional yield and yield after reliability test (R.T.) for the lot A and B.

In order to find an explanation of these observations, the first polysilicon was removed after the wet etching. We have observed by scanning electron microscopy, only for the gate and tunnel oxides of lot B, a multitude of small and shallow craters, as shown in figure 4.

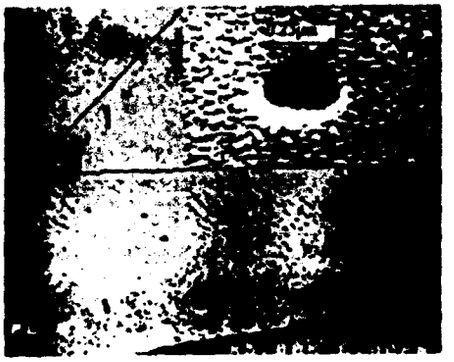


Figure 4: An example of the craters observed after the removal of the polysilicon for the lot B.

It can be assumed that during the wet etching of the lot B, the HF etches partially and locally the gate and the tunnel oxides through pinholes or grain boundaries of the first polysilicon film, contrarily to lot A, where the oxides are not degraded by this etching. These craters could be at the origin of the shifts observed for the breakdown distributions of the lot B and of the reliability yield degradation of this lot.

4 CONCLUSIONS

In this contribution, it appears that the oxide defectivity $D(Q'_{bd})$ is correlated to the standard probe functional yield Y_t and the reliability yield Y_R of EEPROM. The choice of the charge injected before breakdown , Q'_{bd}, depends on the area of the test capacitors and on the oxide thickness. This value of Q'_{bd} must also be different if the defectivity is associated to Y_t or Y_R. We have found for our technology and for our test capacitor area, a good correlation between $D(Q_{bd}=10^{-3}C/cm^2)$ of test G_capacitors and Y_t from one part and between $D(Q_{bd}=10^{-1}C/cm^2)$ of T_capacitors and Y_R for the other part. It appears, consequently, that the ERCS method is suitable to qualify tunnel and gate oxides.

Refereaces

[1] P.Cappelletti, P.Ghezzi, F.Pio, C.Riva Proc IEEE, International Conference on Microelectronic Test Structure (ICMTS) vol 4, N°1, p.81 (1991)

[2] C.M.Osburn, D.W.Ormond J.Electrochem Soc., 119, p.591, (1972)

[3] P.Cappelletti, A.Panchieri, L.Ravazzi Proc ESREF'93, p.77 (1993)

RELIABILIY TEST STRUCTURES FOR ASIC TECHNOLOGY ASSESSMENT

P. Weiss, Swiss Center for Electronics and Microtechnology, Inc., Switzerland
B. Johlander, Components Division, ESTEC,
J. King, GEC Plessey Semiconductor Ltd., England

ABSTRACT

A set of test patterns to qualify the reliability of ASIC technologies has been specifically developped and implemented. Results on bulk CMOS and SOS test lots show that the proposed approach yields a valuable tool.

1. INTRODUCTION

For space systems, Application Specific Integrated Circuits (ASICs) offer key advantages in reducing power consumption and weight. Since ASICs are manufactured in very small series, traditional reliability assurance test methods for large volume electronic components are costly to apply. Alternative methods based on test structures have often been proposed. They are commonly used by semiconductor manufacturers. However, results are usually considered commercially sensitive and therefore frequently unavailable to customers.

The European Space Agency (ESA) has studied the concept of using a set of test structures for assessing a given IC technology instead of considering each particular ASIC. One advantage of such a method is that it allows to apply statistical control methods in order to assure that no reliability hazard has been introduced by the manufacturing process. The results from such tests should form the basis for reducing or completely omitting some lengthy life-tests. This should allow both reducing the parts cost and increasing the confidence in the reliability of ASICs used in space projects.

To implement that concept, a method based on a test chip has been chosen. CSEM, in close collaboration with the ESA and GEC Plessey Semiconductors, has designed such a chip, which has then been integrated and evaluated, both in the CMOS and SOS technologies. This test pattern set, together with the corresponding mesurement methods, provides an useful tool for ASIC technology assessment.

Figure 1. The TAR layout.

2. TEST PATTERNS FOR RELIABILITY

Reliability assessment is normally performed at the product level. To work at the technology level, three different approaches may be used:

- The yield approach
- The physical approach
- The "standard product" approach

The yield approach is based on the idea that yield is obviously related to quality, and that quality is related to reliability. It must then be possible to obtain some information on reliability by measuring the several partial yields which characterize a technology. This sound idea suffers from some drawbacks. In particular, it requires the the disclosure of partial yield data, which may cause some confidentiality problems.

The physical approach is based on the knowledge of the several ASIC failuremechanisms. Small circuits (the "test patterns") are then designed in such a way that they are very sensitive to one particular effect. Of course this relies on the fact that all the possible mechanisms are taken into account. Unsuspected potential failures will not be detected in such a way.

Finally, the "standard product" relies on a well-known IC from the manufacturer's catalogue: in such case any behaviour variation is due to the technology. This has the advantage of beeing realistic, but it is quite specific to a particular circuit, and furthermore the causes of a given failure may be difficult to analyse, due to the complexity of the design.

Thus it may be seen that all these three possibilities have their own advantages and drawbacks. Therefore a global approach, as the one proposed here, should resort to devices from all these three families.

From a general point of view, any reliability test chip should fulfil the following requirements:

- The set of used test structures and measurement methods must provide enough information about all process induced failure mechanisms.
- To gain acceptance, the method should not imply disclosing commercially sensitive information.
- Since yield and quality of a semiconductor process are often related, large, statistically significant, structures should be included.
- The measurements on the chip should be performed to as large an extent as possible at wafer level rather than with packaged devices.
- Besides the chip itself, guidelines for assessment of test results should be available for users.

3. WORK PERFORMED

CSEM had formerly developed a set to study the possible physical failure mechanisms (Ref.1). These existing test structures were first critically screened. New patterns aimed at measuring partial yields (long chains etc.), dynamic properties and stress or failures induced at assembly stage were added. This test set definition was performed in close collaboration between ESA and CSEM. The chosen set is "over complete" in the sense that it includes a number of analytical structures that is normally not used for standard technology assessment but could help in pin-pointing potential problems. Some structures related to radiation tolerance (total ionising dose) were included but specific structures for e.g. Single Event Upsets (SEU) were excluded since they strongly depend on the design of a circuit.

Then, modular designs were proposed, both for a standard 1.5 micron double metal/single poly CMOS process (TAR design) and for a 1.5 micron SOS process (TARSOS design). After design review, CMOS wafers were processed at CSEM while SOS wafer were manufactured by GEC Plessey Semiconductors.

Finally, the completed wafers were characterized with special attention to new structures introduced in the set. The results were then analysed against the objective of possible improvements in reliability assurance methods. Suitable measurement procedures, adapted to each pattern, were established, taking advantage of the observed behaviours.

4. DESCRIPTION OF THE SET OF TEST STRUCTURES

Figure 1 shows an outline of the complete TAR die (8 x 10 mm^2) with the modules using one standard pad-layout to facilitate wafer probing of most structures. Structures which do require packaging have been grouped on the die to facilitate dicing.

The structures were grouped in "analytical" and "physical" sets. The final circuit includes some 100 elementary structures connected by approximately 620 (VANTAR) or 860 (TARSOS) pads for probing or bonding.

From the statistical point of view, several structures with a significant number of samples have been designed for partial yield analysis of continuity or insulation between various layers. These structures are based on the well-known "track-comb"pattern (Ref.2). They involve severe topology and minimum design width to assess the most critical cases of any circuit design. Each pattern comes in four different sizes with area ratio 1, 2, 4, and 200

(Ref. 3). The largest patterns are 900·900 μm^2, which means that they correspond to a realistic case.

The die contains structures to assess several physical phenomena:

- Electromigration, with arrays of parallel metal lines, with and without steps
- Corrosion, with pairs of parallel metal lines
- Oxyde breakdown, with a family of MOS capacitor having all possible edge conditions
- Leakage currents, with a family of diodes, having once again all possible edge conditions
- Hot-carrier degradation, with paired MOS transistors (see figure 2). The idea is to stress one member of the pair, and to keep the other one for reference.

Some simple patterns have also been includes to explore latch-up and electrical overstress effects, whereas these phenomena do depend more on the circuit design than on its technology.

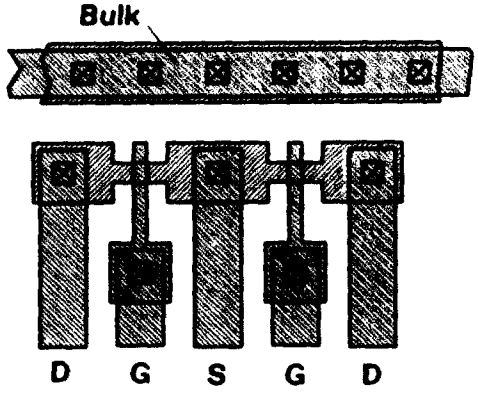


Figure 2, Paired MOS transistor for hot-electron effect characterization.

For the sake of completeness, a number of packaging related problems have been taken into account:

- Stress, which may be estimatode with a strain rosette
- Metal line displacement at the chip edge, detected by possible crack in a thin polysilicon line running along a broad metal band
- Silicon or sapphire edge cracking (due to dicing), detected by a thin metal wire running all around the chip.
- Surface pollution, detected by special parasitic MOS transistor with incomplete gates.

Furthermore, very simple circuits have been introduced as standard devices:

- Ring oscillators
- Inverter chains
- Operational amplifiers

Finally, numerous analytical patterns aim at an in-depth knowledge of the technology, allowing in particular to gauge its behaviour under irradiation:

- MOS transistor matrices
- CV capacitors
- Gated diodes
- Edgeless MOS transistors
- Linewidth measurement structures
- Sheet resistance measurement structures

5. SOME RESULTS

The purpose of the measurements performed on the test lots was to define realistic measurement prescription, and to check the soundness of the proposed patterns.

Several interesting results were obtained. For instance, figure 4 shows the wafer yield for metal 2 to metal 2 shorts on one given fabrication lot. No signficant dependency in the topography may be observed.

Electromigration was measured for both metal layer by resorting to an array of 50 lines stressed in parallel (Ref. 4). A typical result is given in figure 5. The mean time to failure - under accelerated conditions - can be seen to be 2 hours for metal 1 and 90 hours for metal 2. These data compare favorably with results from equivalent technologies.

As a last example, figure 6 gives the hot-electron behaviour characterized with a method described by C. Hu (Ref. 5): a lifetime is defined as the time required for the threshold voltage to shift by 10 mV when the transistor is stressed at constant bulk current. Because of fundamental physics considerations, a log-log plot of the lifetime versus the bulk current should yield a straight line with a slope of -2.9. This is indeed the case as can be seen on figure 6. The bulk current required to get a given lifetime can be therefore be used as a measure of the sensitivity of a technology to hot electron effects.

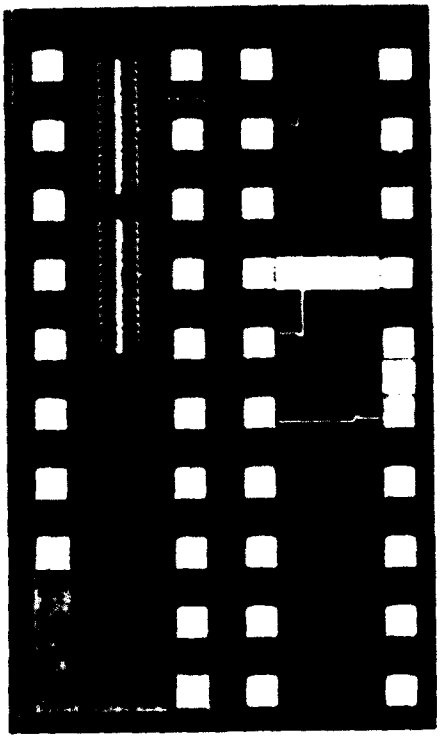


Figure 3. Detail of TARSOS die Package related structures including strain rosette, line displacement detector, edge sensor for chip flaking, and structures for detecting surface ion contamination

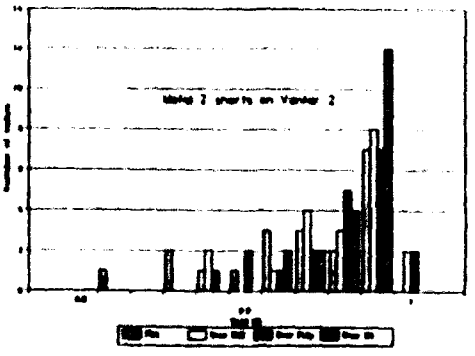


Figure 4. Wafer yield histogram for metal 2 to metal 2 shorts (CSEM data, test patterns 900 X 900 μm^2).

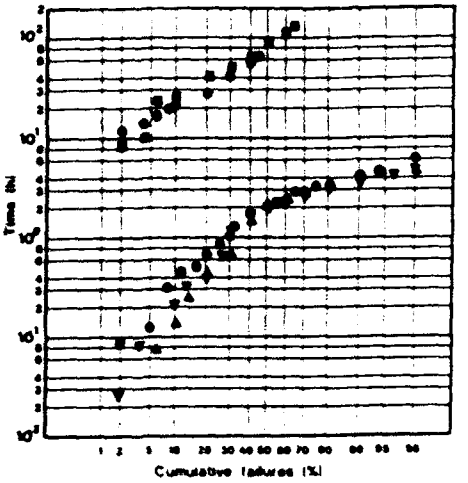


Figure 5. Electromigration failure distribution as measured on a CMOS technology. The upper set of data points are time to failure for Metal 2 lines, (2 samples, 3 µm wide lines) while the lower set is for Metal 1 lines (3 samples, 2 µm wide lines). Stress condition is 150 °C, $1.5*10^6$ A/cm^2. CSEM data

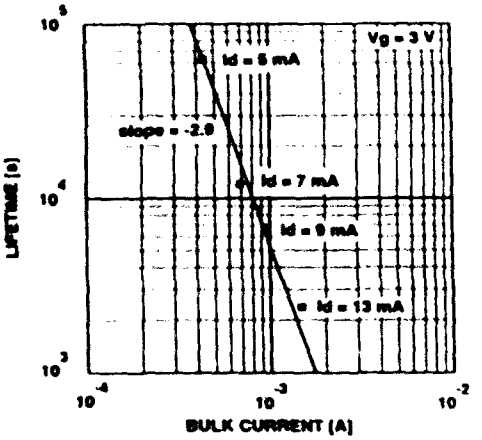


Figure 6. Device lifetime due to hot electron effect vs bulk current. CSEM data

6. CONCLUSIONS

The work performed in this project has shown that the developed test circuit can give very important and useful information about the reliability of a given IC technology. The test circuit used in combination with the measurement and assessment methods developed can form a valuable contribution to the ongoing effort in ESA to improve the efficiency and increase the confidence in the procurement of high reliability ASICs for space applications.

However, relying solely on test structures would have the limitation that they only can address well identified failure mechanisms. Obviously, if a new or unexpected phenomena affects a product, it is not certain that it would be detected. Another area of caution is the correlation between failures on test structures and failures on ASICs. Since this set of test structures has been designed to be particularly sensitive to certain degradation mechanisms, failures in the test structures are likely to occur. However, correlating such failures to field behaviour of a particular ASIC is far from easy, because, among other reasons, the precise cause of the ASIC failure is in general not known.

Nevertheless, the proposed test set, in combination with other "tools" such as audits, line inspections, Total Quality Management (TQM), Statistical Process Control (SPC), failure mode analysis etc., should allow a significant decrease of the required effort.

The recommendation from this work is, as a first step, to apply statistical control methods to the results on the test structures manufactured together with ASICs, i.e. only measure relative "reliability" from one lot to another and establish pass/fail criteria based on the statistical variation of parameters. Data will be stored in a database. Detailed correlation with parts could then later be established based on larger samples. The information in the database would then be used to draw up specific criteria. For successful acceptance of test structures used in the context of this work, complementary classical tests at the product level will probably be needed in the procurement of ASICs for ESA projects.

REFERENCES

1. Weiss, P. and Adams, L. "Development of Semiconductor Test Structures for Reliability Evaluation" Proc. of the 1990 ESA Electronic Components Conference, p 15

2. Mitchell, M. A. "Defect Test Structures for Characterization of VLSI Technologies" Sol. State Technol, May 1985, p 207

3. Mitchell, M. A. et al "Standard Error in Die Yield Projections from Defect Test Structures" Proc. of the 1989 ICMTS Edinburg, 1989, p 189

4. Thomson, C. V. and Cho J. "A New Electromigration Testing Technique for Rapid Statistical Evaluation of Interconnect Technology" IEEE El. Dev. Lett. 7, 12, 1986, p. 667

5. Hu, C. et al. "Hot Electron Induced MOSFET Degradation - Model, Monitor and Improvement" IEEE Trans. El. Dev. 32, 2, 1985, p 375

FAILURE-ANALYSIS-BASED TEST CHIP DESIGN FOR QUICK YIELD IMPROVEMENT

Chisato Hashimoto and Shigeru Nakajima
NTT LSI Laboratories
3-1 Morinosato Wakamiya,
Atsugi-shi, Kanagawa, 243-01 Japan

ABSTRACT

A new design approach for a test chip developed to shorten the debugging cycle time in fabrication is described. This approach meets the requirements for failure analysis as well as parametric and statistical analyses. Particular attention is devoted to accurate defect density estimation and locating individual defects. A specially designed test chip, named YTEG, was used to evaluate 0.5-μm CMOS process technologies, and as a result the effectiveness of the chip has been confirmed.

1. INTRODUCTION

Test chips with a variety of test structures have been widely used to solve debugging process problems (Refs. 1, 2). However, with the increasing level of process sophistication in today's LSI's, obtaining an exact knowledge of what causes failures has become quite difficult using conventional test structures. This difficulty is especially serious in the final stage of process improvement and the production stage to follow, because in these stages basic problems have already been solved, and only complicated ones remain. Each individual failure needs to be precisely analyzed to debug these problems. Conventional test structures, however, lacked this ability, because they attached more importance to parametric and statistical analyses than to failure analysis. To overcome this disadvantage, we developed a specially designed test chip that meets the requirements for not only parametric and statistical analyses but also failure analysis.

In this paper, the strategy of the test chip design we have adopted is described in section 2, and then the actual test chip design procedure is presented in section 3. Finally, some examples of experimental results indicating the effectiveness of this test chip are presented in section 4.

2. STRATEGY OF TEST CHIP DESIGN

Optimal test structures vary with the stage of process improvement, and this factor must be taken into consideration when designing test chips. In this study, we intended to develop a test chip useful for the final stage of process improvement and the following production stage. The roles of a test chip in these stages are described in Fig. 1. The test chip data for individual yield-loss factors with high accuracy obtained through electrical measurements and failure analyses are used for two purposes. One is the yield prediction of real LSI chips.

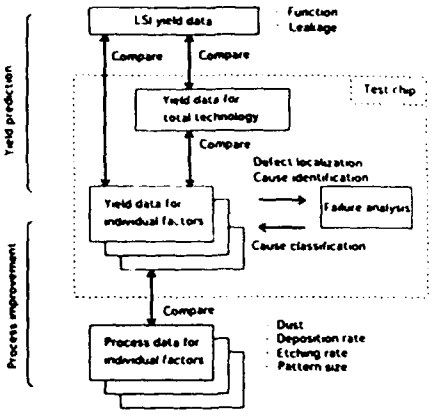


Fig. 1 Yield evaluation using test chip.

For this purpose, a yield model for each yield-loss factor is constructed using the respective fault distribution in a wafer. Then, using the model, each factor's contribution to real chip yield is inferred by comparing the yield of each individual factor and that of the total technology evaluation structure or that of a real chip. The other is the debugging of process problems concerning with each yield-loss factor. For this purpose, the intrawafer fault-distribution data for each individual factor are compared with those for the relevant process steps. Through this study, the process steps in trouble are inferred, and the data are fed back to process engineers. By repeating this process, the total yield is improved. To get the test chip to fully play these roles, the following strategy was adopted for designing it.

2.1 Meaningful test structure

Conventional test chips usually contain a full set of test structures for obtaining parametric data. Considering the necessity of having a rather large area for each structure to evaluate LSI yield accurately, however, this method is not suitable for our chip. Only the most effective test structures were designed based on extensive failure analyses of real chips to prevent missing the structures associated with dominant failure modes.

2.2 Accurate defect density estimation

Multiple defects in a structure must be detected

individually. Otherwise, defect density estimated from statistical test structures includes some ambiguities because of the lack in clustering information. Since it is difficult to solve this problem only by electrical measurements, the test structures were designed to be suitable for both electrical measurements and failure analysis.

2.3 Ease in locating defects

Exact knowledge about failures shortens the debugging cycle time. Therefore, optimal methods for locating defects are to be applied independently or in combination depending on the objective failure mode. The design and size of each test structure was decided according to the defect-locating method chosen.

2.4 Parametric data monitoring

As parametric errors are not assumed in this study, only small test structures were included to monitor parametric data such as threshold voltage, sheet resistance and contact resistance. They were placed in the scribing lines between chips (where the dicing saw cuts), and were completely the same as those in real chips. Placing common monitoring structures enables us to easily detect unusual process conditions.

3. ACTUAL DESIGN PROCEDURE

An actual test chip was designed to examine an SOG gate array process with 0.5-μm-design-rule, 3-layer-metal and N-well CMOS technologies. The chip was named YTEG (Yield evaluation Test Element Group). The yield evaluation procedure using the YTEG is shown in Fig. 2. First, all chips are evaluated using an LSI tester, and fault cell maps are obtained for each structure (step 1). Next, defects are accurately located using light emission microscopy (LEM) or liquid crystal analysis (LCA) (step 2). Then, the defective locations are observed by a photo microscope, and the defects clearly caused by particles or scratches are sorted out (step 3). Finally, the destructive inspection is performed for the residual defects, and the cause of each fault is identified (step 4). Through the step 2 to the step 4, a quick failure analysis system is used (Ref. 3). To facilitate easy adoption of this procedure, the chip and the structures were designed as follows.

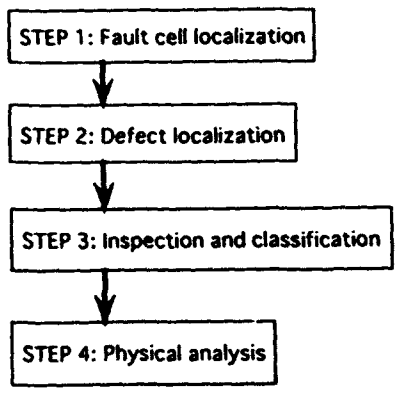


Fig. 2 YTEG evaluation flow.

3.1 Electrical measurement

Both for shortening evaluation time and obtaining statistically valid data simultaneously, the YTEG chip was designed to be applicable to the same LSI tester for LSI evaluation. For this purpose, the YTEG pattern was fabricated on an SOG gate array substrate after contact level, i.e., the chip size and the arrangement of probing pads were the same as those of SOG gate array devices.

Many of the test structures were measured with direct current using two terminals. However, some test structures were divided into small cells (see section 3.3). To overcome the pad number limitation, an array of small cells with X and Y decoders was introduced, and each cell was evaluated by go/no go testing. As shown in Fig. 3, one small cell is selected using X and Y decoders. The cell circuit is rather simple, consisting of a NAND gate and a series of two PMOS transistors and a structure to be evaluated. Since the output voltage Oi varies systematically with the resistance of the structure, the upper and lower limit of Oi for go/no go testing can be calculated from the acceptable value range of the resistance.

3.2 Failure modes and locating defects

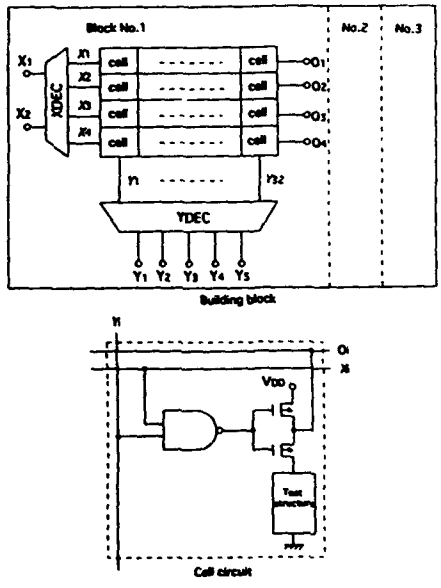


Fig. 3 Test structure cell array with decoders.

Table I Failure modes and analysis methods

Failure mode	Analysis method	Design consideration
Gate oxide leakage	LEM,OBIC	Large cell size, 2 terminal
Junction leakage	LEM,OBIC	(to raise detection probability)
Poor contact	LEM,LCA,EBT	Small cell size with decoder
Poor metalization	LCA,EBT	Series resistance of cell < 10 kΩ
Metalization short	LCA	(to flow 10 mA at 10 V)

LEM: Light emission microscopy
OBIC: Optical beam induced current
LCA: Liquid crystal analysis
EBT: Electron beam testing

As mentioned in section 2.3, different methods for locating defects can be applied independently or in combination depending on the objective failure mode. Gate oxide leakage and junction leakage can be identified by LEM. Failed vias and void generation positions can also be detected by LEM under appropriate joule heating conditions because it is sensitive to IR-rays radiated from materials at temperatures higher than about 250 ℃. Short and resistive positions can also be located by LCA. Whether LEM or LCA is used depends on the degree of resistance increase at these positions. Open positions can be located by observing voltage contrast images from an EB-tester or an SEM.

3.3 Size limit of unit cells

The limit of the unit cell size of a test structure varies with its objective failure mode, since the defect location method differs according to the objective failure mode. For gate oxide and junction failures, large-sized test structures are permissible since defect positions can easily be located by LEM even if multiple defects are generated in a structure. On the other hand, for interconnection failures, although large-sized test structures are desirable from the viewpoint of statistical analysis, the size is limited by series resistance because an adequate amount of current must be applied to heat up resistive portions by joule heating to locate them by LEM or LCA. An array of small cells with X and Y decoders, shown in Fig. 3, was introduced to satisfy the above contradictory demand from statistical and failure analyses. The array configuration is also effective in detecting multiple defects individually in a large-sized test structure.

Table I summarizes the relationship between failure modes and analysis methods with design considerations relevant to cell size included as well.

3.4 Chip layout

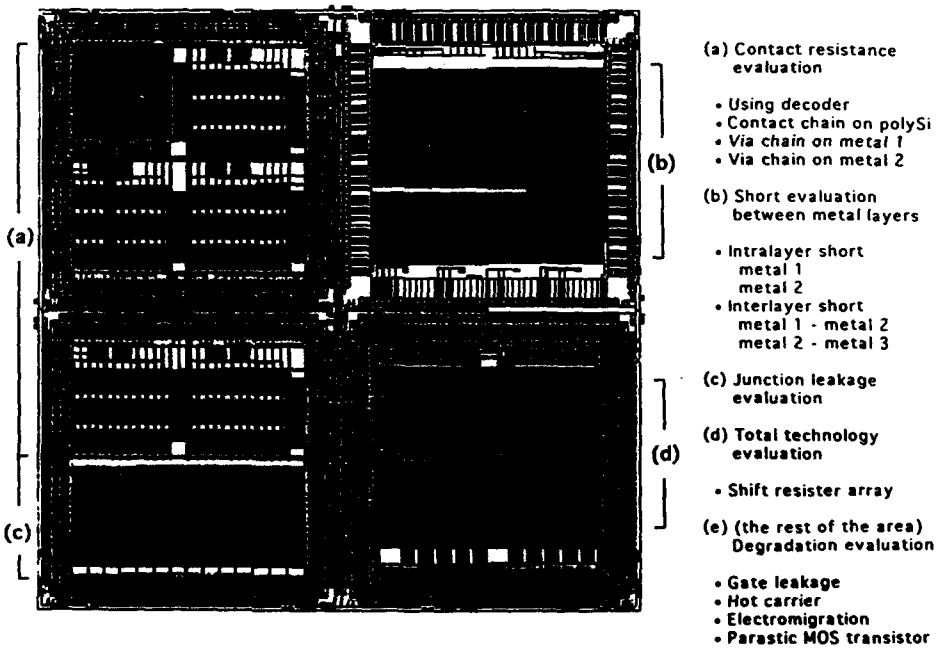


Fig. 4 Layout of YTEG.

Figure 4 shows the layout of the YTEG chip. The chip consists of four subchips, and total chip size is 15 mm x 15 mm. Test structures are categorized into five groups, and each group's objective is described in the figure. Group (a) consists of several serpentine patterns. The underlying topography was varied intentionally from pattern to pattern to obtain different via hole depth. Group (b) consists of several modified comb patterns including the most severe spaces from the viewpoint of design rule and underlying topography. Group (c) consists of several parallel cell arrays in which the positions of metal 1-to-junction contact holes were varied from pattern to pattern to examine the influence of peripheral structures such as field oxide, gate electrode and metal 1. Group (d) is a shift resister array with 2400 stages using D-F/F. The D-F/F is a basic circuit cell in our cell library, which was found by failure analyses of real chips to have a relatively small process margin. Group (e) consists of several special test structures designed to examine various types of degradation. Some of them will be introduced in section 4.

4. EXAMPLES OF EXPERIMENTAL RESULTS

The 0.5-μm CMOS technology was evaluated using the YTEG, and many valuable data were obtained, some of which are presented in this section.

4.1 Data accuracy enhancement

Traditionally test structures have been designed with careful attention to avoid erroneous measurements or erroneous interpretation of data. However, this made it necessary to have a great deal of redundancy when only electrical measurements were employed, which resulted in an increase in the number of probing pads and cell size. Moreover, the possibility of making mistakes remained in spite of these efforts. Using the yield evaluation procedure presented in Fig. 2, this disadvantage was eliminated. Figure 5 shows the ratio of failure per cause evaluated using the group (c) structures. Although there remain some ambiguities because step 4 has not yet been performed, it is clear from the figure that the objective failure cause, alloy spike in this case, is responsible for only 35 percent of the failures at most. If all failures had been taken as metal 1-to-junction leakage, considerable overestimation would have occurred. In this manner, the data accuracy was remarkably enhanced.

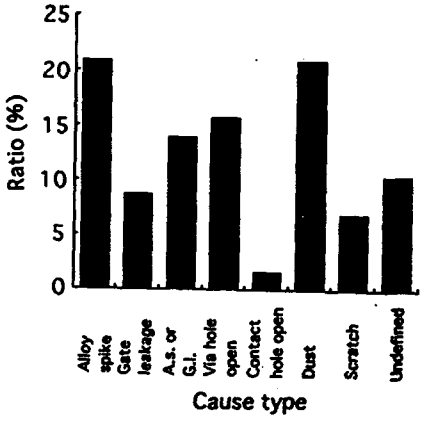


Fig. 5 Ratio of failure causes for junction leakag test structure.

4.2 Assistance for process improvement

Careful examination of the data, especially using the group (e) structures, gave important information for process improvement. Two examples are presented below.

4.2.1 Gate oxide leakage

Electrical measurements were first carried out to find the faulty chips. Then, LEM was performed for the test structures with leakage. The results are shown in Fig. 6 where all emission sites are superimposed on a single basic cell of the gate array. In this structure, the metallization is wired keeping away from the gate oxide area to accurately identify the position of the emission sites. Electrical measurements revealed that many of the leakage paths were in n-type regions, i.e., in the source and drain regions of NMOS transistors and the channel regions of PMOS transistors. On the other hand, LEM results indicate that leakage tends to occur at structurally

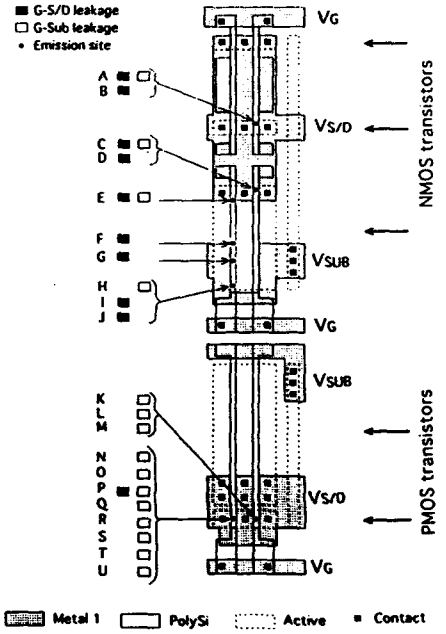


Fig. 6 Structure for gate leakage detection.

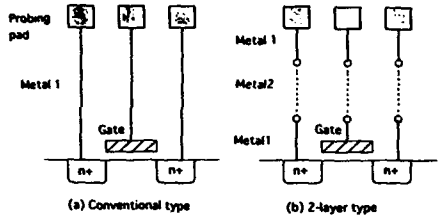


Fig. 7 NMOS transistors for hot carrier immunity evaluation.

the same positions, and these results are the same as those obtained by the failure analysis of real chips. No emission site was observed in gate areas where there is no metal 1 around them. These results indicate the possibility that gate oxide leakage occurred due to the complex effects of charge-up and stress-induced gate oxide degradation.

4.2.2 Hot carrier immunity

The two types of NMOS transistor shown in Fig. 7 were included in the YTEG to examine hot carrier immunity. Figure 8 shows that the hot carrier degradation of the conventional type was larger than that of the 2-layer type. In the figure, the life time is defined as the time required for the threshold voltage shift to reach a certain value (Ref. 4). The generation of damage due to charge-up during a certain process step between metal 1 and first via formation is inferred from this fact.

4.3 Yield prediction

An example of go/no go testing data of the group (a) structures is shown in Fig. 9. In the figure, the data were arranged in order of physical position in a chip. Examining all of these map data strongly suggested that

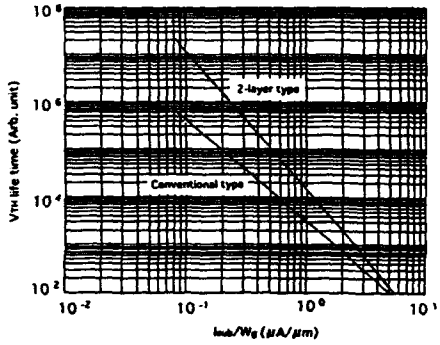


Fig. 8 Hot carrier immunity dependence on transistor type.

no cluster existed at interconnection levels. Furthermore, photo microscope inspection confirmed that all these defects were caused by particles or scratches. These facts indicated that Poisson statistics could be used for yield prediction. Actually, the actual and predicted yield were compared on two levels, a shift resister array in the group (d) and a real LSI chip. First, the defect density of each yield-loss factor was calculated. Then, a device yield due to each yield-loss factor was calculated considering area ratio between the device and each factor. Finally, the product of the device yields for all factors was calculated, and that is the predicted yield (Ref. 5). As a result, wafer-to-wafer agreement was obtained between the actual and predicted function yield of the shift resister array. Also, good agreement was obtained for the average function yield of the real LSI chip. It can be concluded that the function yield of the devices fabricated in our laboratory is dominated by random defects caused by particles.

```
    **********                                                    94/02/28  19:23:39
    ***  AAA TEST  RESULT     PRO : RSUB1DC2  VER : 1.0    ***     WAFER = N03-01
    ************************************************************     CHIP =     18
*** I/O CONTACT CHECK LIST ***   TEST 1000 -
    INPUT  PIN      = OK =    OPEN =   0   SHORT =   0   AVE = -686.0HV   MAX = -688.0HV   MIN = -684.0HV
*** VDD CONTACT CHECK LIST ***   TEST 2000 -          VDD  = OK =         VALUE = -118.0HV
*** DECODER TH CHAIN CHECK LIST ***     TEST 4000 -
    TEST RATE = 10.00US VDD= 3.300V  VIH= 3.300V  VIL= 0.000V  PASS PSI=2.4V,2.7V TH=2.4V,2.7V
    RPSC31A  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC31B  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC31C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC31D  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC32A  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 NG1 0 0 0 0 0
    RPSC32B  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC32C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC32D  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC33A  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC33B  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC33C  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    RPSC33D  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
                                                                  POLY SI CHAIN3    = FAIL FAIL
```

Fig. 9 Example of go/no go testing data. '0' means pass cell, 'NG1' fail cell.

5. CONCLUSION

A test chip that meets requirements for failure analysis as well as parametric and statistical analyses has been developed to shorten the debugging cycle time in fabrication. For this purpose, the following design strategy was adopted. To utilize a restricted chip area, the most effective test structures were designed based on failure analyses of real chips. To estimate accurate defect density, the test structures were designed to be suitable for both electrical measurement and failure analysis. To locate defects easily, the design and size of each test structure was decided respectively according to the defect-locating method chosen. The effectiveness of the test chip for yield prediction and process improvement has been confirmed through application to evaluate 0.5-μm CMOS process technologies.

ACKNOWLEDGMENT

The authors would like to thank Dr. Sunao Ishihara, Dr. Atsushi Iwata, Dr. Eisuke Arai and Kenji Miura for their helpful advice and discussion, and Nobuhiro Shimoyama for evaluating hot carrier immunity. They also wish to extend their thanks to the staff of NTT Electronics Technology for their invaluable assistance.

REFERENCES

1. Lukaszek W., Yarbrough W., Walker T. and Meindl J. "*CMOS Test Chip Design* for Process Problem Debugging and Yield Prediction Experiments", SST-29, No. 3, March 1986, pp.87-93.

2. Beckers S. and Hiltrop C. "The Spidermask: A New Approach for Yield Monitoring Using Product Adaptable Test Structures", Proc. ICMTS, 1990, pp.61-66.

3. Nakajima S., Ueki T., Tazawa S., Ochiai K., Nakamura S., Komine Y. and Takeda T. "A Quick Failure Analysis System for IDD Leakage", Proc. ISTFA, 1993, pp.249-255.

4. Shimoyama N., Machida K., Takahashi J., Minegishi K. and Tsuchiya T. "Enhanced Hot-Carrier Degradation Due to Water-Related Components in TEOS/O_3 Oxide and Water Blocking with ECR-SiO_2 Film", IEEE ED-40, No. 9, September 1993, pp.1682-1687.

5. Stapper C. H., Armstrong F. M. and Saji K., "Integrated Circuit Yield Statistics", Proc. IEEE-71, No. 4, April 1983, pp.453-470.

ASIC DESIGN HOUSE APPROVAL - THE EVOLUTION OF A STANDARD

Clive N. Robinson - Product Reliability and Quality Ltd.
Tony Birks - Defence Research Agency

ABSTRACT

The nature of modern, custom designed, Integrated Circuits, and of the industry which produces them, added to the expectations of the customers, has led to problems in the application of established methods of component approval. Research identified the needs of the industry, and the concepts of ASIC Design House Approval followed by Process Approval and Technology Approval were devised.

Key features of these new types of approval are the use of the company's own quality system allowing rapid approval or extension of approval. The application of Technology Approval in manufacturing will reduce the cost of components, giving greater confidence in their quality and reliability, together with an earlier release.

The schemes are readily applicable to the activities and operational structures of the Small to Medium Enterprises (SMEs) which make up much of the ASIC industry today.

1. WHY IS A STANDARD NECESSARY?

Over the last decade, the increasing use of Application Specific Integrated Circuits (ASICs) and their increasing complexity, has led to concerns over their impact on the reliability of the equipment into which they are designed. By their nature, ASICs are produced in small batches and this leads to major difficulties in both reliability assurance and reliability prediction of the component. Normal techniques such as life testing and burn-in are seldom of value and it is difficult, if not impossible, to compile the life test data upon which failure rates of standard components are based. In any case, these techniques were primarily devised to address random failure mechanisms and those due to manufacturing and processing faults.

A bigger problem with custom components is the possibility of inherent design faults. In an ideal world, these would be uncovered through thorough testing but, given the complexity of modern ASICs, it is seldom economically viable (if indeed possible) to test a device exhaustively.

Several methods devised to enhance the reliability and fitness of purpose through control of the design process itself have been introduced. In certain applications quantitative design techniques such as Formal Methods can be employed to give a device with a totally predictable performance. A similar approach is the use of a so called High Definition Language (of which VHDL is one) as the design tool. Both methods endeavour to ensure the integrity of the device output by adopting a algorithmic design technique combined with a mathematical definition of the device inputs. Both have been used successfully and the use of High Definition Languages in particular are becoming widespread amongst the ASIC design community. The major drawback is that they are applicable only to digital designs and are not appropriate for the analogue or mixed signal devices which are becoming more prevalent. In addition, the integrity of the design will still depend upon the designer, or design team, for the proper use of such tools. Also, the fitness of purpose of the device will depend upon its being correctly specified. The specification is one of the most prominent sources of error in an ASIC design.

In the latter half of the 1980s, in a project initiated by their joint Ministries of Defence, this problem was investigated by a team of ASIC designers and users from the UK, France and Germany. They suggested an alternative approach; that the activities of the design facilities and of the design process itself be more closely controlled (Ref. 1). In the UK, the Defense Research Agency (then the MoD Procurement Executive) were already working on a standard for ASIC design facilities. This was published in 1989 in BS9000 Part 4 and was the first standard in the world to address this activity. The authors were involved in the development of this standard and in steering the first ASIC design house (namely, Walmsley Microsystems Ltd (WML) of Birmingham) through the process to become the first company to gain approval.

The experience gained has contributed to the evolution of Process Approval and, more recently, Technology Approval which have been adopted by the CECC (Refs. 2 & 5). These schemes are now being taken up by other sectors of the electronics industry such as ceramic capacitors and printed wiring boards.

The following sections explain the thinking behind the BS9000 Part 4 standard for ASIC Design Facilities and the lessons learned in its employment. The arguments are then expanded, leading to the concepts of Process Approval and Technology Approval and their wider application. The intention is not to spell out the requirements of the individual standards (for this, see Refs. 2, 3 & 5), rather to give an overview of the basic principles behind them.

2. SETTING THE STANDARD

The fundamental philosophy behind these standards is one of approving the design *process* rather than qualifying a product. The first step must therefore be to examine the ASIC design process, identify the critical areas and then to establish a system of control.

Two important points must be considered. Firstly, custom IC design is typically carried out by a small group of people. This may be a design group within a larger company (perhaps an equipment manufacturer or a semiconductor processing house) or an independent design company. Secondly, 'ASIC Design' often means much more than merely designing a device.

A customer ordering an ASIC will often require the design facility to carry out the design and project manage the component through production, assembly and test, delivering prototype and then production devices. Small design companies usually work within the latter scenario but will, on occasion, carry out a pure design exercise. Both modes of operation must be allowed for if the standard is to be of use. To address this, BS9000 defines two classes of design activity namely, 'Class A' and 'Class B.' Figure 1 illustrates this extended design activity. A Class A design facility is approved to carry out design only, whilst a Class B facility will also carry out or subcontract the other functions, acting as the prime contractor and the interface to the customer. A Class B facility will thus be involved in a more complicated operation. There are certain commercial advantages in being independent from any one manufacturing process and many small companies will operate to this model.

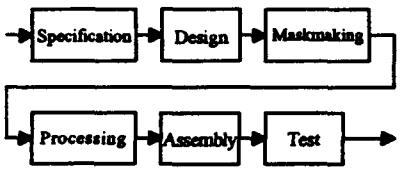


Figure 1. The Extended ASIC Design Activity

2.1 The Design Process

The design activity itself can be abstracted down to the iterative process shown in figure 2. This model is typical of many design situations independent of the circuit type and the technology used and it is here that control must be applied in order to improve design performance.

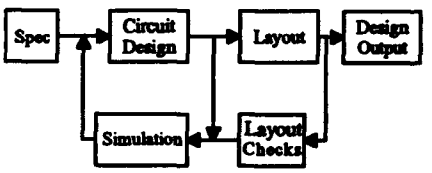


Figure 2. The Detailed ASIC Design Activity

The specification is usually presented by the customer and is translated into a circuit design (i.e. a schematic level) by the designer. This will then be simulated, using a variety of software design tools, and compared with the performance requirements. When satisfied, the designer will proceed to convert the circuit into a chip layout. He will do this for a specific manufacturing process which will have been carefully chosen and he will use electrical and topological (layout) design rules supplied by the processing house. The performance of the circuit in this form will be technology dependent and further simulation will be carried out along with checks to ensure that the layout rules have been obeyed and that the layout reflects the schematic (layout vs. Schematic or LVS checks). This, iterative, stage ends with some final checks and the last step is a translation of the layout data into a standard format for mask making. It is important to note that different design facilities may use different equipment and software and the same facility may use different tools for different processes or design types. The detail of their design procedures may also change but this general design model covers most situations.

Having thus abstracted the design task, we must identify the weaknesses therein. Investigation revealed that the common problems in the design process arise from the following:

1. Poor specification
2. Misinterpretation of the specification
3. Misuse of design tools
4. Use of inappropriate tools
5. Poor control of the design activity (i.e. tasks and checks not carried out)
6. The use of an inappropriate technology.

To address the above, three basic principles can be employed.

Firstly, greater customer involvement in the process. This can be addressed by establishing a series of formal design reviews, initially to ensure that the specification is understood by the designer (and by the customer) and later, to ensure that the design is converging to the specification. Formal design reviews put a certain amount of responsibility upon the customer and this is essential. Often, the customer will not understand the full significance of various details of his specification until well into the design process. Two way communication is required; guidance given to the customer by the design facility and clarification of the requirements by the customer. Commonly, a customer will be well versed in board-level design and works in the situation where design modifications can be carried out on the prototype at a late stage. The point at which commitment is required is blurred and the specification often reflects this. With ASIC design, the option of physical modification is not possible and, once the device has been produced, a redesign is a very costly option. It is of prime importance to ensure that the design meets the specification and the specification meets the customer requirements. In both cases, ultimate responsibility must rest with the design facility being the expert in the field.

Secondly, making sure the designer has the correct design tools and information for the job. The design tools include the software (for simulation, layout and checking the design) and the hardware platforms upon which it runs. Software control methods are essential to ensure the provision and maintenance of appropriate

tools. The other information required will include process design rules which are used in putting together the actual layout and mathematical device models and process characteristics used for simulation. Both of these are process dependent and the information will be obtained from the appropriate processing house. Again, it is up to the design facility to ensure that this information is correct, current and sufficient for the job in hand. Methods of software and information management must be applied.

Thirdly, one must ensure that the designer does his job properly. This introduces the questions of staff selection, qualification and training. But, given an adequate level of competence, measures must be taken to ensure that the designer takes the correct steps, makes the appropriate decisions and addresses all the issues at each stage of the design. One way to achieve this is to establish control gates in the design flow. these can take the form of check points after each functional activity and can take the form of formalised design reviews, internal checks or triggers for information update.

Additionally, some basic quality management practices such as documentation and drawing control and internal auditing are essential to the control of the basic design process.

It can be seen, therefore, that two basic types of control are necessary; the project specific controls which must be addressed at the appropriate points in the design, and the on-going, support activities. The former will be specific to each design, technology and customer and the latter are more general, company practices

2.2 The Extended Design Activity

Some design groups within larger groups, and most independent design houses, will carry out the extended activities of a Class B facility as defined above and illustrated in Figure 1. Thus, they will carry out the design (and perhaps test) in-house and subcontract the other steps to one or more vendors. They will adopt a project management role for the device procurement and act as the interface to the customer.

The design stage will generally fall in line with that discussed above and control can be established in a similar fashion. Control of the other activities leads to additional requirements mainly associated with subcontractor control. On top of this, some of the procedures already mentioned, for example document control and auditing, must be extended. The Class B activity will also involve the receipt, handling, storage and despatch of material. Although most of these can be addressed by the adoption of standard industry practices, it must be recognised that effective control of the extended activity will require a higher level of management commitment and resource as the facility is now responsible, to the customer, for the performance of the entire supply chain and must be both prepared and armed to deal with issues such as customer returns and corrective actions.

These, then are the issues which must be addressed in order to establish control over the ASIC design task. To be acceptable, the standard must have some additional characteristics.

2.3 Further Requirements

If the standard is to be first accepted and then adopted by industry (both customers and suppliers), the following criteria should be met:

1. It should be workable. The standard must be readily achievable by a design facility. Whilst being effective, it should not prove overly onerous, neither should it be expensive for a small company to implement. It should also be acceptable to the design staff. A lot of effort was expended to ensure that the requirements made sense to the designers.

2. It must be flexible. The standard should be applicable and effective in a wide variety of business scenarios. In this case it must be adaptable to the different organisational structures which have evolved to tackle ASIC design, be they large or small, independent or not. Some facilities may have QA systems already in place, some may be approved to other standards (e.g. ISO 9000) and others will be subject to the QA systems of a parent organisation. The standard should readily integrate into these environments.

3. It should address the needs of the industry, supplier and customer alike. This boils down to the question, 'what will the customer gain from placing his order with an approved supplier and what will the supplier gain from approval?'

4. It must be amenable to third party assessment. For a specialised activity such as ASIC design, a high degree of expertise would be required to assess the operation of a company with no formal quality system. Approval to a standard should be carried out in a manner which will facilitate the job of an external assessor without detailed technical knowledge of ASIC design.

5. Rapid approval and maintenance. In the ASIC industry the technologies employed, including those of the design tools, change continually and swiftly. If a standard is to be practical, it is essential that a route is available to quickly extend or change the scope of the approval without major upset to the company concerned.

3. MEETING THE STANDARD

The requirements of BS9000 Part 4 were written to fulfil the above criteria. In January 1992, Walmsley Microsystems Ltd. became the first IC design house to be approved under the scheme (Ref. 4). The company

is a small, independent design house. It operates as a Class B design facility, carrying out the design tasks in-house and, as the prime contractor, subcontracts the mask making and processing etc. A range of different technologies are used and digital and analogue devices of varying complexity are produced. As a case study for the merits and viability of a new standard, the company therefore provides one of the more complex and testing scenarios.

Prior to its BS9000 approval, WML was not approved to any other quality standard although it had established basic QA procedures such as drawing and document control systems. It took the company approximately one year to prepare for approval, putting in place the necessary procedures and documentation. As expected, design control became the focus of these efforts, most of the other requirements being met by adopting standard, industry procedures. The design staff were heavily involved in this development and the design control was eventually refined to a series of checks and reviews which can be reduced to just two, one page control sheets and were later embedded directly into the design software.

WML has been operating the system for over two years and it has seen benefits in the smooth running of their operation and in customer perception of their service.

4. PROCESS APPROVAL AND TECHNOLOGY APPROVAL

Whilst BS9000 Part 4 was aimed at the ASIC design function, the arguments put forward in section 1 for a new type of standard (increasing complexity and customisation with faster turnaround), also apply to the other steps in ASIC procurement.

The structure of the current ASIC industry is modular in form, a different organisation carrying out each step of the process. In order to provide an approved component, these other processes must also be approved. On top of this, as the complexity of devices increased, the cost and time involved in qualifying and maintaining approvals to the existing CECC rules for either qualification or capability approval was proving onerous. A new form of approval was clearly required. The schemes of Process approval and Technology Approval have been developed to meet this need.

4.1 Process Approval

Process Approval addresses the above by careful structuring of the quality system and organisational requirements. For each individual process, a Process Assessment Schedule (PAS) has been written detailing the requirements for that activity. To gain approval, a company is required to prepare a Process Manual to show how their Quality System meets the requirements. This document acts as a route map through their system with reference to the requirements and will be a guide for external auditing. Process control is implemented by identifying the Quality Factors which have a major impact on the quality of the process and to identify the associated Critical Parameters to provide a measurable indication, where possible.

Process Approval is now well established and Process Assessment Schedules have been written for many activities including ASIC design, Mask Manufacture, Ceramic Packaging and Hybrid manufacture. Outside the ASIC industry, PASs have been prepared for Ceramic Capacitors and Printed Circuit Board design, amongst others. So far, some thirteen PASs have been written, allowing many activities to be approved under the scheme. If its activities are not already covered, any company, trade body or other relevant organisation can produce requirements in the form of a Process Assessment Schedule and submit it to the CECC for inclusion in the scheme.

4.2 Technology Approval

As the work described in the previous sections was underway, an extensive programme was set up to investigate the actions component manufacturers carried out to satisfy themselves that the components manufactured were suitable for sale and also the expectations of the users .

The overwhelming response was that Statistical Process Control (SPC) was being used in all areas of manufacture, with all the necessary requirements of feedback and improvement being implemented. Also, the philosophy of Total Quality Management (TQM) was being discussed and encouraged. Although there was considerable interest in these matters in most firms, there were no national or international standards invoking such tools that could form the basis of a third party audit or approval.

It was therefore agreed that the cost effective way forward for a new approval should include the SPC and TQM philosophies.

This led to the development of Technology Approval.. TA retains the structure and modular format of Process Approval but broadens the scope to include the use of SPC and TQM.

The requirements for Technology Approval are written in a Technology Approval Schedule (TAS), as opposed to a Process Approval Schedule, and the guide-lines for this are written in CECC Rules of Procedure RP14 Part VI. There are five mandatory requirements:

1. The use of process control methods (defined in the TAS) and tools to demonstrate the control of processes and product.

i.e. the component manufacturer must be successfully using SPC in all relevant areas of manufacture and test.

2. Continuous quality improvement and its demonstration.

i.e. Using SPC, for example, the company can demonstrate that it has a continuous improvement programme.

3. Monitoring the overall technologies and operation associated with the design and manufacturing processes as well as the components themselves.

i.e. End of line testing is only part of confirming the acceptability of the component. Satisfactory control of the design, manufacture and in-process testing should, almost by default, guarantee that the component will meet its final tests.

4. Process flexibility based on a company's own quality assurance management system and component sector requirements.

i.e. Each manufacturer may have different ways to produce his components and test them. Therefore, each manufacturer defines the actions he takes to satisfy himself that his component is acceptable, and meets his customer's requirements.

5. Rapid approval or extension of approval by use of a Technology Approval Declaration Document (TADD).

i.e. As part of the TADD, the component manufacturer has to describe fully his processes. Therefore, if he wishes to extend his approval for a process, the process extension must be fully documented. The manufacturer then has to demonstrate that he can meet the requirements of the extension and that the process is fully under control. When this has been done, to the satisfaction of the approval authority, the extension is approved.

The TADD can be considered the equivalent of a Process Assessment Schedule in Process approval.

4.3 The Technology Approval Declaration Document

The TADD is similar to a Capability Manual but extends to cover both the technical and managerial aspects of the operation. The following headings (taken from RP14 Part VI) give an indication of the required contents of this document:

1. Manage commitment to Quality
2. Quality Improvement Programme
3. Definition of the relevant Sites and Operations
4. Interface with Subcontractors
5. Internal Audit Plan
6. Test Vehicles
7. Description of Technology
   (This is discussed in section 4.4.)
8. Management of non-conforming Product or Activity
9. New Product Introduction Programme

The manufacturer must document both his new product introduction programme and procedures for its characterisation if it is to be manufactured under the scope of the TAS.

4.4 Description of Technology

The requirements of the Description of Technology are the most searching, because each manufacturer approaches the manufacture and quality of his components, within a component technology, from different directions. The Rules of Procedure provide guide-lines (in the form of TASs) on how to achieve this, for various technologies. Some of these are shown in figure 3(b).

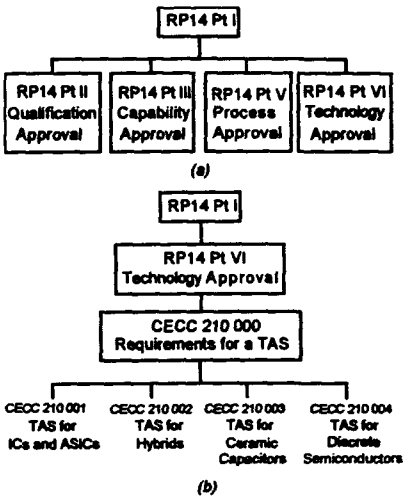


Figure 3. CECC Document Structure

4.5 CECC Implementation of Process and Technology Approvals

Both Process and Technology Approval are incorporated into the CECC system under Rules of Procedure RP14. The family tree of this system is shown in figure 3(a). RP14 Part V deals with Process Approval whilst Technology Approval is addressed in RP14 Part VI.

It will be recognised that many of the general quality requirements for the BS9000 Design House approval are covered by the ISO 9000 standard. This is also true of the other schemes and RP14 incorporates the ISO 9000 requirements in RP14 Part I. A manufacturer must be approved to this before he can gain Technology Approval. Once he has done this, he must then write a Process Assessment Schedule, in order to acquire Process Approval, or a Technology Approval Declaration Document for Technology Approval.

4.6 Benefits to the Customer

What confidence does the user of the component gain when purchasing a component from a manufacturer

with Technology Approval? Some of the advantages are:

1. A knowledge that, as a *minimum*, the supplier has approval to RP14 Part I which encompasses the requirements of ISO 9000

2. A low failure rate through the correct application of Process Control.

3. New qualified product can be released as quickly as commercial product. In fact, all components produced under TA are automatically qualified.

4. The manufacturer will have a Quality Improvement Programme in Place

5. Independent auditing by a third party

6. CECC Detail Specs for all "standard" components, Customer Detail Specs for customer specific components (containing all the customer requirements), or registered data sheets. All components manufactured within a company's Technology Approval Boundary may have CECC release. - Note the significance of this last point.

4.7 The General Applicability of Technology Approval

Although Technology Approval has been driven by the IC and ASIC users and producers, other component areas have realised the cost saving and quality improvements to be gained through the application of TA. Component sectors developing TA or investigating its application are:

- Ceramic Capacitors
- Discrete Devices
- Piezoelectric Devices
- Servos and Synchros
- Opto-couplers

5. CONCLUSIONS AND FURTHER WORK

RP 14 Part VI and CECC 210 000 have been published by the CECC in 1994. A number of companies are, at their own risk, already looking at Technology Approval and have started to apply it before the TAS documentation have been finally published. The benefits of this type of approval have already been shown, to a certain extent, by the application of the principles within the BS9000 ASIC Design House Approval and these have been discused in setion 3. The Design House in question has applied for Process Approval and intends to undergo Technology Approval as soon as possible. This indicates the attraction of the system to the SMEs which make up much of today's industry. Indeed the scheme, with its modular structure, provides a structure within which small businesses can supply approved product in a way not previously possible.

The use of such an approval has ramifications in related areas such as the recognised problems of future supply of devices, but these are outside the scope of this paper.

The Ministries of Defence for Germany, France, Italy and the UK are currently financing an experiment where ASIC designers, manufacturers and users are investigating the application of TA by means of actually producing components under the scheme and comparing this with other forms of release. All the companies involved in the experiment should gain Technology Approval, with obvious advantages to other ASIC users and suppliers.

It is expected that the inherent flexibility of Technology Approval will lead to its widespread use in the European Component Industry., particularly amongst SMEs.

REFERENCES

1. TIEGEC Working Group 2, "Generic Qualification For ASICs and ICs", Final Report, Joint Ministries of Defense, Grenoble Meeting, April 1990

2. CECC 200 000, Requirements for Process Assessment Schedules for Process Approval

3. CECC 200 002, Process Assessment Schedule: Integrated Circuits Design Facility, Final Draft March 1992

4. Robinson, C.N. "Design House Approval - The Walmsley Experience", Presented to CODERM Group, Jan 1994

5. CECC 210 000, Requirements For Technology Approval Schedules (TAS)

Further information can be obtained from the authors
c/o Product Reliability and Quality Ltd.
23 Freelands Road, Ratho,
Edinburgh
EH28 8NP Tel. & Fax 031-333 1162

PLANNING EARLY LIFE RELIABILITY TESTING USING THE 'HUGHES' MODEL

By : Martin Shaw
Product Quality and Reliability Lab
IBM U.K Ltd
Greenock,
Scotland

1.1.1 ABSTRACT

An application of the 'Hughes Test Strength Model' is described in assessing effect of environmental stress testing for Early Life evaluation of video display units. Comparison of different test methods plus results used to validate Hughes model.

1.1.2 BACKGROUND

Over the last two years Greenock Reliability Engineers have worked closely with suppliers to help them introduce well defined Quality / Reliability test strategies. The objective has been to minimise IBM test costs and place the test responsibility where it belongs, at the feet of the vendor.

Not only does this make the vendor more accountable for Quality problems, it provides the right type of motivation to the vendor who is more in control of Quality output than ever before.

Prior to implementing this strategy IBM carried the burden of test costs and also had to co-ordinate sub-system changes to improve Quality / Reliability levels, using up a great deal of 'expensive' man hrs in doing so. This has now changed greatly with the emphasis placed on the vendor and the Greenock engineers doing more of a 'policeman' type role, receiving test information from the vendor rather than creating it.

The remainder of the paper describes in detail a case study of the successful implementation of reliability testing on a recent level Personal System video display and how the test data along with that from controlled testing at IBM sites has enabled a very worthwhile study of the Hughes test strength model.

Generally when an engineer decides to set up an Early Life reliability test strategy, the first stress types to be considered are random vibration and thermal cycling as they enable short duration testing of large sample sizes. This can yield a great deal of information on the Early Life reliability of the product to be tested. This enables the engineer to plan reliability growth testing within a defined period and be in a position to predict the expected Early Life performance of a product prior to volume shipment.

This then begs the question of how long the testing should last and to what effect it will have, i.e - the percentage of 'latent' early life defects that can be precipitated by the combination of various types of stress. As early life defects are predominantly caused by process weaknesses rather than being intrinsic to the design, short, sharp testing is found to be most effective.

When testing was first set up at the video display vendor their test capability restricted the test strength of their early life test. To get over this problem testing was also performed at the two main IBM test sites in Scotland and North Carolina, U.S.A where the equipment is far more advanced and enables achievement of far higher stress levels.

This provided an excellent opportunity to perform testing at different stress levels on similar vintage products which could now be used to evaluate the use of the Hughes test strength model by comparing the types of defect found and the effect that would have on field reliability.

The data used in this paper clearly demonstrates the Hughes model is of great use in evaluating early life stress test methods, showing vendor test effectiveness of 46% versus 89% for that of the more advanced IBM testing. These figures are proven to be an excellent measure of the early life test success when compared to the defect types found in the field versus those found during in-house testing

The old methods of testing often involved long duration elevated temperature testing, but as this was fairly ineffective in highlighting process weaknesses it was restricting the continual improvement in product Quality and Reliability. As a result, the Greenock Reliability Lab was built during 1984 at a cost of 1.5 million U.S Dollars to handle high volume, high stress, early life testing with the emphasis on detecting latent, early life process type defects.

As this type of testing does not lend itself to Acceleration Factor calculations such as the Arrhenius model, applied to steady temperature testing, the simulation time of the testing cannot be easily evaluated.

The alternative is to use the Hughes equation which will provide a level of test effectiveness, independent of simulation time. By doing this the emphasis changes from setting up a test to measure reliability with respect to time to that of capturing as many of the early life process defects in as short a time as possible, most effectively described by the Hughes equation.

Once sufficient confidence in this practice is achieved it becomes a simple task to plan future tests and using actual test / field data, the optimum early life test profiles can be established.

Test profiles will contain length of random vibration time, number of thermal cycles plus temperature limits, number of power cycles, etc. The paper describes in detail the vendor and IBM test profiles with the exact calculation of test strength. Figure 3. illustrates the thermal stress profiles used both in IBM and also by the vendor.

1.1.3 THE HUGHES EQUATION

The Hughes Equation (Ref 1.) provides information on the possible effectiveness of some commonly used stress test methods. This work was carried out under contract for the Rome Air Development Centre (RADC) and referenced in an RADC Technical Report, issued in 1981.

Any Reliability test should include Vibration, Thermal Cycling, Constant temperature testing and also power on periods along with power cycling to optimise the probability of detecting latent process type defects.

Using test strength equations in Fig 1. and Fig 2. the optimum test can be designed to achieve any set level of test strength. Obviously, the higher the test strength required, the longer will be the test profile required. As in most development and manufacturing processes time is limited therefore it makes sense to set a target test strength level and plan appropriate testing well in advance. Using the Hughes equations allows the reliability engineer to do this with minimal time spent on literature search for the most effective test methods that will reduce test times, often a task that takes longer than actual testing !!

Table 1. details the stress test parameters of the IBM and vendor tests respectively.

Table 1. Stress Test Parameters

Parameter	Vendor	IBM
Thermal Cycling Ramp Rate (Deg C per minute)	1.5	5.0
Number of Thermal Cycles	14	48
T/C Temperature Range • (-20 to + 50) • (-10 to + 40) • (- 5 to + 60)	No. of • 6 • 8	No. • 48
Elevated Temp test duration (Performed at 40 deg C)	60 hrs	12 hrs
Continuous Power on duration	60 hrs	12 hrs
No. Power switches	30	80
Power on time between switches (hrs)	2	.03
20 minute Random Vibration Profile (frequency range)	1 *G* rms	6 *G* rms
Random Vibration Test Strength	0.4	0.7

Figure 1 - Test Strength equations for Thermal Cycling and Constant Temperature

Thermal Cycling ;

$$0.8\left|1-\exp\left(-\frac{N \times dT}{dt} \times 11.835 \times 10^{-5} \times \exp.0122(T+273)\right)\right|$$

N = Number of thermal cycles
dT/dt = Rate of temperature change (deg C per minute)
T = Absolute temperature (deg C)
= (|high temp-25| + |low temp - 25|) / 2

Constant Temperature ;

$$0.6\left|1-\exp\left(-t \times 2.63 \times 10^{-5} \times \exp.0122(T+273)\right)\right|$$

T = Constant Temperature (deg C)
t = Time in test

Figure 2 - Test Strength equations for Continuous Power on and Power Switching

Continuous Power ;

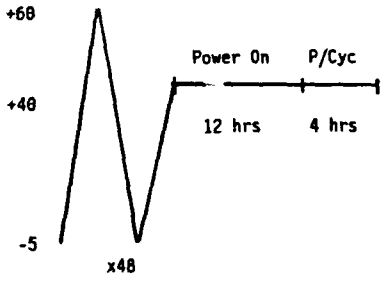
$$\frac{t}{(t+c)} \qquad c = 900$$

Power Switching ;

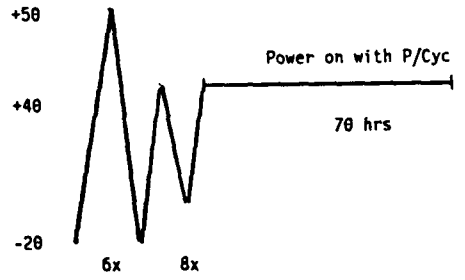
$$1-\left(c/(c+t)\right)^{N}$$

N = Number of power on switches
t = Power on time between switches

Figure 3 - Thermal Cycling test profiles



IBM



Vendor

1.1.4 RANDOM VIBRATION PROFILES

Tables 2 and 3 detail the differing profiles for the IBM and the Vendor vibration tests. Both tests were non-operational with the vendor test constrained by the capability of their random vibration table and controller, hence the frequency range is only up to 200 Hz rather than 2000 Hz for the IBM test.

Table 2. Vendor Random Vibration Test Levels	
Frequency (Hz)	**G2/Hz (PSD Level)**
2.0	0.0010
4.0	0.0300
8.0	0.0300
40.0	0.0300
55.0	0.0100
70.0	0.0100
200	0.0010

Table 3. IBM Random Vibration Test Levels	
Frequency (Hz)	**G2/Hz (PSD Level)**
12.0	0.0030
42.0	0.0030
54.0	0.0100
72.0	0.0100
198.0	0.0010
204.0	0.0200
1998	0.0200

The input energy level for the vendor test is 1 Grms and for the IBM test is 6 G rms. Using this measure and the charts in Ref 2. the test strength is obtained and input in Table 1 and Table 4.

Using the test parameters in Table 1. and applying the Hughes Test Strength equations in Fig 1. it is a simple task to calculate the test strength of the applied Early Life Reliability test profile.

It should be noted that Test Strength for Random Vibration is not derived from any Hughes Equation but is taken directly from an RADC report (Ref 2.). The fixed Test strength values for the random vibration stress is also listed in Table 1.

1.1.5 CALCULATING OVERALL TEST STRENGTH

Using Table 1. plus Figures 1. and 2. the Test Strengths in Table 4. are calculated for each individual stress type.

Table 4. Test Strength Values		
Parameter	**Vendor**	**IBM**
Thermal Cycling	Value	Value
• (-20 to +50)	• 0.048	
• (-10 to +40)	• 0.056	
• (- 5 to +60)		• 0.64
Elevated Temp test	0.042	0.009
Continuous Power On	0.063	0.013
Power Switching	0.058	0.003
Random Vibration	0.40	0.70

As Test strength represents the probability of precipitating a latent defect, the simplest way to calculate the overall test strength is to combine the probabilities of not finding defects with each stress type .

Probability of not finding a defect with any of the stresses imposed

= (1-T.S for T/C) (1-T.S for Vib), etc

= (1-0.048)(1-0.056) ..... (1-0.4)

= 0.46

Prob of finding a defect is therefore ;

= 1-0.46 = 0.54

IBM Test Strength = 0.89

1.1.6 ANALYSIS OF EARLY LIFE TEST RESULTS

Test data was collected over a period of 9 months and field data was used from the first 12 months of installations. To ensure accuracy in the findings the vendor monitor repair centre analysed all U.K field failures and sent information directly back to IBM. They also analysed the failures from IBM Greenock Reliability testing which left in no doubt the conclusions on similarity of failure types found by testing and those experienced by the customer during normal field usage.

Vendor testing was also performed at the manufacturing source with different stress levels in comparison to the Greenock test. All defects were also analysed to root cause providing a database of defects to be compared with ones experienced in the Greenock test and those which were found in the field.

In total, some 44 confirmed monitor failures were returned from customers during the first 12 months of field usage. Of this number, 29 were considered to be of an 'early life' nature caused by inherent weakness in the components or manufacturing processes. Such defects occurred within the first 60 days of customer use. The remainder of the defects (15 of) were either Intrinsic or Dead on Arrival, and not used in the analysis of the early life reliability defect distribution.

Table 5. shows a simple breakdown of failure types experienced during both IBM and also the monitor vendor early life testing.

Table 5. Early Life Failure Types

Defect Type (no. of)	Seen in IBM test	Seen in Vendor test
IC601 failure (6 of)	y	y
F601 failure (6 of)	y	y
D600/D605 fails(5 of)	y	n
Component Solder Land Cracks		
• R315 (4 of)	• y	• n
• CN603 (2 of)	• y	• y
• C316 (2 of)	• y	• n
• C208 (1 of)	• y	• y
Pot Movement (2 of)	n	n
Scan Coil failure (1 of)	n	n

1.1.7 CONCLUSIONS

1. Analysing the data in Table 5. leads one to the obvious conclusion that the IBM Early Life stress test methods were more successful in finding the same type of defects as experienced in the field.
   - During IBM tests, 7 of the 9 defect types were found which would equate to some 26 of the 29 defects, or 90%
   - During Vendor tests, 4 of the 9 defect types were found equating to 15 of the 29 defects or 52%
2. Calculated test strength for vendor test compares closely to the test strength calculated from Table 4. , 54% versus 52%
3. Calculated test strength for IBM test also compares closely to that calculated from Table 2. , 89% versus 90%
4. The Hughes model is by no means perfect but has been shown to be extremely useful in measuring the effectiveness of an early life reliability test proposal.
5. The Hughes model makes no reference to acceleration factor, which for early life defects is very appropriate as the failures would not be intrinsic and thereby difficult to model.
6. Shipped product with any latent early life defects can be considered as 'walking wounded' , requiring the right form of stress test to 'push' the defect mechanism to the failure point when the stresses imposed are greater than the strength of the latent defect.

1.1.8 References

1. Schafer, R.E. ; Gray, S.P. ; James, L.E. ; McMillan, E.A. , Electronic Equipment Screening and Debugging Techniques, Hughes Aircraft Company, Final Report, RADC-TR-78-55, March 1978.

2. Karam, D. , Burn-in : Which Environmental Stress Screens Should be used ? , RADC-TR-81-87, May 1981.

LIFE TIME DETERMINATION BASED ON DEGRADATION KINETICS

Dr. Luc Tielemans
DESTIN N.V.

Dr Luc De Schepper
L.U.C.

Wetenschapspark 1
B-3590 Diepenbeek

Abstract

New strategy based on degradation kinetics is presented . It can be used to determine and control reliability in production and to optimise for reliability in the development.

1. INTRODUCTION

Reliability testing becomes a difficult task due to the ever improving quality and reliability of the electrical components. In the sixties, FIT rates of 1000 FIT were common. Now in the nineties rates reached 10 FIT. In order to verify a failure rate of 10 FIT for a component that operates at 55 °C, a test time of 1000 hours at 125°C on a batch of 4000 samples is required (supposing an activation energy of 0.5 eV and a 60% upper confidence level. The handling of such large test samples over these large test times will become too difficult in the future.

In this article a method based on degradation kinetics is presented and some examples are given to demonstrate the feasibility of this method. This method is capable to overcome the problems mentioned above and can offer more information for reliability control, built-in reliability and feed back to design.

In the following paragraphs this method will be explained for thermal activated degradation mechanisms. In the same way this method can be applied for other stressfactors like current I, voltage V, humidity r.h. etc.

2. DEGRADATION KINETICS AND LIFE TIME

To predict the degradation at a different temperature than the measured ones, at least 2 degradation behaviour curves have to be recorded (until failure). Out of these curves the effective activation energy and the acceleration factor for a failure criterion can be derived.
When more than one degradation mechanism is active, the effective activation energy and the acceleration factor changes for different failure criteria. By changing the failure criterion, the "end of life" changes from one degradation process to the next.
Therefore the effective activation energy has to be determined for each failure criterion when more than one degradation mechanisms are active. This is the case for most products.

3. DEGRADATION KINETICS FOR QUALIFICATION

By qualification of a product it has to be guaranteed that a product will function within its tolerances (= failure criterion) for a certain life time t_{life} . In paragraph 2, the acceleration factor was determined in according to the defined tolerances. To demonstrate that the life time of a product will be at least t_{life} at a maximal operating temperature T_{life}, the product has to be tested during a qualification time t_q at a higher stress level T_q. The qualification time can be derived according the following equation

$$t_q = t_{life} \exp\left\{\frac{E_a}{k_B}\left(\frac{1}{T_q} - \frac{1}{T_{life}}\right)\right\} \qquad (1)$$

To perform qualification tests with fixed test times (according to Total Quality Management TQM) makes no sense, while the demanded life time t_{life} is product dependent and the acceleration factor failure criterion dependent.

4. RELIABILITY INDICATORS FOR LIFE TIME

For a sample lot, tested at temperature T_q , the distribution of the degradation behaviour curves $p^{T_q}(t)$ can be defined. Assuming gaussian statistic, the distribution at the end of the qualification test can be described by the mean: $\langle p^{T_q}(t_q)\rangle$ and the standard deviation $\sigma^{T_q}(t_q)$. From these parameters reliability indicators can be derived (similar to the process indicators for SPC (Statistical Process Control)). The ratio of the mean value to the failure criterion Y is the measure for the intrinsic reliability R_i.

$$R_i = \frac{Y}{\langle p^{T_q}(t_q)\rangle} \qquad (2)$$

An intrinsic reliability factor lower than 1 means that the product has a shorter life time than expected. R_i=1 means that the product fulfils just the required life time. Products with the higher R_i -indicator have the highest life time.
Similar to the definition of the process capability in SPC, the reliability capability factor can be defined as in the following equation (gaussian statistics assumed).

$$R_c = \frac{\left(Y - \langle p^{Tq}(t_q)\rangle\right)}{\sigma^{Tq}(t_q)} \qquad (3)$$

A reliability capability factor larger than 3 means that the products can be considered as a 6 σ-process.
Up to now (cumulative failure method) both lots withstand the test (no failure found) and no difference can be made between these lots in respect to reliability. With these reliability indicators the reliability of products can be characterised and gives the possibility to select the most reliable products. These reliability indicators can be used for purchasing and for reliability SPC, but also as advertising tool.

5. DEGRADATION KINETICS AND RELIABILITY CONTROL

If early failures or unreliable products appear, they have to be selected out using a screening procedure. The test times for the screening tests can be derived from of the distribution of the degradation curves as follows (fig. 1):
1) A theoretical degradation behaviour curve for a product which reachs the failure criterion Y at the qualification time t_q. This curve $Y^{Tq}(t)$ can be named as the "defect curve". All products with a degradation curve above the "defect curve" are failing. The "defect curve" can be calculated by dividing the mean degradation curve by the intrinsic reliability indicator R_i

$$\text{"defect curve"}: Y^{Tq}(t) = \frac{\langle p^{Tq}(t)\rangle}{R_i} = \frac{\langle p^{Tq}(t)\rangle \cdot Y}{\langle p^{Tq}(t_q)\rangle} \qquad (4)$$

2) The screening time t_s can be defined as the minimal time necessary to distinguish is within the distribution of the degradation curves or not. At the end of the screening time t_s , the difference between the defect curve and the upper degradation curve $p^{Tq}_{+3\sigma}(t)$ of the distribution has to be larger than the measuring resolution. (Here the μ+3σ-level is taken).

$$t_s = t \text{ when } Resol = \frac{\langle p^{Tq}(t)\rangle}{R_i} - p^{Tq}_{+3\sigma}(t) \qquad (5)$$

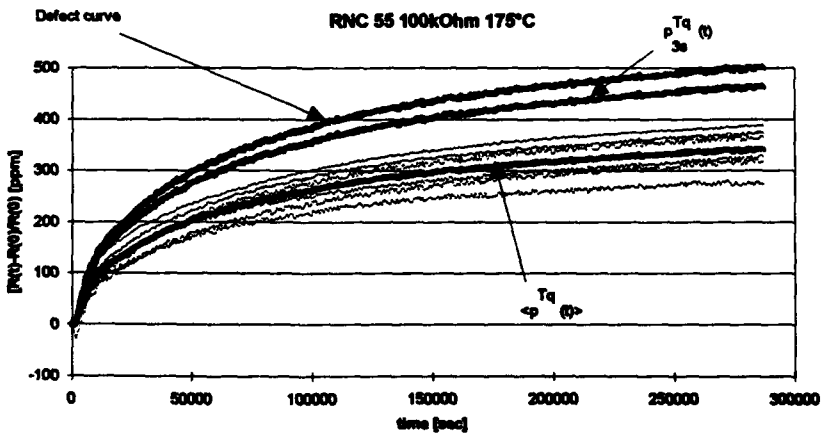
The screening test effort can be reduced by improving the product reliability and by accurate measuring.

To give an example for the validity of the theory explained in the previous sections, the degradation behaviour of RNC 55 resistors is measured. Fig. 1 shows the degradation curves of these resistors. The measuring resolution is in the order of 10 ppm.

From drift measurements at different temperatures, an activation energy of about
1 eV has been derived. Supposing an application with a failure criterion of Y=500 ppm, a required life time of 4 years and a maximal operating temperature of 90 °C, the qualification time t_q is about 300.000 sec (1)

$$t_q = t_{life} \exp\left\{\frac{E_a}{k_B}\left(\frac{1}{T_q} - \frac{1}{T_{life}}\right)\right\}$$

$t_q = 300.000$ sec for this example



The reliability indicators can be derived from the mean and the standard deviation of the drift after 300.000 sec.

$$\langle p^{175C}(300.000\,\text{sec})\rangle = 340ppm$$
$$\sigma^{175C}(300.000\,\text{sec}) = 40,7ppm$$

The reliability indicators can be calculated (2) (3):

$$R_i = \frac{Y}{\langle p^{175C}(300.000\,\text{sec})\rangle} = \frac{500ppm}{340ppm} = 1,47$$

$$R_c = \frac{\left(Y - \langle p^{175C}(300.000\,\text{sec})\rangle\right)}{\sigma^{175C}(300.000\,\text{sec})} = \frac{(500-340)ppm}{40,7ppm} = 3,93$$

These reliability indicators show that the product is intrinsic reliable for the application (life time = 4 years) (R_i>1) (resistors with this life time can be produced) and it can be expected that all resistors of the lot will have the required life time while the capability factoris larger than 3 (R_c=3,9). The resistor lot can be considered almost as a 4 σ-lot.

To control the production (SPC for life time) and to select products which are out of the distribution (early failures and freaks) a screening time has to be derived. As first step the mean degradation behaviour curve $\langle p^{175C}(t)\rangle$ has to be calculated. Further the "defect curve" (4) and the "+3σ-curve" have to be calculated by multiplying the mean-curve with the intrinsic reliability factor R_i for the "defect-curve" and with

$$\frac{\langle p^{175C}(300.000\,\text{sec})\rangle + 3\sigma^{175C}(300.000\,\text{sec})}{\langle p^{175C}(300.000\,\text{sec})\rangle} = \frac{340+122}{340} = 1,36$$

for the "+3σ-curve"

The minimum test time to detect freaks and early failure is the time necessary (screening time t_s) that the defect curve drifts more than the measuring resolution out of the distribution. $t=t_s$ when "defect curve"-"+3s-curve" = measurement resolution (5).

For the in-situ system the measuring resolution was estimated in the order of 10ppm. Out of fig. 2 the screening time t_s can be derived to be 10.000 sec. This means that with the technique based on degradation kinetics the reliability tests times can be reduced (for this application) by a factor of 30.

Using the reliability indicators determined at screening time t_s an SPC-system can be set up to control the reliability of the products.

Transferring this result to the common test times (supposing these are realistic qualification times t_q) a 2000-hour test can be reduced to a 3-days test. This means that after qualification of the product (where the distribution is recorded during over the qualification time t_q) no long duration test have to be performed for lot acceptance. Tests times in the order of days are sufficient to control (SPC) the life time of the product.

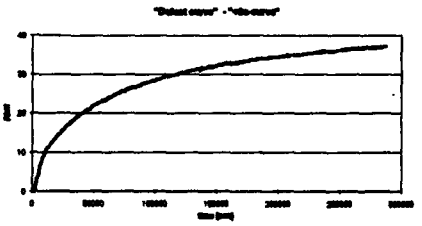


Fig 2: The minimum screening time can be derived as the time where the difference between the 3s-curve and the "defect curve is larger than the measuring resolution

6. CONCLUSIONS

In the standard testing (cumulative failure method) no real screening of the products take place. All products which withstand the test are considered equal. Companies which produce reliable products will benefit from the reliability system based on degradation kinetics:

1) good reliability indicators for products with equivalent specifications.
2) reduced screening times. Products with low intrinsic reliability and/or low reliability capability has to undergo long screening times.
3) "unreliable" products have to be aged more during screening. The life time of "unreliable" products will be reduced even more due to this screening strategy.
4) Companies which invest in good measuring equipment will benefit. Screening times can be reduced due to the small measuring resolution.
5) Degradation kinetics opens the possibility for built-in reliability and robust production (for lifetime)

All these benefits helps the companies, which produce reliable products, to lower their costs and to strengthen their market position. A quality system for reliability has to support this attitude!

A PHYSICS-OF-FAILURE (POF) APPROACH TO ADDRESSING DEVICE RELIABILITY IN ACCELERATED TESTING

Pradeep Lall, Michael Pecht
CALCE Electronic Packaging Research Center
University of Maryland
College Park, MD 20742, U.S.A.
Phone: (301) 405-5342, FAX: (301) 314-9269
email: lall@eng.umd.edu

Michael J. Cushing
US Army Materiel Systems Analysis Activity
Aberdeen Proving Ground, MD 21005, U.S.A.
Phone: (410) 278-2760, FAX: (410) 278-2043
email: cushing@arl.army.mil

ABSTRACT

Tests and screens are usually selected from government/military specifications, such as US MIL-STD-883, in the hope that the stresses and stress magnitudes will target dominant failure mechanisms and reveal the suitability of a technology or design for a given application environment. The testing approach presented in this paper determines test levels based on failure mechanisms, failure modes, and stresses for the application. It uses quantitative failure models and acceleration transforms, and adapts the knowledge of dominant failure mechanisms to the selection of accelerating stress parameters. The stress levels, designed specifically for each test article, are based on manufacturing processes, geometry and materials.

KEYWORDS

Physics of failure, accelerated testing, competing risks, electronics reliability, reliability testing

INTRODUCTION

Tests and screens are selected from government/military specifications, such as MIL-STD-883, based on the assumptions that the stresses and stress magnitudes will target the dominant failure mechanisms and reveal suitability of a technology or design for a given application environment. However, as noted by the US Air Force Rome Laboratory, in the January 1994 IEEE Reliability Society Newsletter, "Although manufacturers continue to use these screens today, most of the screens are impractical or need modifications for new technologies, and add little or no value for mature technologies. . . . With the defense budget/market declining, the DoD cannot afford costly test requirements that add little value to product quality or can be met in other ways."

Present technology is changing at such a pace that devices can be obsolete before they reach maturity. Also, new development in products and advances in technology, coupled with the need to improve product reliability, have increased demand for the application of accelerated tests. With the use of modern technology, products can be designed to operate for tens of years without failure. Accelerated tests are thus needed by design, reliability and manufacturing engineers to obtain timely information on the reliability of components and materials. Accelerated tests are needed in the reliability-design process to assess, demonstrate and improve component and subsystem reliability, certify components, detect failure modes, compare different manufacturers, and improve overall quality.

Satisfactory approaches to these issues have been elusive, which has precluded the widespread use of accelerated reliability testing. Failures must be identified with respect to the failure mechanism and stresses in order to address failures generically coupled with an understanding of the failure mechanism. Increased use of physics-of-failure concepts during product design & manufacturing may soon help reduce some of these difficulties, thereby increasing the usefulness & acceptance of accelerated reliability testing of electronic products.

APPROACH

Difficulties encountered in accelerated reliability testing of electronic products have limited its application & acceptance. Crucial, but difficult issues associated with the evaluation of accelerated reliability tests include:

- determination of the dominant failure mechanisms, sites and modes that are the weakest links in the product under intended life-cycle loads and how best to improve the design and/or manufacture.
- determination of appropriate stress levels for accelerated tests, such that, the product that passes the accelerated tests will have no dominant failure mechanisms that are likely to occur under intended life-cycle loads
- assessment of product reliability under intended life-cycle loads from data obtained under accelerated tests.

One of the keys to the proposed acceleration modeling approach is the explicit treatment of failure processes or mechanisms. Separate treatment of failure mechanisms, a central feature of the physics-of-failure approach, is recommended by leading authorities on the accelerated reliability testing of electronic devices since failure mechanisms can have *different life distributions* and acceleration models (e.g., [Nelson 1990; Tobias & Trindade 1986]). "Just as, in general, different failure mechanisms follow different life distributions, they may also have different acceleration models...we can study each failure mode and mechanism separately...This method is virtually the only way to do acceleration modeling successfully" [Tobias & Trindade, 1986, p.137]

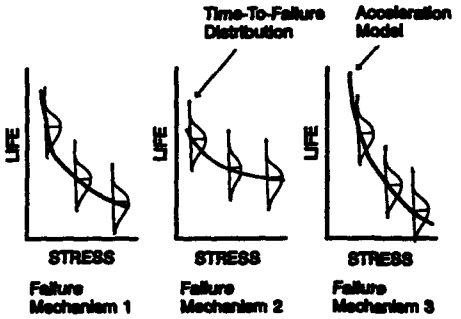


Figure 1 Time to failure versus stress for a hypothetical microelectronic device

Consider a microelectronic device which is subject to failure due to three dominant failure mechanisms acting at associated failure sites within the device. While the time to failure of each of the failure mechanisms may be influenced by changes in several parameters, including loads, geometries, material properties, defect magnitudes and stresses, focus is placed on the impact of stress on time to failure. Figure 1 depicts the relationship between the time-to-failure distribution of each of the three failure mechanisms, and the stress that each mechanism has a dominant dependence on. The failure mechanisms that cause hardware failures in electronic products generally do not have identical stress dependencies. For example, the temperature-stress dependencies of microelectronic failure mechanisms are known to vary considerably [Pecht, Lall 1992]. The relationship between the device time-to-failure and stress may be elusive unless failure mechanisms receive explicit treatment. Explicit consideration of dominant failure mechanisms, and their dependencies on loads, geometries and material properties, will provide the insight required to design & build reliability into electronic products, and compress reliability test time. The proposed accelerated test approach will include:

- Determine the likely failure mechanisms, based on current models for each potential failure mechanism. Include information on model applicability and limitations.
- Review the various stresses and their limits for the models of each failure mechanism in order to determine the accelerated stresses to be used in the test.
- During the test, perform a thorough failure analysis of each part, interconnection, etc. until the root-cause failure mechanism(s) is determined.
- Using the competing-risk analytical approach, estimate model parameters for each failure mechanism using only failures due to that failure mechanism, treating failures due to other failure mechanisms as censoring times.
- Extrapolate down to use stress for each failure mechanism.
- Use the competing-risk model to combine the times-to-failure distributions, of the failure mechanisms. This would result in a composite function appropriate for a non-repairable electronic item or the time-to-first-failure of a repairable electronic system. Any reliability parameters of interest can be calculated from this composite function.

ROLE OF THE PoF APPROACH IN ACCELERATED TESTING

What is the PoF approach? PoF is an approach to aid in the design, manufacture and application of a product by assessing the possible failure mechanisms due to expected life-cycle stresses. PoF is an approach to reliability assessment, not reliability prediction. Reliability assessment involves the evaluation of product potential to survive for the mission life in the application environment. Attributes for addressing product potential to survive mission life include: dominant failure mechanisms and the stress drivers for failure, and a pareto ranking of the times to failure of the dominant failure mechanisms.

How do microelectronic device failures manifest themselves? Failures can be broadly categorized by the nature of the loads—mechanical, thermal, electrical, radiation, and chemical—that trigger or accelerate the mechanism. Mechanical failures, for example, can result from elastic or plastic deformation, buckling, brittle or ductile fracture, interfacial separation, fatigue crack initiation and propagation, creep, and creep rupture. Thermal failures can arise by operating the component outside its thermal-performance specifications, through heating a component beyond its critical

temperature (such as the glass-transition temperature, melting point, or flash point), or by severely cycling the temperature. Electrical failures include those due to electrostatic discharge, dielectric breakdown, junction breakdown, hot electron injection, surface and bulk trapping, surface breakdown, and electromigration. Radiation failures are caused principally by uranium and thorium contaminants and secondary cosmic rays. Chemical failures arise in environments that accelerate corrosion, oxidation, and ionic surface dendritic growth (Table 1).

The PoF approach aids in determining potential causes and the location of failures and in developing effective accelerated tests. The PoF approach supports good engineering judgement regarding the impact of stresses on the product or elements of the product. Stresses can arise throughout the life-cycle, in manufacturing, handling, storage, and operation. The PoF approach has been implemented in the Computer-aided Design of Microelectronic Packages (CADMP-II) Software. CADMP-II is a set of integrated software programs that can be used to design and assess integrated circuit (IC), hybrid, and multichip module (MCM) packages. CADMP-II may be used in several ways:

- PoF approach will assist in rapidly assessing alternative microelectronic-package design solutions during qualification. Reliability assessment for each design alternative consists of a pareto ranking of the dominant failure mechanisms in terms of time-to-failure. The ability to satisfy mission life requirements, and avoid failure due to any of the dominant failure mechanisms during mission life, will provide a measure of relative worth and aid in selecting the best of the available options.
- Once a design is selected, the PoF approach may be used as a guide to design improvement by identifying the drivers for the dominant failure mechanisms during accelerated test design. The drivers may include package architecture, material properties and application environment. The merit of various design trade-offs can then be evaluated by determining the sensitivity of the dominant degradation mechanisms to failure mechanism drivers.
- The ability of a proposed design to resist any damage during accelerated testing may be assessed. This information can be used to modify a design or derive appropriate screens for minimal residual damage to good components.
- The PoF approach is useful to various other engineering analyses. Examples include: evaluation of new materials, structures, and technologies; assessment of packages designed by other software programs or manufacturers; maintenance strategy planners can make use of information concerning the potential failure sites, to minimize down time. due to fault isolation

ACCELERATED TEST MODELS: EMPIRICAL VS. POF MODELS

An accelerated test model is a mathematical equation used to relate the behavior of an item at one stress level to its behavior at another stress level. Models which are entirely based on curve fitting of test data are called empirical models. Models which describe behavior at the structural, or atomic level consistent with experimental results, are called structural, closed-form, constitutive, or physics-of-failure models. PoF models characterize various degradation mechanisms operational in microelectronic devices. The degradation mechanisms address mechanical, electrical, and chemical degradation. The mechanical degradation mechanisms address thermal mismatches at mating interfaces, spatial temperature gradient, time-dependent temperature-change induced stresses, vibration, and mechanical shock stresses. Electrical degradation mechanisms address temperature, ionic contamination and charge anomaly induced parameter drifts including intrinsic carrier concentration, thermal voltage, mobility, current gain, leakage current, collector-emitter saturation voltage, and VTC shift for Bipolar Devices; and threshold voltage, mobility, drain current, time delay, strong inversion leakage, subthreshold leakage, and chip availability for MOSFET devices. Chemical degradation mechanisms address moisture and contamination-induced corrosion. In addition, mass-transfer degradation mechanisms involving electromigration, stress driven diffusive voiding, metallization migration, contact spiking, and hillock formation have also been addressed. PoF models address potential degradation mechanisms in various package elements including the element attach, substrate, substrate attach, lead, lead seal, lid, lid seal, case, and interconnects.

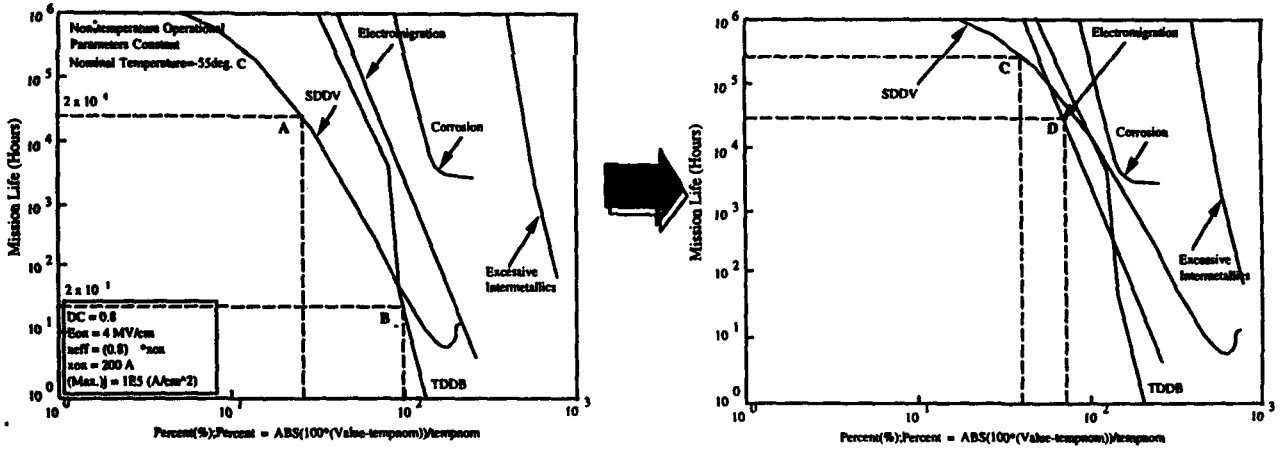
The mechanical, electrical and chemical degradation induced mechanisms have been classified into wearout and overstress mechanisms. Wearout mechanisms address the time to failure under user-specified defect magnitudes due to the accumulation of damage, while the overstress mechanisms address the failure of the component under the first application of the load. While, it would be physically impossible to list all the models in this paper, a representative sample of the models is given in table 2.

In general, a typical failure model can be represented as

$$t_f(i,j) = F\begin{pmatrix} g_{11} & g_{12} & g_{13} & \cdots & g_{1n} \\ m_{21} & m_{22} & m_{23} & \cdots & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots & \cdots \\ e_{41} & e_{42} & e_{43} & \cdots & \cdots \\ \delta_1 d_{51} & \delta_1 d_{52} & \delta_1 d_{53} & \cdots & \delta_1 d_{mn} \end{pmatrix} \quad (1)$$

where $t_f(i,j)$ is the time to failure due to a mechanism, g_{11}(for

Figure 2: The graph illustrates the method for identifying the optimal values of accelerated stress loads such as temperature, based on the sensitivity of dominant failure mechanisms. The accelerated test temperature at point A corresponds to failure due to SDDV, while the test temperature at point B corresponds to a much lower time-to-failure due to TDDB. Notice the failure mechanism shift, as the accelerated stress condition changes. The acceleration transform itself can be modified by changing the appropriate geometry, material, and defect magnitude parameters.



i = 1 to n) are the geometric parameters, m_{pi}(for i = 1 to n) are the material properties, p_{ri}(for i = 1 to n) are the operational parameters, e_{si}(for i = 1 to n) are the environmental stresses, d_{ai}(for i = 1 to n) are the defect magnitudes and l and r are user specified indexes to indicate the existence or non-existence of a defect; $\delta_{lr} = 0$ for $l \neq r$ and $\delta_{lr} = 1$ for l=r. For example, Black's electromigration model can be represented as

$$t_{fail\text{-}mech} = F \begin{pmatrix} w_{mat} & t_{mat} & 0 \\ A_{para} & E_a & n \\ j & 0 & 0 \\ T & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad (2)$$

In this representation of the Black's equation, parameters such as activation energy, current exponent and Black's electromigration coefficient are treated as material properties because their values vary for different metallization materials.

The models are based on the physics or mechanics of the degradation mechanisms and have been derived from the literature. To ensure that the equations are not applied beyond their ranges of validity, the references, along with the range over which the equation has been validated by experimentation, has been cited in the failure mechanism library. PoF models are used to identify the dominant degradation mechanisms of the component in the application environment specified by the life-cycle stress profile. Life-cycle stress profile refers to the magnitude and duration of all the loads the package is subjected to during its life-cycle including data on package manufacturing, assembly, storage, transportation, usage, and repair environments.

ACCELERATED TEST STRESS DESIGN

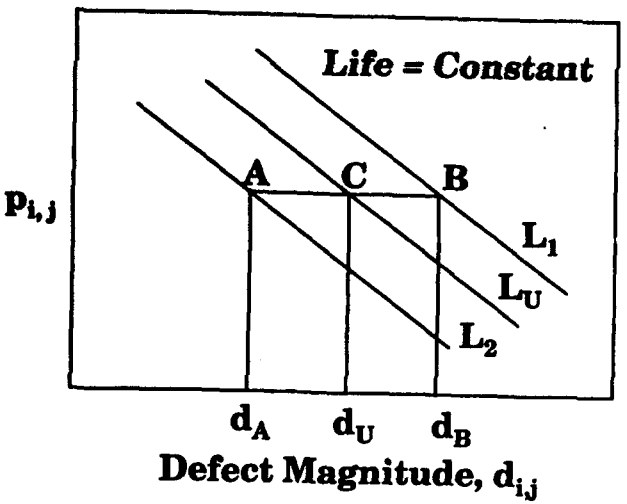
Sensitivity analysis is used to identify the optimal values of accelerated stress loads, such as temperature, based on the sensitivity of the dominant failure mechanisms to the loads. The methodology has been illustrated by an example. Sensitivity analysis involves identification of the functional relationships for each of the dominant failure mechanisms (listed in Table 2). Consider a device in which the dominant failure mechanisms consist of electromigration, stress driven diffusive voiding (SDDV), time-dependent dielectric breakdown (TDDB), corrosion, and excessive intermetallics. Based on the functional relationships in table 2, the time to failure due to various mechanisms is then calculated, versus change in the nominal value of temperature. The relative dominance of the failure mechanisms and the maximum operating temperature for desired mission life can then be assessed at various temperatures. Accelerated test temperature at point A will result in a failure due to SDDV after a test time of 2×10^4 hours (given by point A on figure 2). The dominant failure mechanism in this accelerated stress test is SDDV. The dominant mechanism however, will change to TDDB, if the test temperature were to be raised to point B. The dominant mechanism in this case will be TDDB with a accelerated test time of 20 hours. The acceleration transform such as shown in figure 2 can be used to relate the accelerated test time to time-to-failure under the intended life cycle stress profile. Further, the accelerated stress magnitudes can be derived from the acceleration transform to target various dominant mechanisms. The test time due to any one environmental stress can be further optimized based on the dependence of the dominant failure mechanisms on operating stresses. As an example, the dominant mechanisms in this case have a dominant dependence on the following operating stresses: duty cycle (DC), electric field across oxide (E_{ox}), effective oxide thickness (x_{eff}) and maximum current density in metallization (Max.j). A test time and the dominance of the failure mechanisms can thus be changed by changing the value of the operating stresses during accelerated test, to target various failure mechanisms and achieve optimal time compression.

DEFECT RELATED RELIABILITY

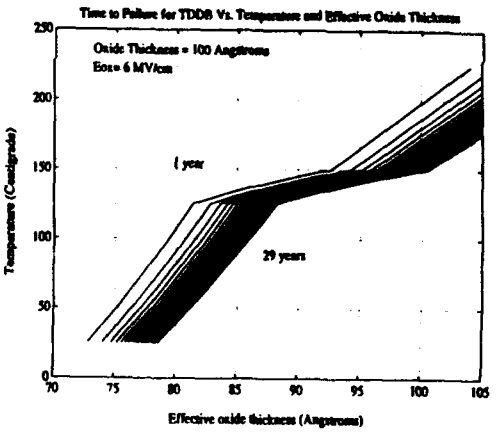
Defects are classified in three major categories: intrinsic material defects (including improper material properties); improper geometry (including improper size, cracks, improper thickness); and improper location (including misregistration errors). Effects of defect magnitudes can be addressed in terms of effect on time-to-failure or effect on performance parameters.

Defect information is used for both designing test structures for accelerated tests and for device screening. While the basic geometry, material, and operation parameters determine the nominal reliability, the worst-case defect magnitudes may be used to determine a defect-related reliability. Most users of microelectronic packages do not design their package, so the liberty of specifying the defect magnitudes does not generally exist. While some degree of defects exist in all product within the limits of the statistical process control, the allowable defect magnitudes can be derived based on life cycle stress profile and mission life requirements. Susceptibility to various failure mechanisms in a design can be assessed by evaluating the time-to-failure for potential failure mechanisms, during the life time of the product. Failure mechanisms which are most likely to result in device failure during mission life are then termed as the dominant failure mechanisms. Once the knowledge of the dominant field failure mechanisms has been obtained, the acceleration transforms relating the device life in operation to that in accelerated testing can be derived. Defect related reliability can then be addressed by screens derived by assessing the effects of the "worst case defect magnitudes"

Figure 3: Allowable defect magnitudes are determined based on the sensitivity of the dominant failure mechanisms from PoF models. The mathematical expression can plotted to evaluate the variability in life $L_1 < L_U < L_2$ versus operational stress parameter $p_{i,j}$ and defect magnitude $d_{i,j}$



(a)



(b)

during accelerated tests.

Allowable defect magnitudes are determined based on the sensitivity of the time-to-failure, due to potential failure mechanisms, from PoF models as follows:

$$For \quad j = k$$
$$If \quad \left.\frac{\partial t_f}{\partial d_{sj}}\right|_{j=k} > 0 \qquad (3)$$
$$Calculate \quad \min(d_{sj})|_{j=k} \; .AND. \; t_f > ML$$

$$For \quad j = k$$
$$If \quad \left.\frac{\partial t_f}{\partial d_{sj}}\right|_{j=k} < 0 \qquad (4)$$
$$Calculate \quad \max(d_{sj})|_{j=k} \; .AND. \; t_f > ML$$

where ML is the mission life. The mathematical expression can be plotted as figure 3 which shows the variability in life $L_1 < L_n < L_2$ versus an operational stress parameter p(i,j) and a defect d(i,j). The allowable defect magnitude is selected such that the lower bound on the variability in life due to the defect is greater than the mission life in the application environment. The minimum and maximum defect magnitudes which will not allow device failure during mission life (in the application environment specified in the life-cycle stress profile) is thus evaluated. These defect magnitude are then used to design screens. An example application of the concept is TDDB. TDDB is the formation of low resistance dielectric paths through localized defects in dielectrics such as thermally grown oxides or other oxides in MOS devices. Failures typically occur at weaknesses in the oxide layer due to poor processing or uneven oxide growth. Various studies have demonstrated a correlation between the low breakdown strength and presence of stacking faults in the oxides [Liehr, 1988]. Other studies have attributed early oxide breakdown to charge accumulation in oxide [Lee, 1988; DiStefano, 1975; Harai, 1978; Ricco, 1983; Holland, 1984] and to local thinning and discontinuities in the oxide caused by metal precipitates [Honda, 1984; Honda, 1985; Wendt, 1989]. The mechanism is characterized by the sudden, usually permanent, d.c. conduction in the dielectric of MOS capacitors. Typically, thin FET dielectric materials exhibit this breakdown failure mechanism depending on latent defect density, temperature, electric field intensity, ratio of the device operating potentials to the intrinsic dielectric strength, and the distances between the conductors in electronic packages which are defined by technology limits or electrical requirements. The screen for effective oxide thickness can be devised based on the life curve for TDDB versus temperature and effective oxide thickness (derived based on the [Moazzami, et al. 1989] model; Table 2).

Conversely, the adequacy of MIL-STD-883 can be assessed for a particular design. An example of defect-detection oriented non-stress screen is Method 2010; MIL-STD-883 which addresses scribing and die defects, diffusion and passivation faults, metalization — non-adherence, bridging, misalignment, scratches and voids; gate alignment and voiding; and dielectric isolation defects for active devices.

COMPETING CAUSES OF FAILURE

An important issue that arises when reliability assessment begins at the failure-mechanism level concerns development of a probabilistic time to failure for an electronic product. In particular, given information on probabilistic times to failure for relevant failure mechanisms acting at associated failure sites, a suitable approach is needed to developing probabilistic, product time to failure. An approach to this problem is competing-risk modeling.

Reliability quantification has frequently focused on the approximation of a constant hazard rate value for each device, which are then summed to obtain a constant hazard rate value for the circuit card assembly. The summing of constants continues up the indenture levels. This process focuses attention on quantifying the reliability of a product in terms of reliability parameters for each item contained within the product. By contrast, the competing-risk model focuses attention on the times to failure of the failure mechanisms & sites that dominate a competition to cause product failure. These times to failure are treated as individual random variables, regardless of whether the product is a device, circuit card assembly, or system, and a constant hazard rate is not assumed. Reliability, regardless of indenture level, is quantified directly from the times to failure of the failure mechanisms & sites that are dominant at that indenture level. Use of the competing-risk model supports & encourages modeling of the dominant product failure mechanisms & sites, which can provide extraordinarily powerful input to various reliability, maintainability and logistics analyses throughout the product life cycle. Though the competing-risk model is of general applicability, the following sections will focus on microelectronic devices due to their importance to electronic products.

Competing-Risk Modeling

A microelectronic device has a potential, random time to failure T_{device} from each of n root-cause failure mechanisms acting at associated failure sites within the device. In the series case, the microelectronic device fails when the first of the n times to failure occurs. This model is often called the competing-risk or weakest-link model [Tobias & Trindade 1986, p.164; Nelson 1990 p.378; Mann et al. p.142]. A list of the failure mechanisms & sites that are key to microelectronic-device reliability can be found in [Lall, et al. 1993]. If T_1, T_2,

..., T_n are the potential, random times to failure for the n failure mechanisms, then the time to failure of the microelectronic device is

$$T_{device} = \min(T_1, T_2, \cdots, T_n). \quad (5)$$

For a particular set of loads, the reliability of the device, as a function of time ($R_{device}(t)$), is given by

$$R_{device}(t) = Pr(T_{device} \geq t). \quad (6)$$

Thus

$$R_{device}(t) = Pr[(T_1 \geq t) \cap (T_2 \geq t) \cap \cdots \cap (T_n \geq t)]. \quad (7)$$

This is the general form of the competing-risk model and can perhaps be best evaluated with a monte carlo simulation, given sufficient data concerning the variation in the failure-mechanism times to failure. Sources of dependency between competing failure mechanisms include variation of local loads, geometries and material properties which affect more than one of the n failure mechanisms. The sources of dependency can be modeled in a simulation through appropriate distribution of the value selected for a load, physical dimension, or material property during a particular monte carlo trial. Because loads and geometries are stress analysis inputs, load and geometry variations affect the stress analysis. It would not be practical to embed a stress analysis within a monte carlo trial, but a load/geometry to stress transformation could be developed from stress analyses and embedded within the simulation.

Independent Case

If the n times to failure are statistically independent, then the reliability of the microelectronic device simplifies to

$$R_{device}(t) = Pr(T_1 \geq t) Pr(T_2 \geq t) \cdots Pr(T_n \geq t). \quad (8)$$

Thus

$$R_{device}(t) = R_1(t) R_2(t) \cdots R_n(t). \quad (9)$$

This is the special case of the competing-risk model for independent competing risks. It can be shown (*eg*, [Mann et al. pp.142-145]) that, in this case, the hazard functions of the n times to failure can be summed:

$$h_{device}(t) = h_1(t) + h_2(t) + \cdots + h_n(t). \quad (10)$$

If the n times to failure are not statistically independent, then the independent competing-risk model is not entirely appropriate, but can provide a lower limit on the reliability of the device when the times to failure are positively correlated. The reliability of the shortest time to failure can be used as an upper limit on the device reliability. This crude bracketing may be sufficient for some practical purposes [Nelson 1990 p.382]. It may also be possible to reduce sources of dependency, such as load variation, by performing separate physics-of-failure reliability assessments at selected load values.

Reliability Bounds

If the life distributions of the n failure mechanism times to failure are not known, it may still be possible to bound the reliability (or other parameters of interest) of a microelectronic device based on partial information concerning the n life distributions. The more complete the information about the n life distributions, the tighter the bounds can be made.

Reliability Bound Based on Means

Perhaps the simplest reliability bound available is based on first moments (*ie*, means) of the n failure-mechanism times to failure, provided the hazard rates of the failure mechanisms do not decrease with time. Or more formally stated, if the means of the n times to failure are known, and if:

(1) the microelectronic device is "coherent" (*ie*, improving the reliability of one failure mechanism can not decrease the reliability of the device)
(2) the failure-mechanism times to failure are either statistically independent or "associated" (*ie*, the times to failure tend to act similarly which is often the case because external loading affects many failure mechanisms adversely)
(3) the failure-mechanism times to failure have increasing or constant hazard functions (*ie*, wearout or constant hazard function applies to all failure mechanisms) with known mean lives $\mu_1, \mu_2, \cdots, \mu_n$
(4) the time period of interest, t, is less than the minimum of the known failure-mechanism mean times to failure,

then the reliability of the microelectronic device can be related to the mean failure-mechanism times to failure by [Barlow & Proschan, pp. 113-114]:

$$R(t) \geq e^{-t \sum_{i=1}^{n} \frac{1}{\mu_i}}. \quad (11)$$

While this bound will often turn out to be quite conservative in many applications, it offers a few distinct advantages:

- the failure-mechanism times to failure need not be statistically independent
- the precision of the probabilistic quantification of microelectronic-device reliability is consistent with the available information
- it provides an auditable link between probabilistic

microelectronic-device time to failure and analysis of the dominant failure mechanisms & sites.

Improved bounds which take advantage of additional information about the life distributions of the failure mechanism times to failure are available.

Reliability Bound Based on Mean and Variance

Barlow & Marshall developed a method [Barlow & Marshall 1965] that provides upper & lower reliability bounds for time-to-failure distributions with non-decreasing hazard functions or non-increasing hazard functions, given that the mean and variance (or standard deviation) are known. The bounding method for the non-decreasing hazard function is of greater interest here since a failure mechanism that is either subject to wearout or has a hazard function that is constant in time can be addressed. Others have found this method to provide tight reliability bounds given the available information on the time to failure distribution (*eg*, [Gertsbakh 1989, p. 64]).

Because no explicit analytical forms are available, these reliability bounds have been tabulated in [Barlow & Marshall 1965]. In order to use the tables, the mean must be normalized to 1 and the second moment about the origin must be calculated. A general relationship between moments of a distribution is [Hastings 1974, p 14]:

$$\mu_r = \sum_{i=0}^{r} \binom{r}{i} \mu'_{r-i} (-\mu_1')^i \,. \qquad (12)$$

Notation

μ_r	r^{th} moment about the mean
μ_r'	r^{th} moment about the origin.

By setting r equal to 2 in equation (12), the following relationship can be readily obtained:

$$\mu_2' = \mu^2 + \sigma^2 \,. \qquad (13)$$

Notation

μ	mean (or first moment about the origin)
μ_2'	second moment about the origin
σ^2	variance (or second moment about the mean)
σ	standard deviation.

Rearranging the terms of equation (13) gives

$$\mu^2 = \mu_2' - \sigma^2. \qquad (14)$$

The impact of normalizing the mean to 1 is readily seen when equation (14) is divided through by the square of the mean:

$$1 = \frac{\mu_2'}{\mu^2} - \frac{\sigma^2}{\mu^2} \,. \qquad (15)$$

In order to use the tables in [Barlow & Marshall 1965], the time t must also be divided by the mean.

With this method, the reliability of each of the dominant failure mechanisms can be bounded. Regardless of the indenture level of the item to be quantified, each of the dominant failure mechanisms for the item should be bounded and then the competing-risk model applied. This method would be best applied to only the dominant product failure mechanisms because of the conservative nature of bounds. In the independent case of the competing-risk model, for example, bounds will be multiplied together. Hazard functions should not be approximated for low indenture-level items (*eg*, devices) and then summed to obtain a composite hazard function for higher indenture-level items (*eg*, a circuit card assembly), as this procedure will incur unnecessary inaccuracies.

NUMERICAL EXAMPLE

For the purposes of illustrating these concepts & methods, consider a microelectronic device with two dominant failure mechanisms. While it may be optimistic to hypothesize a device with only two dominant failure mechanisms, the clarity of the illustration will be improved.

A typical physics-of-failure reliability assessment output ranks the dominant failure mechanisms. When only the mean or median time to failure is considered, the failure mechanisms are ranked from the shortest to the longest mean or median time to failure, with the shortest given highest priority. However, if the failure mechanism with the second shortest mean time to failure has a time-to-failure distribution that is much more dispersed than the failure mechanism with the shortest mean time to failure, then it should be given higher priority for corrective action. This is one shortcoming of ranking failure mechanisms according to mean times to failure.

Figure 4 depicts Weibull probability density functions (PDF) for the two dominant failure-mechanism times to failure of the microelectronic device. Since both of the failure mechanisms are subject to wearout, the Weibull shape parameters are greater than one. Characteristics of the Weibull time-to-failure distributions are:

PDF_1

- scale parameter η = 100,000 hours
- shape parameter β = 2
- mean μ = 88,600 hours

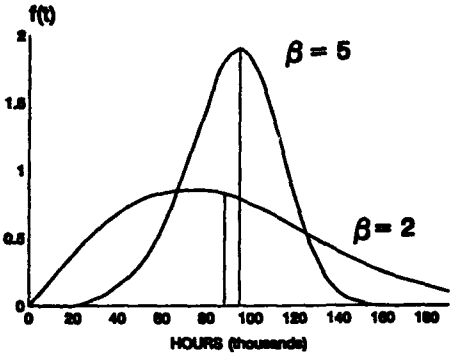


Figure 4 Weibull probability density functions

- variance $\sigma^2 = 2.146 \times 10^9$ hours^2.

PDF_2

- η = 100,000 hours
- $\beta = 5$
- μ = 91,800 hours
- $\sigma^2 = 4.423 \times 10^8$ hours^2.

If only the means are considered, the failure mechanisms appear to be:

- roughly comparable in importance
- not likely to result in failure of the device provided the desired lifetime for the device is significantly shorter than 90,000 hours.

For desired device lifetimes of 50,000, 60,000 or 75,000 hours, the Weibull reliability values are provided in table 3. The reliability values of failure mechanism 1 are much lower than those of failure mechanism 2 indicating that corrective action should focus on failure mechanism 1.

It is frequently the case that only partial information is available for failure mechanism times to failure. If the mean and variance for the failure mechanism times to failure are known, the reliability of the failure mechanisms can be bounded using the method and tables of [Barlow & Marshall 1965]. The second moment about the origin for failure mechanisms 1 and 2 can be calculated from equation (13):

- failure mechanism 1 $\mu_2' = 10^{10}$
- failure mechanism 2 $\mu_2' = 8.873 \times 10^9$.

t (hours)	$R(t)$ Failure Mechanism 1	$R(t)$ Failure Mechanism 2
50,000	0.78	0.97
60,000	0.70	0.93
75,000	0.57	0.79

Table 3 Time-dependent reliability using Weibull distributions

In order to use table III in [Barlow & Marshall 1965], μ_2' must be divided by the square of the mean and time t must be divided by the mean. The lower bounds for failure mechanisms 1 and 2 are listed in table 4. These lower bounds are quite close to the reliability values calculated from complete information on the time-to-failure distributions (table 3). Consequently, a more accurate assessment of these two failure mechanisms can be made using this approach compared with the case when only the means are considered. Failure mechanism 1 would be readily recognized as the weaker link.

t (hours)	$R(t)$ Failure Mechanism 1	$R(t)$ Failure Mechanism 2
50,000	0.72	0.91
60,000	0.64	0.86
75,000	0.49	0.71

Table 4 Time-dependent reliability using bounds

Provided the failure mechanism times to failure are statistically independent, the device reliability for desired lifetimes t equal to 50,000, 60,000 and 75,000 hours are calculated using equation (9), and are listed in table 5.

SUMMARY

This paper addresses the infrastructure for screening and accelerated testing methodology. The methodology will facilitate the derivation of appropriate test and screen stresses and their magnitudes. Use of the physics-of-failure approach

t (hours)	$R_{device}(t)$ (Weibull)	$R_{device}(t)$ (Bound)
50,000	0.75	0.65
60,000	0.65	0.55
75,000	0.45	0.35

Table 5 Time-dependent device reliability

during product design can supply the needed models. Because the acceleration models of failure mechanisms vary, multiple accelerated test profiles may be preferable in order that some failures or assurance is obtained on each of the dominant failure mechanisms. A physics-of-failure reliability assessment can be used to improve fault detection during an accelerated reliability test, which is often conducted periodically and at ambient conditions. Use of physics-of-failure concepts can provide the dominant failure mechanisms, sites, modes, and the stress conditions under which those failure modes are most easily detected. This information can be used to tailor and prioritize the fault detection provisions.

REFERENCES

Barlow, R.E., Proschan, F., *Statistical Theory of Reliability and Life Testing*, Holt, Reinhart & Winston, 1975.

Barlow, R.E., Marshall, A.W., "Tables of bounds for distributions with monotone hazard rate", J. Amer. Statist. Assoc., 60, 1965, pp 872-890.

Black, J. R.,"Physics of Electromigration," IEEE Proceedings of the International Reliability Physics Symposium, pp. 142-149, 1983.

Chen, I.C., Holland, S.E., and Hu, C., "Electrical Breakdown in Thin Gate and Tunneling Oxides.", *IEEE Transactions on Electron Devices*, ED-32 (1985), pp. 413.

DiStefano, T.H., Shatzkes, M., "Dielectric Stability And Breakdown In Wide Bandgap Insulators", *J. Vac. Sci. Technol.*, Vol.12, No.1, pp.37, Jan./Feb. 1975.

Gertsbakh, I.B., *Statistical Reliability Theory*, Marcel Dekker, Inc., New York, 1989.

Harari, E., "Dielectric Breakdown In Electrically Stressed Thin Films of Thermal SiO_2", *J. Appl. Phys.*, Vol.49, No.4, pp. 2478, April 1978.

Hastings, N.A.J., Peacock, J.B., *Statistical Distributions*, John Wiley & Sons, New York, 1974.

Ho, P.S. and Howard, J.K. *Journal of Applied Physics*, 45 (1974), 3229, quoted from Ho, P. S. and Kwok, T. Electromigration on Metals, *Reports on Programs in Physics*, 52 (1989), pp. 301-348.

Holland, S., Chen, I.C., Ma, T.P., and Hu, C., "On Physical Models For Gate Oxide Breakdown", *IEEE Electron Device Letters*, Vol. 5, No. 8, pp. 302, August 1984.

Honda, K., Ohsawa, A., Toyokura, N., "Breakdown In Silicon Oxides(II)-Correlation With Fe Precipitates", *Appl. Phys. Lett.*, Vol.46, No. 6, pp. 582, March 1985.

Honda, K., Ohsawa, A., and Toyokura, N. Breakdown in Silicon Oxides: Correlation with Cu Precipitates. *Applied Physics Letters*, 45 (1984), 270.

Howard, Robert T., "Packaging Reliability: How to Define and Measure It", *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, (1982), 376-384.

Kato, M., Niwa, H., Yagi, H., and Tsuchikawa, H. Diffusional Relaxation and Void Growth in an Aluminum Interconnect of Very Large Scale Integration. *Journal of Applied Physics*, 68 (1990), pp. 334-338.

Kidson, G.V., "Some Aspects Of The Growth Of Diffusion Layers In Binary Systems", Journal of Nuclear Materials 3, No.1, pp. 21-29, 1961.

Kopanski, J., Blackburn, D.L., Harman, G.G., Berning, D.W., "Assessment of Reliability Concerns for Wide-Temperature Operation of Semiconductor Device Circuits," *Transactions of the First International High Temperature Electronics Conference*, Albuquerque, New Mexico, 1991, pp. 137-142.

Lall, P., Pecht, M., "An integrated physics-of-failure approach reliability assessment", Proc. 1993 ASME [illegible] Electronics Packaging Conf., EEP-vol 4-1, 1993, pp 509-524.

Lee, J., Chen, I.C., and Hu, C., "Modeling And Characterization Of Gate Oxide Reliability", *IEEE Transactions on Electron Devices*, Vol.35, No. 12, pp.2268, December 1988.

Lee, J., Chen, I.C., and Hu, C. Statistical Modeling of Silicon Dioxide Reliability. *26th Annual International Reliability Physics Symposium*, (1988), p.. 131-138.

Liehr, M., Bronner, G.B., Lewis, J.E., "Stacking-Fault-Induced Defect Creation in SiO_2 on Si(100)", *Appl. Phys. Lett.*, Vol. 52, No. 22, pp.1892, May 1988.

Mann, N.R., Schafer, R.E., Singpurwalla, N.D., *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley & Sons, New York, 1974.

Moazzami, R., Lee, J., Hu, C., "Temperature Aceleration Of Time-Dependent Dielectric Breakdown", *IEEE Transactions on Electron Devices*, Vol. 36, No. 11, July 1989.

Moazzami, R., Lee, J., Chen, I.C., Hu, C., "Projecting The Minimum Acceptable Oxide Thickness For The Time-Dependent Dielectric Breakdown", *IEDM Tech. Digest*, pp.710, 1988.

Moazzami, R., Hu, C., "Projecting Gate Oxide Reliability And Optimizing Reliability Screens", *IEEE Transactions on Electron Devices*, Vol. 37, No. 7, July 1990.

Nelson, W., *Accelerated Testing, Statistical Models, Test Plans, and Data Analyses*, John Wiley & Sons, New York, 1990.

Niwa, H., Yagi, H., and Tsuchikawa, H. Stress Distribution in an Aluminum Interconnect of Very Large Scale Integration. *Journal of Applied Physics*, 68(1) (1990), pp. 328-333.

Pecht, M.,(editor), *Quality Conformance and Qualification of Microelectronic Packages and Interconnects: A Focus On Reliability*, John Wiley and Sons, New York, New York, June 1994.

Pecht, M., Lall, P., Hakim, E., "Temperature Dependence on Integrated Circuit Failure Mechanisms," Chapter 2, *Advances in Thermal Modeling III*, ed. by A. Bar-Cohen and A. D. Kraus, New York, IEEE and ASME Press, Dec. 1992.

Pecht, M.,(editor), *Integrated Circuit, Hybrid and Multichip Module Package Design Guidelines: A Focus On Reliability*, John Wiley and Sons, New York, New York, June 1994.

Philosky, E., "Intermetallic Formation In Gold-Aluminum Systems", Solid State Electronics, 13, pp. 1391-1399, 1970

Philosky, E., "Design Limits When Using Gold Aluminum Bonds", 9th Annual Proceedings IEEE Reliability Physics Symposium, Las Vegas, Nevada, pp. 11-16, 1971

Philosky, E.,"Purple Plague Revisited", 8th Annual Proceedings of the International Reliability Physics Symposium, Las Vegas, Nevada, pp. 177-185, 1970

Ricco, B., Azbel, M.Ya.,and Brodsky, M.H., "Novel Mechanism For Tunneling And Breakdowon of Thin SiO_2 Films", *Phys. Rev. Lett.*, Vol. 51, No. 19, pp.1795, Nov. 1983

Texas Instruments, private communications with Donald Lipps, 1993.

Tobias, P.A., Trindade, D.C., *Applied Reliability*, Van Nostrand Reinhold, New York, 1986.

Wendt, H., Cerva, H., Lehmann, V., Pamler, W., "Impact Of Copper Contamination On The Quality Of Silicon Oxides", *J. Appl. Phys.*, Vol. 65, No. 6, pp. 2402, March 1989

ACKNOWLEDGEMENTS

This project has been funded by a grant from the US Army with additional support from NASA and the twenty-eight members of CALCE EPRC. CADMP-2 software has been developed jointly by a team of researchers and computer programmers at the CALCE Electronic Packaging Research Center (NSF SIUCRC Center of Excellence in Electronic Packaging), University of Maryland.

Table 1. Failure sites, operational loads, and failure mechanisms

		Contaminants	Relative humidity	Steady-state temperature	Relative humidity cycle	Temperature cycle	Thermal shock	Radiation*	Vibration/mechanical shock	Voltage/current/charge
Case		J	J	Z			Q, J	O, G	Q, R, M	X
Die		Y, U	J	Y, X, T, U, V, W		A, R, M	A, R, M	F, P, G, S	A, R, M	J, K, L, X, T
Attachment				Z		A, B, Q, R, M	A, B, Q, R, M	G	A, B, C, R, M	
INTERCONNECTS	Flip-chip solder joints	J	J	L		A, B, H, R, M	A, B, C, Q, R, H, J, N	G	A, C, Q, R, M	N
	Tape automated bonds	J, K	J,K	L		B, E, Q, R, M	A, B, C, E, Q, R, J, K, N	G	A, C, D, E, R, M	K, N, X
	Wirebond interconnects	Q, J, L	J, Q	L		A, B, E, Q, R, M,	A, B, C, E, Q, R, M, L		A, C, D, E, Q, R, M	J, L, X
Leads		J, K	J,K			B, R, M	A, B, C, M, J		A, B, C, M	J, K, M, X
Lead seals		Q, J	J		J	A, Q, R, M	A, Q, R	O	A, Q, M	
Lid seals						A, Q, H	A, Q, H	O	A, H	
Substrate						A, R, M	A, B, R, M		A, B, R, M	J

Overstress failure mechanisms: A. brittle fracture, B. ductile fracture, C. yield, D. buckling, E. large elastic deformation, F. single-event upset/soft-error, G. radiation-induced thermal breakdown

Wear-out failure mechanisms: H. creep, J. corrosion, K. dendritic growth, L. interdiffusion, M. fatigue crack propagation, N. diffusion, O. insulation failure, P. excessive leakage currents, Q. interfacial de-adhesion, R. fatigue crack initiation, S. Frenkel defects, T. electromigration, U. metallization migration, V. stress-driven diffusive voiding, W. charge anomalies (hot electrons, slow trapping), X. EOS/ESD, Y. ionic contamination, Z. depolymerization.

* Radiation stress may include gamma rays, alpha particles, fast neutrons, cosmic rays, electrons, and protons.

Table 2: A Representative Sample of the Models Used for PoF Approach

Failure Mechanism	Failure Model Equation/Reference	Model Type	Assumptions and Range of Validity
Electromigration Blacks Model	$$t_{i\sigma\text{-}mean} = \frac{w_{\text{-}met}\, t_{\text{-}met}}{j^{n} A_{para}\, e^{-E_a/K_B T}}$$ where $t_{i\sigma\text{-}mean}$ is the mean time to failure (hours), $w_{\text{-}met}$ is the metallization width (cm), $t_{\text{-}met}$ is the metallization thickness (cm), A_{para} is a parameter depending on sample geometry, physical characteristics of the film and substrate, and protective coating, j is the current density (A/cm^2), K_B is the Boltzmann constant, n is an experimentally determined exponent (n = 2 for Black's model), E_a is the activation energy (eV) and T is the steady-state temperature (Kelvin). [Black 1983].	Wearout	Not valid for Bamboo structures; value of current exponent is typical of Al-1%Si metalization; current exponent and geometry parameter value changes with material. Model assumes metallization open as failure mode; more typical of single-layer metallization rather than multilayer metallizations where failure results in resistance increase.
S.D.D.V. Kato and Niwa Model	$$t_{if} = \frac{K_B T\, l^3}{20 D_{\sigma gb}\, t_{\text{-}gb}\, \sigma_{\text{-}met} V_{at\sigma atm}}$$ where t_{if} is the time to failure in seconds, K_B is Boltzmann's constant, T is the temperature in Kelvins, $D_{\sigma gb}$ is the grain-boundary diffusivity (meter^2/sec), $t_{\text{-}gb}$ is the grain-boundary thickness (meter), $V_{at\sigma atm}$ is the atomic volume for conductor atoms (meter^3), $\sigma_{\text{-}met}$ is the stress in metallization, and 2l is the void separation. [Kato et al. 1990] [Niwa et al. 1990].	Wearout	Assumes plastic deformation due to passivation cooling — occurs instantly; failure is assumed a function of c[illegible]ional relaxation.
TDDB Fowler-Nordheim Tunneling Model	$$t_{\sigma ox,bd}(T) = \pi_{prop} e^{\gamma_f \frac{t_{\sigma ox}}{V_{\text{-}ox}}\left(1 + \frac{C_{para\text{-}ox}}{E_a}\left(\frac{1}{T}-\frac{1}{300}\right)\right) - \frac{E_a}{K_B}\left(\frac{1}{T}-\frac{1}{300}\right)}$$ where, $t_{\sigma ox,bd}$ is the time to breakdown, π_{prop} is the room temperature value of the pre-exponential, γ_f is the electric field acceleration parameter, $t_{\text{-}ox}$ is the effective oxide thickness E_a is the activation energy, $V_{\text{-}ox}$ is the voltage across the oxide, $C_{para\text{-}ox}$ is an empirical factor determined experimentally, T is the steady state operating temperature, K_B is Boltzmann's constant. [Lee et al. 1988] [Holland et al. 1984] [Chen et al. 1985] [Moazzami et al. 1988].	Wearout	Charge build up at defects in oxides; Breakdown of the oxide occurs when a critical charge density is reached in the oxide to trigger the breakdown process; defects are modeled as localized oxide thinning.
Excessive Intermetallics Kidsons Model	$$t = \frac{x^2}{D_0\, e^{\frac{Q}{RT}}}$$ where, t is the time to failure due to intermetallic formation, x is the critical intermetallic layer thickness, and T is the steady state operating temperature, Q is the activation energy for intermetallics R is the rate constant. [Kidson 1961].	Wearout	Assumes parabolic diffusion law; has been shown to be valid for gold-aluminum bonds [Philofsky 1970,1971].

FAULT DETECTION IN ANALOGUE CIRCUITS BY DYNAMIC SUPPLY CURRENT MONITORING

M. A. Al-Qutayri

Dept. of Electronic & Elect. Engineering
De Montfort University, The Gateway
Leicester, LE1 9BH, U.K

e_mail: maq@dmu.ac.uk

Abstract

This paper compare transient voltage and current monitoring for analogue circuits testing. A coefficient of variation parameter is derived to perform the comparison. A simple fault model is described to evaluate the effectiveness of the time-domain based testing technique.

1. Introduction

The demand for analogue ICs, especially those realised in CMOS and BiCMOS technologies, is steadily increasing due to new application areas in fields such as telecommunications, automotive control systems, consumer electronics ... etc. The increase in demand requires the reliability of the analogue ICs to match that of the digital ICs.

However, the state-of-the-art computer tools that are used in the design and test of digital circuits are much more advanced than those for analogue circuits. This is due to the complex characteristics of analogue circuits which depend on continuous descriptive variables, the problem of tolerance, and the lack of an adequate fault model.

To enable the assessment of the effectiveness of a testing strategy in terms of fault-coverage a fault-model is required. In this paper a simple fault-model at the transistor level is derived for analogue circuits. The model is then used to assess a the performance of a time-domain testing technique. In this technique both the transient output voltage and transient supply current are monitored to detect the presence of a fault.

2. Fault Model

For digital ICs a number of fault models are available. The models enable the generation of tests and evaluation of the quality of tests. However, for analogue circuits an adequate fault model virtually does not exist, and this is one of the main reasons for the lack of automatic test pattern generation algorithms for such circuits.

Therefore, to evaluate the effectiveness of the proposed time-domain testing technique, a fault model at the transistor level was derived to synthesis catastrophic faults in an analogue IC. The model, shown in Fig.1, is based on the studies of ICs yield [1] and prominent modes of failure of MOS transistors [2]. It indicates that the likely single faults are: drain-open, source-open, gate-drain-short, gate-source-short, and drain-source-short. These faults are caused by open circuits in the diffusion and metallization layers, and short circuits between adjacent diffusion and

metallization layers. No probabilities are associated with each fault.

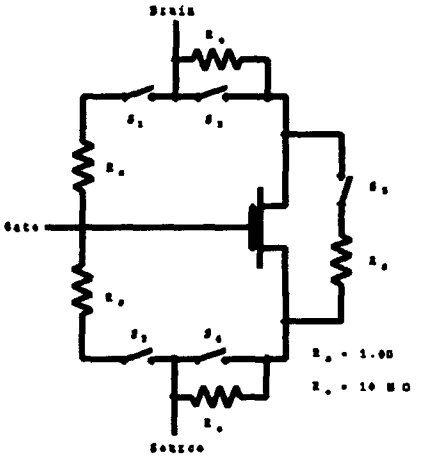


Figure 1: MOS Fault Model

3. Time-Domain Technique

The time-domain testing strategy [3] is based on the excitation of the circuit-under-test (CUT) with a pseudo random binary sequence (PRBS) of pulses, and subsequent measurement of the transient response at the output node/s. Both the transient voltage at the output node/s and the transient current i_{DDT} of the supply current are measured. The transient current (i_{DDT}) testing technique is similar to the IDDQ testing [4] used for digital circuits, except that IDDQ testing is performed under static conditions while i_{DDT} testing is performed under dynamic conditions.

The PRBS was selected as a test vector [5] because it can be readily generated by a digital tester, such sequences have well defined properties, can be used, if necessary, to extract the impulse response of an analogue CUT and the pulse width can be tailored to force large Fourier components to fall within the sensitive region of the circuit response. If the CUT is a mixed-signal network, the response generated by applying a PRBS test signal can be used as a signature to characterize the network under-test.

4. Voltage and i_{DDT} Monitoring

To establish whether a fault was detected or not, the response of the CUT is compared with that of the fault-free toleranced response. If the CUT response falls outside the bounds of the fault-free toleranced response at least at one sampling instant then the fault is considered detectable. Otherwise, it is assumed undetectable.

The number of instances (i.e samples) at which the CUT response falls outside the tolerance envelope are then counted, and the percentage of deviation from the ideal response is accumulated. A parameter called the Coefficient of Variation (CV) is calculated for each fault that was detected at least at one instant. The objective of calculating CV is to determine which type of measurement, voltage or current, detects a particular fault with higher degree of confidence. The higher the difference between the fault-free response and CUT response the higher the confidence in the measurement. CV is then normalised to make it easier to compare the results of the two sets of measurements.

The time-domain testing technique and the fault detection process outlined above, were applied to the BiCMOS low-pass filter circuit shown in Fig.2. A total of 70 single fault conditions were introduced to the network. Of these faults 60 were catastrophic faults in the MOS op-amps, and 10 soft faults in the resistive and capacitive components. The soft faults ranged in variations between ±25% and

±50% of a component fault-free value.

When comparing the CUT responses monitored at Vout and i_{DDT} with the respective fault-free toleranced responses, the results indicate that a fault-coverage of 100% was achieved by each measurement. However, computation of the normalised CVs for Vout and i_{DDT} showed that 50 faults are best detected by Vout while the other 20 faults are best detected by i_{DDT}. The results are summarized in Fig.3. Note that due to the scaling factor some faults appear as if they were not detected.

5. Conclusion

The results presented above show that the time-domain testing technique is an effective method for the detection of faults in analogue circuits. The use of the PRBS as a test vector resulted in a high fault-coverage. However, the PRBS performance will need to be compared with other forms of test vectors. This will be the subject of future work. The paper also showed the need for a simple yet effective fault-model for analogue ICs

Comparison of the transient voltage and current results indicate that the two methods are complementary. Therefore, i_{DDT} monitoring is expected to play a major role in testing analogue ICs once the technique is developed further. The major advantage of i_{DDT} monitoring is that it does not require access to internal circuit nodes.

6. References

[1] W. Maly, Strojwas and S. Director, "VLSI yield prediction and estimation: a unified framework", IEEE Trans. Computer-Aided Design, Vol. CAD-5, No.1, pp. 114-130, January 1986.

[2] J. Galiay, Y. Crouzet and M. Vergniault, "Physical versus logical fault model MOS LSI circuits: impact on their testability", IEEE Trans. Computers, Vol. C-29, No.6, pp. 527-531, June 1980.

[3] M.A. Al-Qutayri and P.R. Shepherd, "Go/No-Go testing of analogue macros", IEE Proc. Circuits, Devices and Systems - Part G, Vol.139, pp.534-540, August 1992.

[4] C. Hawkins, J. Soden, R. Fritzemeier and L. Horing, "Quiescent power supply current measurement for CMOS IC defect detection", IEEE Trans. Industrial Electronics, Vol.36, No.2, pp. 211-218, May 1989.

[5] M.A. Al-Qutayri and P.R. Shepherd, "PRBS testing of analogue circuits", IEE Colloquium on Testing Mixed-Signal Circuits, Savoy Place, London, May 1992.

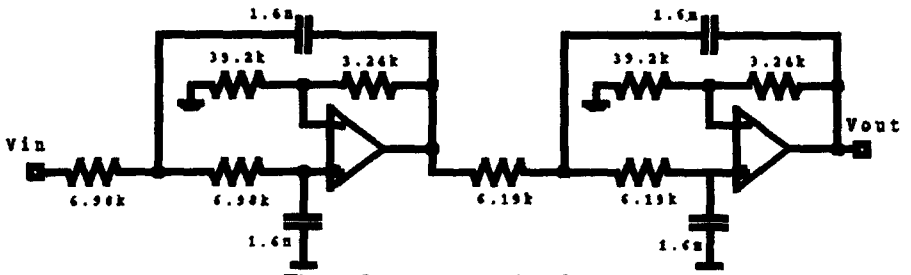


Figure 2: Low-Pass Filter Circuit

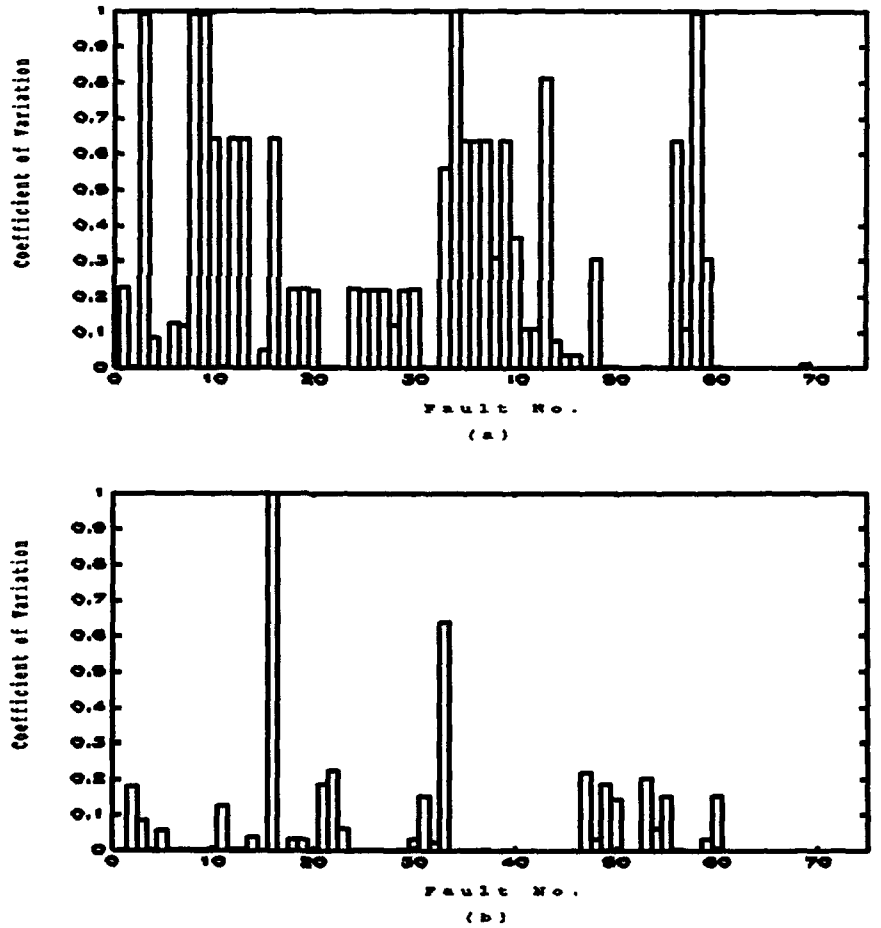


Figure 3: Normalised CVs of low-pass filter circuit. (a) at Vout (b) at i_{DDT}

TESTING OF FLASH EEPROM STRUCTURES BY MEANS OF Co^{60} γ IRRADIATION AND CONSTANT CURRENT STRESS

A. Paccagnella
Istituto di Elettrotecnica — Università di Cagliari
piazza d'Armi — I-09123 Cagliari, Italy

G. Crisenza, G. Ghidini
Central R&D — SGS-Thomson Microelectronics
via C. Olivetti, 2 — I-20041 Agrate Brianza, Italy

A. Rizzato, A. Scarpa, E. Zanoni
Dipartimento di Elettronica ed Informatica — Università di Padova
via G. Gradenigo, 6/A — I-35131 Padova, Italy

P.G. Fuochi
CNR-FRAE — Area della Ricerca
via P. Gobetti, 101 — I-40129 Bologna, Italy

ABSTRACT

The effects of Co^{60} γ irradiation and Costant Current Stress on MOS capacitors with tunnel oxide have been investigated. Four different device technologies have been used. We have observed a qualitative agreement between the results obtained with the two methods, indicating that also radiation testing can be a valuable tool to investigate the quality of thin tunnel oxide layers.

1. INTRODUCTION

The thin tunnel oxide is one of the most important technological steps of the structure of the flash memory cells [1, 2]. Testing of the oxide quality, in terms of breakdown voltage, trap and defect density, represents one of the key issues for the assessment of the production yield, device performance and long-term reliability. Different electrical testing methods have been implemented and are widely applied for these purposes [3, 4].

The operating conditions of a flash memory require to sustain the injection of large current densities into the tunnel oxide by Fowler–Nordheim injection of free carriers [5]. The study of the evolution of the trapping mechanisms active in the dielectric film under these conditions is then a key issue for device performance and reliability.

In this work, we have compared for the first time results obtained on MOS capacitors through Fowler–Nordheim (F.N.) stresses with data achieved through exposure to γ rays. Ionizing radiation induces in fact generation of hole–electron pairs in the SiO_2 film, which can be subsequently trapped. This procedure has been considered as an alternative technique to F.N. for the analysis of trapping phenomena in SiO_2.

Both F.N. and radiation stresses have been performed on four different structures based on thin tunnel oxide.

2. TEST STRUCTURE DEVICES

The devices used throughout this work were fabricated by SGS–Thomson Microelectronics (Agrate Brianza, Italy) and consist of MOS capacitors on ⟨100⟩ p-Si, $N_A = 7 \cdot 10^{15}$ cm^{-3}, with area 10^{-3} cm^2. The oxide has been grown in dry O_2 ambient at 950 °C, and has passed a post-oxidation annealing step in N_2 for 30 minutes; its thickness is 125 Å.

Two different device structures have been considered:

1. poly-Si-2 / ONO (Oxide-Nitride-Oxide) / poly-Si-1 / tunnel oxide / p-Si; (ONO-type devices);

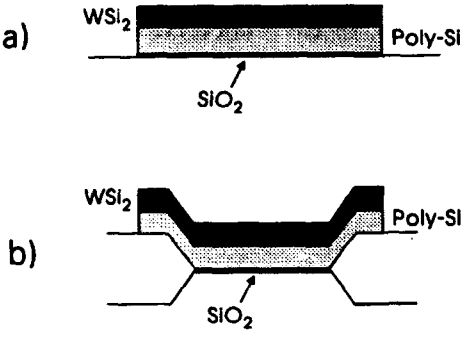


Figure 1: *Different periphery technologies. a) Self-Aligned (SA); b) Field Oxide (FOX).*

2. poly-Si–2 / poly-Si–1 / tunnel oxide / p-Si; (DPCC-type devices).

The devices' periphery was either Self-Aligned (SA) without bird's beaks, or formed by a Field Oxide (FOX), as depicted in Fig. 1.

The capacitors used in our experimental works are listed below:

- TA1: ONO-type with FOX periphery;
- TA2: ONO-type with SA periphery;
- TA3: DPCC-type with FOX periphery;
- TA4: DPCC-type with SA periphery.

These devices are part of test structures currently used in the reliability evaluation of the flash memory tunnel oxides.

The ONO layer, used to isolate the floating gate of the flash device, also acts as a barrier against diffusion of process-induced contaminants (mainly F) from WSi_2 into the tunnel oxide. Such impurities could adversely affect the oxide quality [6, 7].

3. EXPERIMENTAL PROCEDURES

Ten samples for each technology have been irradiated with Co^{60} γ at CNR–FRAE up to a cumulative dose of 700 krad(Si). Devices have been purposely kept unbiased during irradiation, in order to avoid problems due to device microbonding, which has been observed to induce some modifications of the device electrical properties.

The electrical measurements have been completed within few hours after irradiation, in order to minimize the room-temperature annealing phenomena of

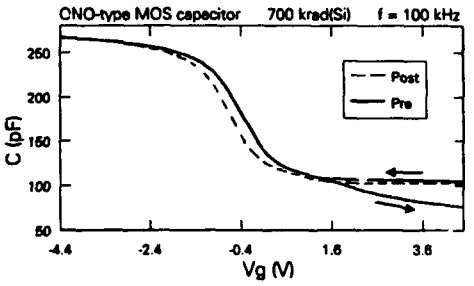


Figure 2: *C–V curves at 100 kHz of an ONO-type device before irradiation (solid line) and after Co^{60} γ irradiation at 700 krad(Si) (dashed line).*

the radiation induced charge. Experimental characterizations consist of C–V measurements at various frequencies in the range 1 kHz $\div$ 800 kHz [8]. It was not possible to achieve quasi-static C–V curves, as the devices could not reach the thermal equilibrium condition with acceptable sweep rates [9].

Other samples have been subjected to Constant Current Stress (CCS) with an injection current density $J = 10^{-2}$ A/cm^2. This value has been chosen in order to simulate the operative stress conditions of the tunnel oxide in flash EEPROM memories.

During the CCS tests, the voltage V_{FN}, required to mantain the constant injection throught the Fowler–Nordheim tunneling process, has been carefully monitored.

These stresses have been conducted up to the injection level of 1.5 C/cm^2, which approximately corresponds to the charge injection conditions of a flash memory after 10^5 write/erase cycles.

4. RESULTS

4.1. Radiation Tests

Gamma irradiation of a MOS structure induces the creation of electron/hole pairs in the SiO_2 film. Due to the low mobility, practically only holes are trapped in the oxide, thus giving an oxide positive charge (Q_{oxr}). Correspondingly, new Si/SiO_2 interface states (N_{it}) are also created [10].

The positive charge trapped in the oxide shifts the C–V curves towards negative voltages, while the interface states give rise to a stretch-out of the characteristics [10].

Assuming the approximation that, in a MOS structure, the interface states in the upper half of the silicon band-gap are acceptor-like [11], while those

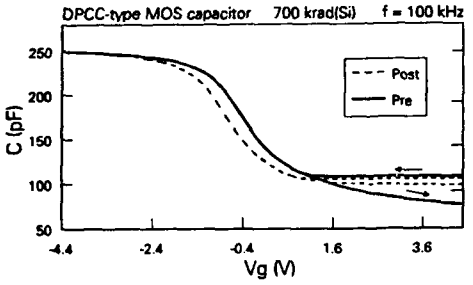


Figure 3: *C–V curves at 100 kHz of a DPCC-type device before irradiation (solid line) and after Co^{60} γ irradiation at 700 krad(Si) (dashed line).*

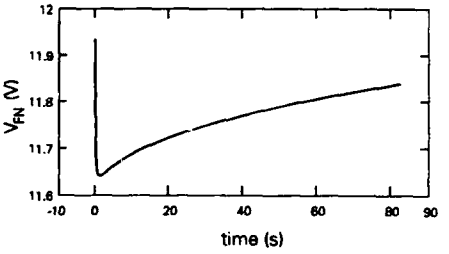


Figure 4: *CCS curve of an ONO-type device, negatively stressed with a 0.01 A/cm² injection current density; V_{FN} is the voltage applied between gate and substrate.*

in the lower half are donor-like, when the Fermi level at the semiconductor surface matches the intrinsic Fermi level (midgap condition), all the interface states are in a neutral charging condition. Thus the measured shift of the applied voltage in the midgap condition induced by irradiation, ΔV_{mg}, takes into account only the oxide trapped charge [10]:

$$\Delta V_{mg} = -\frac{\Delta Q_{oxt}\bar{x}_t}{C_{ox}T_{ox}} \qquad (1)$$

where $\bar{x}_t$ is the centroid of the corresponding charge distribution starting from the poly/oxide interface, C_{ox} the oxide capacitance and T_{ox} the oxide thickness.

Figures 2 and 3 respectively show the shift of ONO-type and DPCC-type C-V curves due to the radiation stress. Neither frequency dispersion nor hysteresis phenomena affect the charactersistics of the irradiated capacitors. However, the deep inversion condition cannot be reached any more. This indicates that γ radiation tests have produced only a small amount of SiO_2/Si interface states below the gate. Minority carriers affecting the deep inversion condition can be supplied from adjacent regions, likely below the thick field oxide, where a much higher N_{it} can be expected. In fact, N_{it} strongly scales with the oxide thickness [12].

Mean values and data dispersions of ΔV_{mg} are reported in Tab. I: DPCC-type devices show a larger charge trapping than ONO-type ones, while no other effects related with different periphery technologies can be detected.

Differences observed in Tab. I between ONO and DPCC samples depend on the DPCC F contaminations, which lead to a higher level of hole trap density, at least nearby the Si/SiO_2 interface.

Table I: *Mean values $\langle \Delta V_{mg} \rangle$ and corresponding dispersion $\sigma_{\Delta V_{mg}}$ of the shift of the midgap voltage after γ irradiation, averaged on ten samples, for all studied technologies.*

	ONO		DPCC	
	FOX	SA	FOX	SA
$\langle \Delta V_{mg} \rangle$	−0.248	−0.244	−0.370	−0.364
$\sigma_{\Delta V_{mg}}$	0.0065	0.0056	0.0108	0.0145

4.2. Constant Current Tests

The CCS can produce creation of new Si/SiO_2 interface states and bulk electron traps, and mainly trapping of negative and positive charge in the oxide film [13].

Figure 4 shows a typical CCS curve on an ONO-type device. Increasing the injection level (that is the total injected charge quantity), a continuous build-up of negative charge takes place in the oxide (Q_{ox}), and, correspondingly, the Fowler–Nordheim threshold voltage V_{FN} shifts to more negative values. If the capacitor is in accumulation, this quantity can be related to the stress-induced oxide charge ΔQ_{ox} by means of the the following relation, neglecting any possible contribution coming from interface states [14]:

$$\Delta V_{FN} \simeq \frac{\Delta Q_{ox}}{C_{ox}}\left(1 - \frac{\bar{x}}{T_{ox}}\right) \qquad (2)$$

where $\bar{x}$ is the centroid of the charge distribution

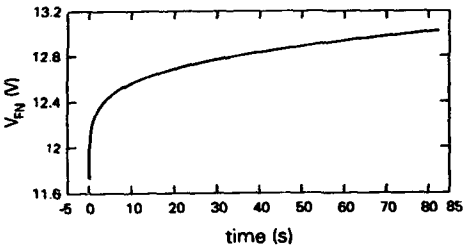


Figure 5: *CCS curve of a DPCC-type device, negatively stressed with a 0.01 A/cm² injection current density; V_{FN} is the voltage applied between gate and substrate.*

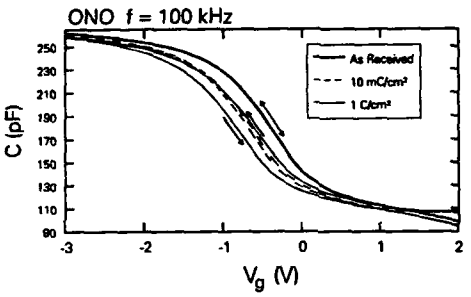


Figure 6: *C–V curves at 100 kHz of an ONO-type device after different injection levels; stresses have been conducted with negative gate bias.*

calculated with respect to the poly/oxide interface. Actually, at the beginning of the stress the so-called "turn-over" takes place: ΔV_{FN} becomes positive, indicating hole trapping in the oxide and after an injection of 1.3 mC/cm² the negative trapping is increasingly dominating. On the other hand, the same injection on a DPCC structure gives completely different results: almost no turn-around is visible for the enhanced negative charge trapping in the presence of F contaminats (Fig. 5)

In Tab. II, the mean ΔV_{FN} and the corresponding dispersion are reported, measured on ten samples for each technology, after a CCS up to an injection level of 1.5 C/cm². DPCC-type structures show a larger shift, indicating a higher negative trapped charge, than ONO ones.

While no effect due to interface states was observed after γ irradiation, it can be clearly detected by C–V measurements after CCS tests. Figures 6

Table II: *Mean values (ΔV_{FN}) and corresponding dispersion $\sigma_{\Delta V_{FN}}$ of the shift of the F.N. voltage after a 1.5 C/cm² CCS, averaged on ten samples, for all studied technologies.*

	ONO		DPCC	
	FOX	SA	FOX	SA
(ΔV_{FN})	−0.197	−0.186	−0.687	−0.682
$\sigma_{\Delta V_{FN}}$	0.0057	0.0121	0.0743	0.0323

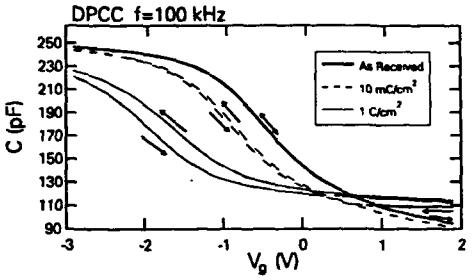


Figure 7: *C–V curves at 100 kHz of a DPCC-type device after different injection levels; stresses have been conducted with negative gate biasing.*

and 7 show the C–V characteristics at 100 kHz, obtained after different injection levels. The curve shift towards negative voltages, apparently due to positive oxide trapped charge, is larger in DPCC capacitors. However, a hysteresis of the characteristics appears and increases with the CCS injection level.

C–V measurements at different frequencies have also been made on electrically-stressed devices. Figures 8 and 9 show C–V curves obtained at 1 and 100 kHz on capacitors which had been previously stressed at an injection level of 1.5 C/cm². A big frequency dispersion appears, which is larger on DPCC devices, and also the curve shift to negative voltages is strongly frequency dependent. These data indicate that Fowler–Nordheim stresses introduce a substantial increase of the interface state density, clearly controlling hysteresis and frequency dispersion. Even the shift of the C–V curves should be related not only to oxide trapping of positive charge, but also to interface states which control the results at such a high injection level.

Several γ-irradiated samples have been also subjected to CCS tests. We have only observed a slight

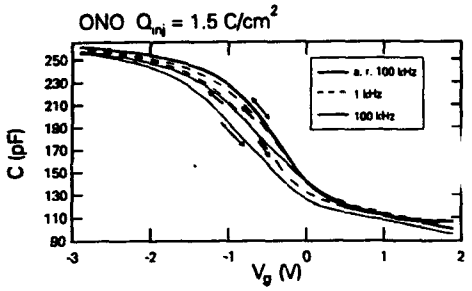


Figure 8: *Frequency dispersion of C–V curves of an ONO-type device after a 1.5 C/cm^2 negative injection; the 100 kHz curve of the unstressed device is also reported (there was not frequency dispersion before CCS test.)*

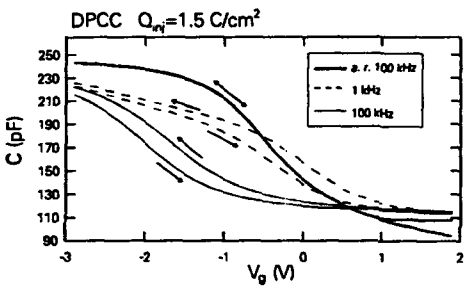


Figure 9: *Frequency dispersion of C–V curves of a DPCC-type device after a 1.5 C/cm^2 negative injection; the 100 kHz curve of the unstressed device is also reported (there was not frequency dispersion before CCS stress).*

decrease of the $|V_{FN}|$ starting value of about 50 mV due to the positive charge trapped in the oxide. However, ΔV_{FN} values follow the same curves of the unirradiated devices, suggesting that no new oxide electron trap has been created by the irradiation.

5. DISCUSSION

ΔV_{FN} measurements have all been performed with a negative gate bias: they are more sensitive to charge trapped in the oxide region closer to the poly electrode. Our results reveal a trapping of negative charge after CCS tests, apart from a short "turnover" effect on ONO devices. The negative charge is much larger on DPCC than on ONO devices for a given injection level. Only minor modifications of ΔV_{FN} are observed on irradiated devices.

C–V curves are more sensitive to oxide charge trapped near the silicon substrate. DPCC samples show a larger Q_{ox} then ONO ones after both irradiation and CCS tests, at least at moderate injection levels, as at high injection levels (> 1 C/cm^2), V curves show dramatic deformations. This means that both radiation and CCS tests can supply valuable informations about the trapping characteristics of injected holes for a given oxide technology. Incidentally, the oxide with more effective hole trapping at the Si/SiO_2 interface, i.e. the DPCC one, is also the most effective in electron trapping at the poly/SiO_2 interface.

6. CONCLUSIONS

We have compared for the first time two methods to study charge trapping in very thin tunnel oxides: γ irradiation and Constant Current Stress.

Both of them can lead to hole trapping at the Si/SiO_2 interface, even though CCS induces also a high density of interface states, as demonstrated by C–V measurements.

Devices fabricated with different technologies have been tested, and results obtained with the two methods are in agreement: DPCC-type capacitors have a higher density of hole traps than ONO ones, likely due to F contaminations in the oxide, without any effects related to the technology of the periphery.

CCS stress can also supply valuable informations on electron traps, which cannot be studied by using γ radiation tests.

The investigation method based on γ irradiation can be improved in order to achieve a better knowledge of oxide hole traps, for instance by irradiating devices under bias, by annealing studies or quasi-static C–V measurements.

It is worth to remark that no traps have been introduced into the oxide film at the radiation levels used in this work, indicating that no noticeable perturbation of the oxide structure is introduced.

REFERENCES

[1] P. Cappelletti, A. Panchieri and L. Ravazzi, "Mastering Key Factors Which Affect Flash Memory Reliability", in *Proc. of ESREF 93*, Bordeaux, France, 1993.

[2] A. Watts, "Built-in Reliabilty for 10 FITs Performance on EPROM and FLASH MEMORY", Technical Article TA109, SGS–Thomson Microelectronics, November 1991.

[3] P. Cappelletti, P. Ghezzi, F. Pio and C. Riva, "Accelerated Current Test for Fast Tunnel Oxide Evaluation", in *Proc. of IEEE Int. Conf. on Microelectronic Test Structures*, volume 4, March 1991.

[4] P. Cappelletti, P. Ghezzi, F. Pio and C. Riva, "Wafer Level Tunnel Oxide Reliability Evaluation By Means of the Exponentially Ramped Current Stress Method", *Microelectronics and Reliability*, vol. 33, n. 10, pp. 1579–1596, 1993.

[5] M. Lenzlinger and E. Snow, "Fowler–Nordheim Tunneling Into Thermally Grown SiO_2", *Appl. Phys. Lett.*, vol. 40, pp. 278–283, 1969.

[6] P. Ghezzi, F. Pio and C. Riva, "Impact of Fluorine of LPCVD WSi_2 Deposition on Thin Films", in *Proc. 3rd Workshop on Thin Dielectrics in Microelectronics*, Bordeaux, France, October 1991.

[7] C. Papadas, G. Ghibaudo, G. Pananakakis, F. Pio, C. Riva and P. Ghezzi, "On the Endurance performance of FLOTOX EEPROM Cells with WSi_2 Overcoated Floating Gate Electrode", *Microelectronics Journal*, vol. 24, n. 4, 1993.

[8] B.J. Gordon, "C–V Plotting: Myths and Methods", *Solid-State Technology*, page 57, January 1991.

[9] M. Kuhn, "A Quasi-Static Technique for MOS C–V and Surface State Measurements", *Solid-State Electronics*, vol. 13, pp. 873, 1970.

[10] F.B. Lean, Jr. H.E. Boesch and T.R. Oldham, "Electron-Hole Generation, Transport and Trapping in SiO_2", in T.P. Ma and P.V. Dressendorfer, editors, *Ionizing Radiation Effects in MOS Devices and Circuits*, ch. 3. John Wiley & Sons, New York, 1989.

[11] K. Ziegler, "Distinction Between Donor and Acceptor Character of Surface States in the Si-SiO_2 Interface", *Appl. Phys. Lett.*, vol. 32, n. 4, pp. 249–251, 1978.

[12] C.R. Viswanathan and J. Maserjian, "Model for Thickness Dependence of Radiation Charging in MOS Structures", *IEEE Trans. on Nuclear Science*, vol. NS-23, n. 6, pp. 2151–2156, 1982.

[13] M.-S. Liang and C. Hu, "Electron Trapping in Very Thin Thermal Silicon Dioxide", in *IEDM Technical Digest*, pp. 396–399, 1981.

[14] C. Papadas, P. Morfouli, G. Ghibaudo and G. Pananakakis, "Analysis of the Trapping Characteristics of Silicon Dioxide After Fowler–Nordheim Degradation", *Solid-State Electronics*, vol. 34, n. 12, pp. 1375–1379, 1991.

UPPER VOLTAGE AND TEMPERATURE LIMITATIONS OF STRESS CONDITIONS FOR RELEVANT DIELECTRIC BREAKDOWN PROJECTIONS

R.-P. VOLLERTSEN
SIEMENS Components, Inc., c/o IBM, 1000 River Street, Essex Junction, VT 05452
W. W. ABADEER
IBM Microelectronics Division, 1000 River Street, Essex Junction, VT 05452

1. ABSTRACT

Parameters and their dependence on the stress conditions are discussed using a three-dimensional model, including all possible degrees of freedom. The results, applied to breakdown data for 10 nm oxide, reveal the upper limits for stress conditions. A physical explanation for the upper voltage limit is presented. Furthermore, the thermal activation energy and the slope of lifetime distributions (e.g., the shape factor β of the Weibull distribution) are investigated as functions of electric field and temperature. The results are compared to available literature data.

2. INTRODUCTION

Projection of dielectric reliability to field conditions is one of the key issues during MOS-IC technology development. For dielectric reliability projection, the 1/E model (Ref. 1) is mostly used for field acceleration and the Arrhenius model for thermal acceleration. The 1/E model correlates the breakdown time to the Fowler-Nordheim tunneling current. Both the 1/E model and Arrhenius model achieve satisfactory results if failure times of an individual failure percentage (e.g. t_{50}) are analyzed. Furthermore, it has been reported that the acceleration parameters of the 1/E model are temperature-dependent (Ref. 2) and the activation energy changes with electric field (Ref. 2-4). However, the behavior at different failure percentages, i.e., the behavior of the entire lifetime distribution which is reflected by the slope of the distribution, is seldom considered. Data suggesting that the slope changes with field and temperature are available for extrinsic (Ref. 3, 4) as well as for intrinsic distributions (Ref 5). A relationship has been derived, based on the 1/E-model, which describes the dependence of the slope of the distributions on the electric field (Ref 6).

This paper discusses the temperature-field dependence of the oxide breakdown. It highlights the importance of considering the behavior of the slope of lifetime distributions as a function of both field and temperature. The results are used to analyze breakdown data. The upper stress-condition limits for the applicability are revealed for both field and temperature. An explanation of the upper limit for accelerated field stress is presented and supported by data. Furthermore the dependence of the thermal activation energy is investigated, resulting in a new aspect.

3. THEORETICAL CONSIDERATIONS

The relation between breakdown times at different electric fields according to the 1/E model is given by:

$$t_i = C\,(t_a / C)^{(E_a / E_i)} \qquad (1)$$

where t_i represents the breakdown time of interest at the field E_i, and t_a represents the breakdown time measured at accelerated field E_a. C is a time constant for a given temperature and failure percentage.

3.1 Projections with dependent parameters

Usually the breakdown times at one percentage, e.g., 50% or 63.2% for log normal or Weibull, respectively, are used to derive the acceleration parameters. The combination of the 1/E- and the Arrhenius plot results in a three-dimensional diagram (Fig. 1). Figure 1 illustrates two ways to project from highly accelerated conditions to operation conditions .

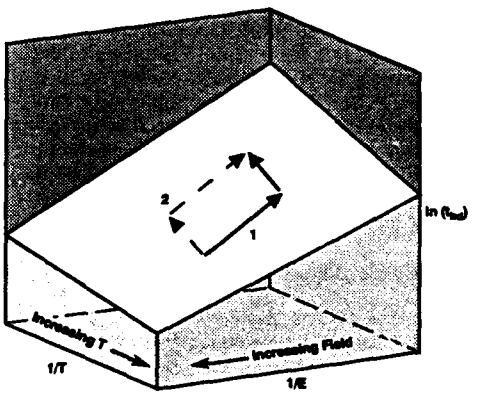


Fig. 1: Schematic dependence of time to breakdown on field and temperature at one failure percentage.

On path 1 the field is extrapolated first, succeeded by the temperature extrapolation for which it is required to know the dependence of the thermal activation energy on the electric field. This path is described by:

$$t_i = C\,(t_a / C)^{(E_a / E_i)} \cdot \exp\{\Delta H(E_a) \cdot (1 / T_i - 1 / T_a) / k\} \quad (2)$$

where k is the Boltzmann's constant, T_i and T_a are the temperatures and $\Delta H(E_a)$ is the field-dependent activation energy at field E_a. The second path first extrapolates to the new temperature and then to the field of interest:

$$t_i = C(T_i)\,\{t_a \cdot \exp[\Delta H / k \cdot (1 / T_i - 1 / T_a)] / C(T_i)\}^{(E_a / E_i)} \quad (3)$$

with the temperature dependent parameter $C(T_i)$ at the temperature T_i. From Fig. 1, it can be concluded that C, which is the time at 1/E=0, also has a thermal activation energy that follows the Arrhenius model as proposed in Ref. 2:

$$C(T_i) = C(T_a) \cdot \exp\{\Delta H_C \cdot (1 / T_i - 1 / T_a) / k\} \quad (4)$$

Because C decreases with temperature, a negative activation energy ΔH_C is determined. From equations 2 and 3, the function for the field-dependent thermal activation energy can be derived:

$$\Delta H(E_i) = \Delta H(E_a) \cdot E_a/E_i + \Delta H_C \cdot (1 - E_a / E_i) \quad (5)$$

This equation yields the same form as the expression given in Ref. 2 but does not contain the parameter G, which is constant as long as the same failure percentage is considered.

The plane (for each failure percentage) in the $t_{bd}(E, T)$ diagram (Fig. 1) can be described using equations 1-3. In most cases it is necessary to extrapolate to low percentages, i.e., to a lower plane. The planes of the $t_{bd}(E, T)$ diagram are not necessarily parallel, e.g. as a consequence of the 1/E dependence. Furthermore, data are available which suggest that the slope of breakdown distributions can deviate from the behavior predicted by the 1/E and Arrhenius models. Therefore, the slope is an additional parameter which requires consideration for lifetime projections.

3.2 Modelling the slope of breakdown distributions

It has been shown (Ref. 6) that the degradation behavior could change with failure percentage F; i.e., C is not a constant for all percentages of one distribution, and is rather substituted by C(F). If C is known for one percentage (e.g., at 50% or 63.2% for log normal or Weibull, respectively), C(F) can be expressed in a linear system by:

$$C(F) = C \cdot \exp(y \cdot K_{Er}) \quad (6)$$

where y is the value on the linear y-axis of the probability net and K_{Er} is a parameter to account for the deviation of the slope from the one predicted by the 1/E-model with C(F)=const.. K_{Er} can be determined from at least two distributions at different fields at the same temperature:

$$K_{Er} = (1/s_1 - E_r / (E_1 \cdot s_r)) / (1 - E_r / E_1) \quad (7)$$

where s_1, s_r are the slopes of the two distributions at the corresponding fields E_1, E_r. K_{Er} is independent of the acceleration but associated with the reference field E_r and slope s_r. The slope s_i at the field of interest E_i can be calculated by:

$$s_i = (E_r / (E_i \cdot s_r) + K_{Er} \cdot (1 - E_r / E_i))^{-1} \quad (8)$$

The effect of the temperature on the slope can be treated analogously. Applying the Arrhenius model to each percentage of a distribution results in parallel-shifted distributions. If data show a change of the slope with temperature, this could be interpreted as a (slight) change of the activation energy along the distribution. To account for this change of the activation energy ΔH, the percentage-dependent term $y \cdot K_T$ is added to the Arrhenius equation:

$$t_2 = t_1 \cdot \exp((\Delta H + y \cdot K_T) \cdot (1 / T_2 - 1 / T_1) / k) \quad (9)$$

where y is the value on the linear y-axis of the probability net and K_T is a parameter to account for the deviation of the slope from a parallel line. K_T can be determined from at least two distributions at different temperatures at the same electric field:

$$K_T = (1/s_2 - 1/s_1) / ((1 / T_2 - 1 / T_1) / k) \quad (10)$$

where s_1, s_2 are the slopes of the two distributions at the corresponding temperatures T_1, T_2. K_T is independent of the acceleration. The slope s_i at the temperature of interest T_i can be calculated from the slope s at temperature T by:

$$s_i = (1/s + K_T \cdot (1 / T_i - 1 / T) / k)^{-1} \quad (11)$$

For K_T=0, the distributions are parallel. s in equation 11 could be substituted for the s_i of equation 9, with the new equation describing the combined effect of temperature and field acceleration on the slope. Our treatment did not include a possible dependence of K_{Er} on the temperature and K_T on the electric field. The

number of parameters to be determined for an accurate projection with the highest degree of freedom is already large. A smart selection of the projection path and starting point can reduce the number of parameters without affecting the accuracy.

4. EXPERIMENTAL

n^+-polysilicon gate capacitors on n-substrate were fabricated with a standard MOS process. The oxide thickness was 10 nm with an active oxide area of 0.002 cm^2. The capacitors were stressed at wafer level, with temperatures ranging from 30 to 190°C and constant electric fields from 9 to 13 MV/cm. Constant current stresses at 30 and 150°C at various current densities were also performed on transistors with $3.4 \cdot 10^{-7}$ cm^2 gate areas.

5. RESULTS

5.1 Electric field dependence

t_{bd}-distributions from different electric fields are shown in Fig. 2. The model parameters were determined by multiple linear regression, with the regression equation derived from equations 1, 6 and 8. A low field (e.g. operation conditions) was used as a reference field and the four parameters $t_{63.2\%}$, s_r, C and K_{Er} were calculated simultaneously for that condition by the multiple regression from all data in the field range of 9 to 12 MV/cm. The parameters for the straight lines were then calculated using equations (1) and (8). The resulting values of C and K_{Er} are $2 \cdot 10^{-10}$s and 0.05, respectively. C is in good agreement with the value expected at 150°C according to Ref. 2. The small value of K_{Er} indicates almost no deviation from the 1/E model. The slope of these distributions is plotted versus the electric field in Fig. 3. Equation 8 is used to calculate the solid line. The slope increases with increasing field as expected (Ref. 6). The stronger increase of the slope at fields above 12 MV/cm suggests a change of the degradation mechanism. This marks one upper limit for relevant stress conditions because a too-steep distribution would yield over-optimistic projections. The observed steep increase of the slope is contrary to reported data in Ref. 6-7 where the slope decreased at a certain high field. Plotting t_{bd} versus 1/E reveals another upper limit.

Figure 4 shows the $t_{63.2\%}$ from Fig. 2 with data from other works (Ref. 6, 8, 9). In all cases, it is more appropriate to fit the data by two straight lines rather than by one, i.e. the parameter C decreases at high fields and yields a too optimistic projection. This limit for relevant stress fields depends on oxide thickness and

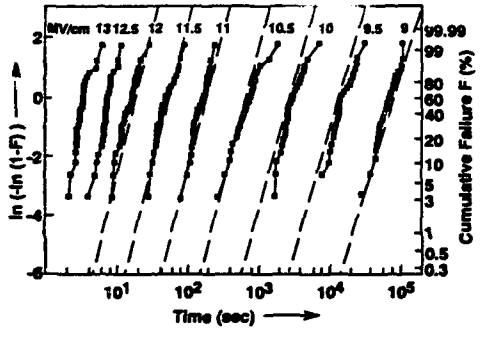


Fig. 2: t_{bd}-distributions of 10 nm SiO_2 at different fields. The straight lines represent results from multiple linear regression.

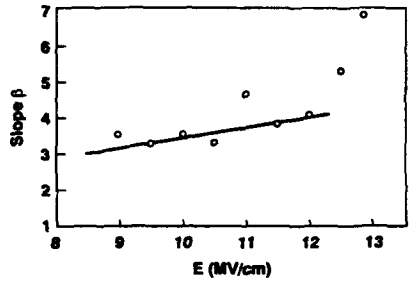


Fig. 3: Slope β of the Weibull distribution versus the electric field. Data from Fig. 2 compared to prediction from equation (8).

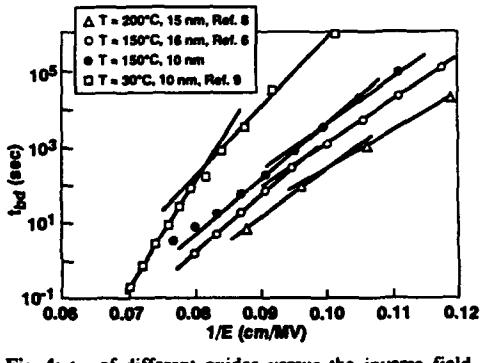


Fig. 4: t_{bd} of different oxides versus the inverse field show a temperature and oxide thickness dependent transition from low-field to high-field degradation behavior.

temperature. It has been reported that a drop of Q_{bd} coincides with the field where the transition occurred (Ref. 6). Therefore, constant current stresses of NMOS-transistor gates with small oxide area were performed. Fig. 5 shows the resulting t_{bd} plotted versus 1/E at 30°C and 150°C. The 150°C data clearly show a transition at a breakdown field of about 14 MV/cm; however, the transitions is not as clear for the 30°C case and may be easily overlooked. This demonstrates the temperature dependence of the transition. According to a physical model (Ref. 10), the observed transition occurs when the dominant degradation mechanism changes from trap creation at low fields to band-gap ionization due to carriers with energies exceeding 9 eV with respect to the bottom of the oxide conduction band (Ref. 10). Curves resulting from this physical model are shown for the 30°C data in Fig. 5 and demonstrate good agreement. As long as there is no way to separate the two mechanisms, product-relevant projections should be based only on data in the field range below the transition point.

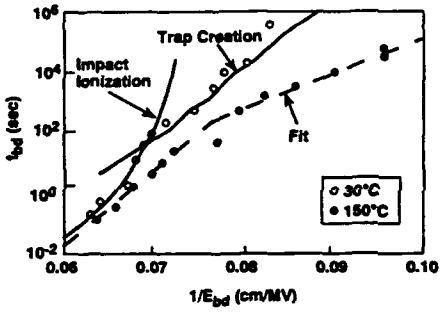


Fig. 5: t_{bd}-values from two temperatures plotted versus the inverse electric field. A physical model fits the 30°C data well. The 150°C data clearly show the upper stress limit.

5.2 Temperature dependence

Fig. 6 shows $t_{63.2\%}$ from different temperatures and stress fields in an Arrhenius plot. Two straight lines are required to fit the data of each field, which indicates a change of the activation energy. The transition occurs at around 90°C at high fields and shifts to higher temperatures as the field decreases. These results for the first time show a field dependency of the transition point for the activation energy. Data reported so far indicated changes of the activation energy at about 130°C (Ref. 11) or at 150°C (Ref. 2, 12). In addition, Fig. 6 shows that the slope of the fitted lines increases with decreasing field in both the lower and the higher temperature range. This reflects the field dependence of the activation energy previously reported (Ref. 2-4, 8).
The activation energies determined from Fig. 6 for the lower and higher temperature range are plotted versus the electric field in Fig. 7. For comparison, data (Ref. 3-4, 8, 13) representing different oxide thicknesses as well as intrinsic and extrinsic breakdown, are also shown. The activation energies determined in this work in the higher temperature range join well the data from intrinsic breakdown at high temperatures (125-400°C) of 15 nm oxide with a comparable oxide area (Ref. 8). The activation energies determined in this work in the lower temperature range rather follow the data extracted from Ref. 13. The different field dependencies of both temperature ranges restrict the determination of the thermal acceleration parameters to a range which includes the operation temperature. If data of the higher temperature range are included into the calculation of the activation energy, over-optimistic projections will result.
It's interesting to note that the activation energies reported for extrinsic breakdown (Ref. 3-4) have a significantly stronger dependence on the electric field than data from intrinsic breakdown. A communality of the activation energies given for a wide range of

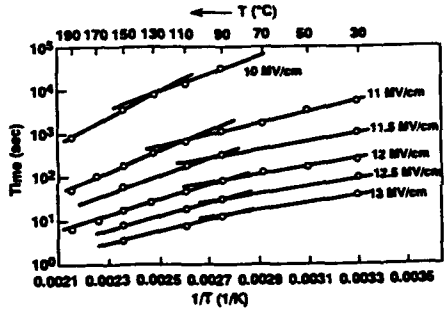


Fig. 6: The Arrhenius plot of t_{bd}-data shows an activation energy dependence on the electric field and the temperature range.

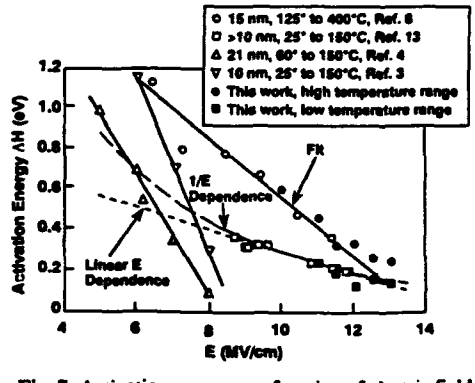


Fig. 7: Activation energy as a function of electric field.

electric fields is that a linear dependence is shown on the field rather than a dependence on the inverse field. Especially if the extrapolation over a wide field range is required, this could result in significantly different activation energies (compare in Fig. 7 dashed line to dotted line from equation 5). Consequently the more conservative approach of using equation 2 with a linear dependency of the activation energy on the electric field may be preferred rather than applying equation 3.

The slope β, as determined from the individual distributions at one electric field, is plotted versus the temperature in Fig. 8. No change in slope is observed up to 130°C (solid line), which means the oxide behaves like that expected from the Arrhenius model (K_T=0). But an increase of the slope is observed at higher temperatures; therefore, the upper limit set by the activation energy determination is confirmed by the slope's behavior. No variation of the slope with electric field was observed in Ref. 8; however, defect-related breakdown showed a change of the slope with temperature (Ref. 3-4).

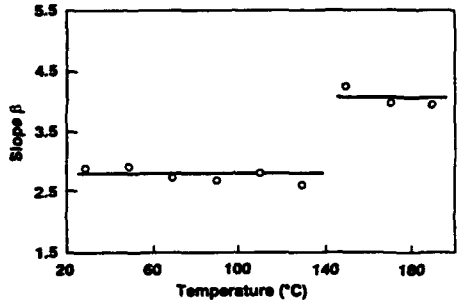


Fig. 8: Slope β of the Weibull distribution versus temperature. The slope changes significantly at high temperatures.

5. DISCUSSION

The presented results revealed an upper limit for each parameter required for the projection of accelerated stress results to operation conditions. Although the sample size for each condition was relatively small, the variation of the data was smaller than the deviation at certain high-stress conditions, e.g., the variations of the slope in Fig. 3 and 8. On the other hand, the results were in good agreement with available literature data, so it is realistic to establish upper limits for the stress conditions. The data suggest that there is no fixed upper limit. The limits change with stress conditions and also most likely depend on the oxide process. Therefore, the upper limits need to be determined or verified for almost each projection.

The behavior of the slope of breakdown distributions at the upper limits is not consistent. Available data show contradicting results, so that it can only be stated that the slope shows a significant disturbance at high-stress conditions compared to low-stress conditions. From the presented results, it is obvious that the slope is an additional indicator that is affected after changes at other parameters (C, ΔH) already occurred. Therefore, it is necessary to ensure that C and ΔH are determined in a range that includes projection target conditions. It is important to avoid mixing data with data beyond this range, to prevent over-optimistic projections.

For a conservative projection with a minimum number of measurements for model parameters and limits, one should stress at the target temperature or, if appropriate, at a worst-case temperature and deal only with field acceleration. For consideration of different temperatures, e.g., for a burn-in simulation, the activation energy dependence on the electric field is preferable, rather than dealing with the temperature dependence of C. The activation energy does not always show a field dependence (Ref. 8, 11); but, if it has such a dependency, a linear dependency on the field is more consistent with available data over a wide field range and leads to more conservative projections than an inverse field dependence as given in Ref. 2 or by equation 5.

The presented data do not show significant deviations of the slope from the model behavior in the relevant range. However, other available data (Ref. 3-4, 6-7) give evidence that such a deviation can occur. Therefore we recommend that the behavior of the slope is checked before extrapolations to smaller percentages are performed, so that deviating behaviors can be considered instead of forcing a certain behavior to the data. The variations of the breakdown behavior reported in the literature also suggest that there be some flexibility for relevant projections of dielectric breakdown.

6. CONCLUSIONS

The discussion of the projection from accelerated to operation conditions identified two possible extrapolation paths which conserve all degrees of freedom. This results in an impractically large number of parameters and their dependencies to be determined. The data also revealed upper temperature and voltage limits for all important parameters (field acceleration parameter C, activation energy ΔH and the slope of the breakdown distributions). A physical model for the transition point of the field acceleration relates this limit to the change from trap creation at low fields to band-gap ionization at higher fields; the latter mechanism is not relevant at operation conditions. The revealed limits are not fixed but depend on the stress conditions. The limits and acceleration parameters also depend on oxide thickness and technology. Available data suggest a wide variation of oxide breakdown

behavior rather than a unique behavior. Therefore, it is necessary to consider the behavior of the distribution slopes for projections and to determine or verify the parameters and limits for each process.

The number of measurements could be reduced if measurements were performed at the projection target temperature or, if appropriate, at worst-case temperature; only the field acceleration for one fixed temperature would then need to be determined.

To utilize a higher acceleration than that provided within the upper limits, the results from the dominant degradation mechanisms at that range would need to be transform to the one dominant at operation conditions. Until this is possible, projections including data from beyond the upper limits will lead to over-optimistic results.

6. REFERENCES

1. Chen I.C., Holland S. and. Hu C "A quantitative physical model for time-dependent breakdown in SiO_2", Proc. IRPS, 1985, pp. 24-31

2. Moazzami R., Lee J.C. and Hu C. "Temperature Acceleration of Time-Dependent Breakdown", IEEE ED-36, No. 11, Nov. 1989, pp. 2462-2465

3. McPherson J.W. and Baglee D.A. "Acceleration Factors for thin Gate Oxide Stressing", Proc. IRPS, 1985, pp. 1-5

4. Boyko K.C. and Gerlach D.L. "Time Dependent Dielectric Breakdown of 210 Å Oxides", Proc. IRPS, 1989, pp. 1-8

5. Fellinger J. and Vollertsen R.-P. "Consistency of Lifetime Extrapolation Models applied to ONO-Tripple Layers", Proc. ESREF, 1991, pp. 247-254

6. Vollertsen R.-P. "A New Approach of Statistical Modelling the Time Dependent Oxide Breakdown", Proc. ESREF, 1992, pp. 97-100

7. Nafria M., Suñé J. and Aymerich X. "Limitations of oxide breakdown accelerated testing for reliability simulation", Proc. ESREF, 1992, pp. 101-104

8. Suehle J.S., Chaparala P., Messick C., Miller W.M. and Boyko K.C. "Field and Temperature Acceleration of Time-Dependent Dielectric Breakdown in Intrinsic Thin SiO_2", Proc IRPS, 1994, pp.120-125

9. Chen I.C. and Hu C. "Accelerated Testing of Time-Dependent Breakdown of SiO_2", IEEE EDL-8, No. 4, April 1987, pp. 140-142

10. DiMaria D. J., Cartier E. and Arnold D. "Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon", J. Appl. Phys. 73, No. 7, April 1993, pp. 3367-3384

11. Monseric C., Mortini P., Ghibaudo G. and Panankakis G. "Breakdown Characteristics of Gate and Tunnel Oxides versus Field and Temperature", Quality and Reliability Engineering International, Vol. 9, No. 4, July-Aug. 1993, pp. 321-324

12. Chen C.-F., Wu C.-Y., Lee M.-K. and Chen C.-N. "The Dielectric Reliability of Intrinsic Thin SiO_2 Films Thermally Grown on a Heavily Doped Si Substrate - Characterization and Modeling", IEEE ED-34, No. 7, July 1987, pp. 1540-1551

13. Schuegraf K.F. and Hu C. "Effects of Temperature and Defects on Breakdown Lifetime of Thin SiO_2 at Very Low Voltages", Proc IRPS, 1994, pp.126-135

7. ACKNOWLEDGMENTS

The authors appreciate the discussions with D. J. DiMaria and E. Cartier from Yorktown and acknowledge the model calculations for Fig. 5. They thank R. DuFresne and D. Brochu for performing the electrical measurements.

NOISE INVESTIGATION OF METAL-GaAs OHMIC CONTACTS

Zs. Kincses, B. Kovács and I. Mojzes

Technical University of Budapest
Department of Electronic Technology
H-1521 Budapest, HUNGARY

ABSTRACT

The metal-semiconductor contacts are essential building blocks of any semiconductor devices. This work describes a method that is applicable to evaluate the noise spectra of ohmic contacts applying the widely used TLM (Transmission Line Model) test pattern. The results of the noise measurement are compared with results of the contact parameter measurement. The comparison shows that the contact noise has an 1/f component exceeding the thermal noise attributed to the contact resistance (R_C).

1. INTRODUCTION

The metal-semiconductor contacts are important part of semiconductor devices. Although a lot of effort was devoted to study the physics and the technology of metal-semiconductor junction, there are certain open questions of this field. One of these questions is the relation between the obtainable noise spectra and the state of degradation of the ohmic contacts. To study of this question an accurate method is required to separate the noise spectra of the ohmic contact from the noise generated in other parts of the device or the test pattern.

Several device models are widely used. These models generally do not care about the location of noise sources. Localisation of the noise source can be carried out by measuring the bias or temperature dependence of the noise. In contrast to that in the present we propose a new method based on comparison of noise spectra measured on different points of a TLM (see Ref. 1) which is in this respect a multi-contact test pattern. According to our new method proposed, the noise data measured on different parts of the TLM pattern will be evaluated by a subsequent numerical calculation, that is capable to separate the noise of the ohmic contact from the complete noise of the test pattern.

2. THE NEW CONTACT NOISE SEPARATION METHOD

During the measurement of the contact noise a low noise voltage source (as alkaline dry batteries) and a low noise preamplifier (fabricated by Ambrózy as described in Ref. 2) were applied. The noise spectrum was measured in frequency range of 1.6 Hz ... 20 kHz with real-time digital frequency analyzer type Brüel & Kjær 2131 controlled by PC 486's.

The theoretical background of our new method is based on the next reasoning. In a device several noise voltages are combined. Measuring the noise of the device we can measure the combination of these noise voltages. For the measuring of this voltage a simple test lay-out can be used, which is a plain homogeneous resistive layer on the semiconductor wafer with two ohmic contacts. Connecting of this device into the measuring circuit, the contacts are flowed through with current. Comparing to the IEC recommendation (Ref. 3), in the frame of this work a new modified measurement philosophy was used, where in the device under test (DUT) the driving and measuring points are separated.

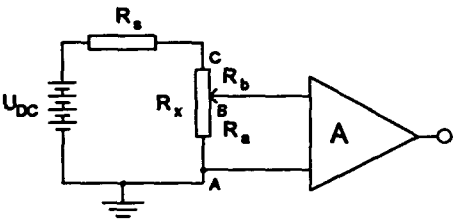


Fig. 1: Proposed measurement setup

Considering a currentless probe taken into the circuit, we shared the resistance of DUT described above. (See Fig. 1.) The biasing DC current is flowing through the test pattern from the extreme contact to the other extreme contact. The noise signal is measured between the first (extreme) contact and the currentless probe. According to the placement of this probe the obtained noise signals are different. Assuming the noise of the auxiliary parts of the measuring circuit to zero, the measured noise origins from the two contacts and the resistive layer is as follows in (Eq. 1):

where:

R_s is the separator serial resistance in the circuit, R_x is the resistance of the resistive layer, R_a is the resistance of the part of the resistive layer between A and B points, R_b is the resistance of the other part of the resistive layer (between B and C points), u_c^2 is the noise power of the contact flowed through by current, u_{r0}^2 is the noise power of the unit resistance of the resistive layer.

Curves calculated in the function of the ratio R_a/R_x are shown in Figs. 2 and 3. The parameter of the curves is

the noise level of u_c^2 or u_{r0}^2, when the other parameter is assumed to zero.

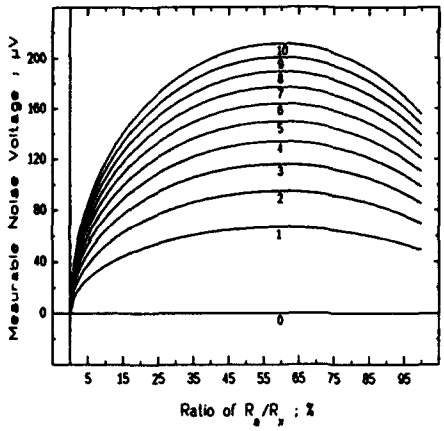


Fig. 2: resistor noise parameter is the resistor noise level: u_{r0}^2, index step is 1 μV^2 contact noise $u_c^2=0$

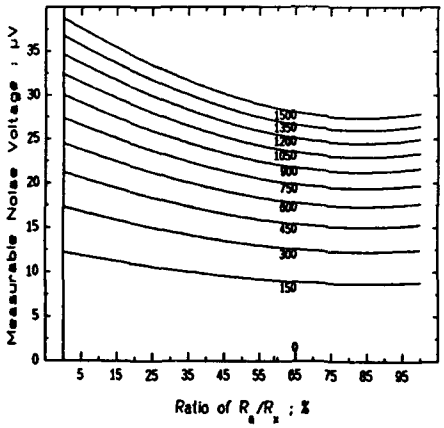


Fig. 3: contact noise parameter is the contact noise level u_c^2 index step is 150 $\mu V^2/k\Omega$ resistor noise $u_{r0}^2=0$

Using this equation the contact noise (u_c^2) and the noise of a unit resistance (u_{r0}^2) can be calculated by curve-fitting. This formula is very simply, it does not take any consideration for non-ideal properties of the test pattern.

3. OHMIC CONTACTS AND TLM

In practical aspects a potentiometer-like probe cannot be used because of its higher noise level. This contact must be deposited onto the surface as contacts are usually made. Transmission Line Model is a widely used multicontact test pattern (Fig. 4). There is a need of a new formula taking the impact of metal overlayer into consideration.

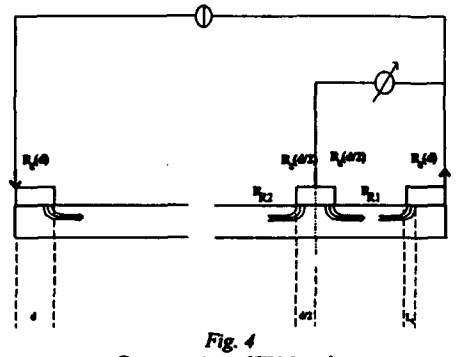


Fig. 4
Cross-section of TLM probe

In this case the biasing DC current is flowing through the TLM pattern from the extreme contact to the other extreme contact. The noise signal is measured between the first (extreme) and an other contact. Fig. 5 shows the equivalent circuit of the noise measurement on the TLM pattern.

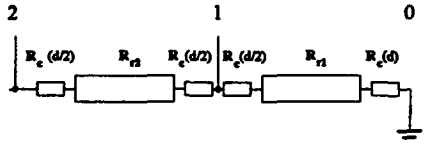


Fig. 5
The equivalent circuit of the TLM pattern

As Figs. 4 and 5 show, there are two different type of contacts: extreme and internal. In practice the internal contact cannot be currentless. When a metal overlayer is deposited onto the semiconductor surface the current flowing through the semiconductor resistive layer it goes in and out the metal. The internal contact can be understood as two serial contact with half of contact length. Contact resistances can be evaluated (Ref. 4).

In this point of view the equation for the given noise to the outlet is modified. Assume that R_a is the resistance between contact No. 0 and i, R_x is the resistance from the beginning to the end of TLM structure, u_{c1} is the noise voltage of each extreme contact, u_{c2} is the noise voltage of each internal

contact, u_{r0} is the noise voltage of a unit resistance of the resistive layer. The n is the number of internal contacts.
$R_b = R_s - R_u$, and R_s is the serial separator resistance in the measuring circuit. In the case of this model the following noise power can be measured:

Noise power originated from contacts is: (See Eq. 2)

Noise power originated from resistive layer is (since the noise of R_s resistance is $R_s u_{r0}^2$): (See Eq. 3)

Noise powers denoted by S_c and S_r are added to the output.

4. SAMPLE PREPARATION

The applied test pattern chip is described in (Ref. 5). S doped, n-type GaAs epitaxial layer grown by Effer-Nozaki type VPE method was used to prepare the samples. The thickness of the active layer is 3 μm, the free carrier concentration at room temperature is 1.5×10^{15} cm^{-3} and drift mobility of the active layer is 10000 cm^2/Vs at the same temperature. The multilayer ohmic contact metallization contains AuGe(eutectic)/Ni/Au layers with thickness 75 nm, 12.4 nm and 20 nm, respectively. This metal composition has a rather low optimal heat treatment temperature (Ref. 6). The applied chemical surface preparation before the evaporation was finished with rinsing in high purity (18 MΩcm) water (Ref. 7). Finishing the test pattern with the preparation of the Schottky gate of FATFET structures, the layer parameters were measured as it was described in (Ref. 5). The DC parameters of the ohmic contact structures were obtained using the conventional TLM method (Ref. 1). In this case the conventional TLM method is applicable due to the presence of very high epitaxial sheet resistance ($\rho_\square = 5$ kΩ) comparing to the metal sheet resistance ($\rho_M = 1$ Ω) (Ref. 4).

The evaluated specific contact resistance was $\rho_C = 1.8 \times 10^{-3}$ Ωcm^2. Although the real specific contact resistance is a little bit lower as it was proved (in Ref. 4) the obtained value is very high comparing our previous results obtained during MESFET-like device preparation (Ref. 7). However, in this case, the free carrier concentration of the applied epitaxial layer is lower with two ranges of magnitude than in the case of MESFET devices. This is why, the optimized contact preparation process results so poor contact. (See Braslau-theory in Ref. 8.)

5. RESULTS AND DISCUSSION

Since in our cases the transfer length (Ref. 4) is $L_t = 6$ μm and the contact length (see Fig. 4) is $d = 100$ μm, it can be accepted that $u_{c1} \approx u_{c2}$. (See Fig. 6.) Applying these assumptions the noise signal of the contact can be evaluated knowing the noise signals, measured on the different contact pairs as it was described above.

Fig. 7 shows the measured noise spectrums of a TLM pattern based on GaAs. Fig. 8 shows the evaluated noise spectra of the ohmic contact. Taking into account that R_C is 300 Ω in the investigated TLM pattern, it should be emphasized, that the noise spectrum of the ohmic contact has a significant 1/f component exceeding the thermal noise attributed to the contact resistance of the investigated contact.

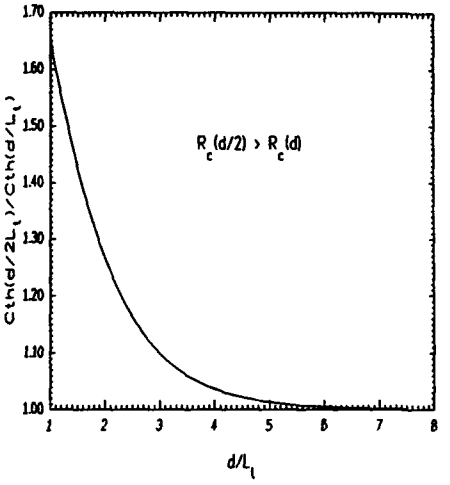


Fig. 6:
contact resistance ratio of extreme and internal in function of contact length comparing to transfer length

6. SUMMARY

In the frame of this work a new method was introduced for measuring and evaluation of the noise of metal–semiconductor contacts used in semiconductor devices. The measurement is based on a multi-contact test pattern allows to measure the distribution of noise in the function of the localization. A formula is presented which is capable to determine the noise parameters by curve fitting. Our results show that this method is useful in many cases we investigated. The spectrums of frequency distribution were found are nearly 1/f.

7. ACKNOWLEDGEMENTS

This work was supported by the Hungarian National Scientific Research Foundation through the Grant No. 1237 and 7557.

The sample, being measured in the course of this investigation, were prepared in the Research Institute for Technical Physics of the Hungarian Academy of Sciences.

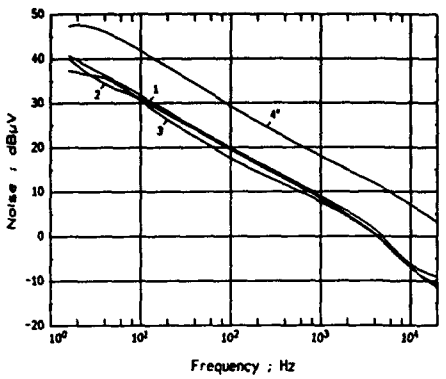


Fig. 7:
The noise spectrums of the TLM pattern

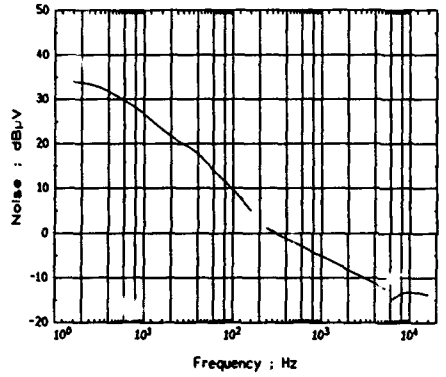


Fig. 8:
The separated noise spectra of the ohmic contact

8. REFERENCES

1. H.H. Berger: "Contact resistance and contact resistivity", J. Electrochem. Soc., 119, pp. 507-514, (1972)
2. A. Ambrózy: "Noise Measurement on Thick Film Resistors", Proc. of 8th Noise in Physical Systems and 1/f Noise, pp. 65-69, Rome (1985)
3. IEC, Publication 195: "Method of measurement of current noise generated in fixed resistors", Genève, Suisse, (1965)
4. B. Kovács, I. Mojzes: "Influence of finite metal overlayer resistance on the evaluation of contact resistivity", IEEE Transactions on Electronic Devices, ED-33(9), pp. 1401-1403, (1986)
5. B. Szentpáli, B. Kovács, F. Riesz, V.V. Tuyen: "Measurement techniques for the determination of doping and mobility profiles", Proc. of Hungarian–Greek Workshop on the Electronic Properties of Irradiation Induced Defects in III-V Semiconductor Structures, 9-14 Sept. 1993, Budapest, Hungary
6. B. Kovács, I. Mojzes, R. Veresegyházy, M. Németh-Sallay, S. Bíró, B. Pécz: "The influence of ohmic metal composition on the characteristics of ohmic contacts" Vacuum TAIP, 40(1/2), pp. 179-181, (1990)
7. B. Kovács, M. Németh-Sallay, I. Mojzes: "The effect of chemical treatment of the GaAs surface on the ohmic contact properties", Mat. Res. Soc. Symp. Proc., 181, pp. 549-552, (1990)
8. N. Braslau: "Alloyed Ohmic Contacts to GaAs", J. Vac. Sci. Technol., 19(3), pp. 803-807, (1981)

9. EQUATIONS

$$S = u_c^2\left[\frac{(R_s+R_b)^2+(R_s)^2}{(R_s+R_x)^2}\right] + u_{r0}^2\left[\frac{R_s(R_s+R_b)^2+R_b(R_s)^2}{(R_s+R_x)^2}\right] \quad (1)$$

$$S_{c_i} = \left[\left(\frac{R_s+R_b}{R_s+R_x}\right)^2\left[u_{c1}^2+(2i-1)u_{c2}^2\right]+\left(\frac{R_s}{R_s+R_x}\right)^2\left[u_{c1}^2+(1+2n-2i)u_{c2}^2\right]\right] \quad (2)$$

$$S_r = \left(\frac{R_s+R_b}{R_s+R_x}\right)^2 R_s u_{r0}^2 + \left(\frac{R_s}{R_s+R_x}\right)^2 R_b u_{r0}^2 \quad (3)$$

IDENTIFICATION OF FAULTS IN ANALOG CIRCUITS

Y. Maidon, P. Fouillat, Ch. Bouvet, S. Lesage

Laboratoire IXL-Microélectronique
CNRS/URA 846 - Université Bordeaux I
351 cours de la Libération 33405 TALENCE

Abstract : We proposed an original method, based on spectral analysis, for fault diagnosis on analog circuit. Detection and location of the fault are possible by the means of a dictionary in which are stored harmonic amplitudes connected with different values of each accused parameter of the transistors in the circuit. By interpolation or extrapolation, we compute the actual value of parameters.

1. COMPILATION OF A FAULT DICTIONARY [1]

The dictionary contains the faults, stimuli and responses of each component.

It is necessary to define the faults to be isolated in the test circuit as only those present in the dictionary can be identified. Should the dimension of the dictionary be exaggerated the methods of detection could be affected ; however the increasing reliability of the wafers will favour progressive size reduction.

The method used for compiling the dictionary depends on the type of analysis performed on the circuit under test : DC, transient, frequency, noise.

All construction methods proceed in the following way, as an example in DC analysis :

- description of the test circuit
- selection of faults
- choice of stimuli
- stimulation of the circuit
- introduction of faults
- evaluation of the efficiency of the stimuli to detect faults
- compilation of the dictionary

It is necessary to select an optimum number of stimuli/responses [2] in order to store the minimum amount of data whilst maintaining a sufficient degree of detection and localisation of faults. This method of fault research is limited in the case of circuits presenting rapid variations of gain in accordance with the frequency. During the test the circuit is excited by the same stimuli as those used in the compilation of the fault dictionary. The signatures are then compared. A criterion for the isolation of faults is implanted in order to detect in the test circuit a prestocked fault or an ambiguous group corresponding to a possible fault group.

In the field of transient analysis, circuit response is examined by pseudo-random excitation [6] or by a grade. The number of ambiguous groups recorded during the compilation of the fault dictionary is often quite high.

Different methods of frequency analysis can be envisaged using sinusoidal excitation as a starting point. The fault dictionary could, for example, be compiled in the following way :

a- choice of test frequencies.
b- measurements of gain and/or of the phase at these frequencies at different points in the circuit.
c- coding of the error or creation of a variation matrix of the response of the circuit at the nth fault and the kth frequency [3] [4].

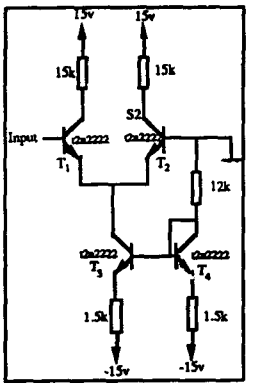
The number of ambiguous groups generated in this operation is high [5].

It is always necessary to select the test nodes. There exists a method which favours the minimisation of the number of nodes needing testing [2].

2. PROPOSED METHOD : SPECTRAL ANALYSIS OF THE OUTPUT OF AN ANALOG CIRCUIT

In order to avoid defining specific criteria for each circuit we suggest a method of fault research independent of the wiring tested, not requiring intermediate test points and easy to put into practice.
To do this, a sinusoidal or square signal is applied to the input of the circuit and a spectral analysis of the response of the global assembly is carried out when the output tested has reached its steady state. This study can be achieved out on weak signals or on strong signals, and at different frequencies.

The first approach to validate the method consists in carrying out FFTs on the following assembly.



iss	=	0.	xtf	=	1.	ns	=	1.00000
cjs	=	0.	vjs	=	0.50000	ptf	=	0.
mjs	=	0.	eg	=	1.10000	af	=	1.
itf	=	0.50000	vtf	=	1.00000	bf	=	betaf
br	=	3.16954	is	=	$1.8413e^{-14}$	vaf	=	65.38
var	=	11.26266	ikf	=	1.00000	ise	=	$3.3793e^{-15}$
ne	=	1.31911	ikr	=	1.00000	isc	=	$4.3948e^{-13}$
nc	=	1.14075	irb	=	$1.2552e^{-03}$	nf	=	1.00500
nr	=	1.00748	rbm	=	$1.0000e^{-02}$	rb	=	12.24
rc	=	0.31004	re	=	$3e^{-3}$	mje	=	0.26480
mjc	=	0.34896	vje	=	0.58792	vjc	=	0.47002
tf	=	tauf	tr	=	$305e^{-09}$	cje	=	$2.4147e^{-11}$
cjc	=	$1.1200e^{-11}$	fc	=	0.95000	xcjc	=	0.94836
subs	=	1						

Figure 1 : Circuit under test and SPICE Model of t2n2222 transistor

The spectral response of this perfectly balanced circuit is uneven (FFT simulation with the software package HSPICE). If a fault is introduced into one of the differential pair transistors, there follows a modification of the harmonic response. The following table gives an example of our technique when the current gain bf equal 10 (instead of 153 in the original model). Others kinds of degradations have been tested with the following parameters : vaf, re, rb.

s2	Reference	bf = 10	Vaf = 10 (V)	re = 1 (Ω)	rb = 120 (Ω)
H_0 (v)	7.11	7.38	8.46	7.04	7.07
H_1 (v)	689m	680m	564m	675m	677m
H_2(v)	16u	848u	3.6m	248u	132u
H_3(v)	381u	380u	677u	367u	373u
H_4(v)	18n	728n	3u	223n	92u
H_5(v)	212n	212n	30n	196n	203n

If we compare the results of columns bf and V_{af} with the reference column, we can clearly distinguish the default value (see table above).

3. VALIDATION OF THE METHOD

3.1 Sensitivity

In order to prove the sensitivity of fault diagnosis by harmonic analysis we compute factors such as :

$$S_{i,j,k} = \frac{\Delta H_i}{H_i} / \frac{\Delta x_{j,k}}{x_{j,k}} \qquad i \in [0,5] \quad k \in [1,4]$$

where : H_i = amplitude of the i^{eme} harmonic
$x_{j,k}$ = value of the j^{eme} parameter on the k^{ieme} transistor

So, for $x_{1,1}$ = bf of T_1 and $x_{2,1} = V_{af}$ de T_1, we find $S_{i,j,1} \simeq 3.10^{-3}$ for the worst case.

Simulation results show that harmonic and DC analysis give the same degree of sensitivity.

3.2 Fault location

Different methods allow fault detection and location [7] [8] [9] [10] but we operate on an other way.
Successively, one after one and one by time, all parameters of the transistors of the circuit are modified and respective values H_i, given by HSPICE simulation, are stored.
For a "one fault" circuit, dictionary exploitation permits to find only specified faults.
If parameter value is not present in the dictionary, interpolation or extrapolation permits to compute a closely value.
Least square method gives this value with a good precision. It supplies also the possibility of H_i identification with polynominal or exponential function.

4. CONCLUSION

This method, efficient for "one fault" circuits, must be modified in the case of multifaults circuit. We are currently working on new developments to address the identification and location of faulty points.

5. BIBLIOGRAPHY

[1] Salama and Bender,
Fault Diagnosis of Analog Circuits
Proc. IEEE, vol. 73, n° 8, pp 1279-1286, Aug. 1985.

[2] V.C.Prasad and S.N.R. Pinjala,
Boolean Method for Selection of Minimal Set of Test Nodes for Analog Fault Dictionary
Electr. Letters, vol. 29, pp. 747-749, Apr. 1993.

[3] S. Seshu and R. Waxman,
Fault Isolation in Conventional Linear Systems - A Feasibility study
IEEE Trans. Reliab.vol R-15, pp 11-16, 1966.

[4] A. Sriyanada and D.R. Towill,
Fault Diagnosis via Automatic Dynamic Testing : A voting Technique
The Automation Testing (Inst. Elec. Eng. Conf. Publ. n° 91), pp 196-201, 1972.

[5] K.C. Varghese, J.H. Williams, and D.R. Towill,
Simplified ATPG and Analog Fault Location via a Clustering and Separability Technique
IEEE Trans. Circuits Syst., vol CAS-26, pp 496-505, 1979.

[6] C.J. Macleod,
Parameter estimation using Pseudorandom Binary Sequences
Electr Letters, vol 5, pp 35-36, 1969.

[7] M. Slamani and B.Kaminska,
Analog Circuit Fault Diagnosis - Based on Sensitivity Computation and Functional Testing
IEEE Design and Test of Computers, vol 9-1, pp 30-39, 1992.

[8] Hong Dai and T.M. Souders,
Time Domain Testing Strategies and Fault Diagnosis for Analog Systems
IEEE Trans Instru. & Meas., vol 39, n°1, Feb 1990.

[9] G.N. Stenbakken and T.M. Souders,
Test-Point Selection and Testability Measures vi QR Factorization of Linear Models
IEEE Trans. Instru. and Meas., vol IM 36, n°2, June 1987.

[10] G.N. Stenbakken, T.M. Souders, G.W. Stewart,
Ambiguity Group and Testability
IEEE Trans. on Instru. and Measurements, vol 38, n°5, Oct 1989.

A COMPARISON OF STATIC AND DYNAMIC TECHNIQUES FOR THE DETECTION OF GATE OXIDE SHORTS IN DIGITAL MOS CIRCUITS

L Dong and V M Dwyer

Department of Electronic and Electrical Engineering
Loughborough University of Technology
Loughborough
LEICS, LE11 3TU, U.K.

Tel: +44 509 222856, Fax: +44 509 222854

ABSTRACT

Gate oxide shorts (GOS) represent a reliability threat in MOS based ICs and commonly used static tests for such shorts are not always effective. In this paper, a GOS was intentionally induced in an inverter circuit resulting in an extra time delay which is passed on to the output pin. Using a simple model for the defective oxide PSPICE simulations were performed and a good agreement was achieved between the experimental and simulation results. As a result, a simple dynamic test, measuring the signal delay through the circuit also may indicate the presence of a GOS.

1. INTRODUCTION

Gate oxide shorts in MOS transistor circuits have long been recognised as a serious reliability problem, especially for scaled-down devices (Ref.1). They can be caused either during manufacturing process or by electrostatic discharge (ESD) after packaged. A large amount of research has been carried out, both experimentally and theoretically.

GOS defects cannot be detected by electrical test programs based on stuck-at fault models since the circuits with GOS defects generally do not lose their function. However, electrical parametric changes do occur after GOS present in circuits.

GOS fault can be easily detected in CMOS circuits by measuring the static power current, however, this method does not work in nMOS circuits. In this paper, a simple dynamic test is proposed and its effectiveness is evaluated.

2. STATIC TEST FOR GOS

The most obvious characteristics for a MOS transistor with GOS is that there is gate current when the transistor is on. Since there is no static power dissipation in a GOS-free CMOS circuit (the leakage current is in the order of 10^{-11}A (Ref.2)), the presence of GOS gives rise to a conductive path between the power supplies as shown in Fig.1.

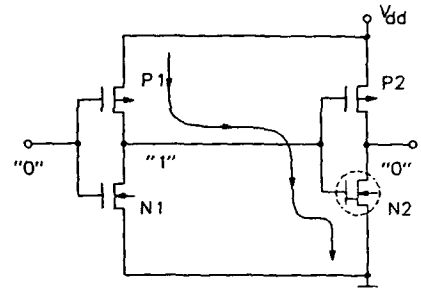


Figure 1: Schematic diagram of current drawn from power supply for a two-cascaded CMOS inverters with GOS defect in one of the transistors.

This current varies from 10^{-9}A to 10^{-3}A (Refs.1,2) and depends on the size and the extent of the oxide failure. Even the lightest GOS defect will increase the static power consumption up to 2 orders. Hence, static power consumption is a very effective way to detect GOS faults in CMOS inverter circuits.

It would not be so simple for a very complicated logic circuits. However, it can be solved by measuring the power supply during the logic test procedure.

However, problems arise for the detection of nMOS circuit with GOS since there is a static power dissipation even for a GOS-free circuit. In a two-cascaded nMOS inverter GOS-free circuit, for any of the static states, both of the pull-up transistors are on and one of the pull-down transistors is on. This

determines the static power current is the 'on' current of pull-up transistor. While the presence of GOS only adds a gate leakage current (in the order of μA) to the power current. Consequently, if any of the transistors contain a GOS, no noticeable difference is detected.

3. DYNAMIC TEST FOR GOS

3.1 Theory of circuit delay

In MOS circuits, the outputs at any internal node connect to the inputs of the next gate. The oxide between gate and channel in MOS transistors forms a capacitor C_i between the gate and substrate, such that the load at the internal node appears as C_l. This internal circuit can be modelled as a circuit shown in Fig.2.

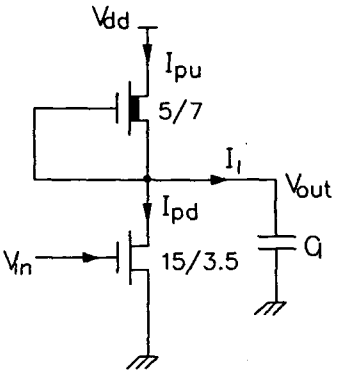


Figure 2: A nMOS inverter.

The charge and discharge times t_c and t_d can be determined by equation (1,2).

$$C_l(V_{dd} - V_{lo}) = \int_0^{t_c}(I_{pu} - I_{pd})dt \quad (1)$$

$$C_l(V_{dd} - V_{lo}) = \int_0^{t_d}(I_{pd} - I_{pu})dt \quad (2)$$

Normally the transistor current characteristics degrades upon creation of a GOS, and since the delay time of a gate depends on both the load capacitance and the charge or discharge current, a change in time delay in expected.

3.2 Test samples

A 4" wafer containing a nMOS inverter as shown in Fig.2 on each die was used throughout the experiments. The wafer was held upon a brass chuck by a vacuum pump, and four adjustable microprober were used to access the device bond-pads. The pull-up transistor is a depletion-mode transistor with a dimension of 5/7 (μm), while the pull-down transistor is a enhancement-mode transistor with a dimension of 15/3.5 (μm).

3.3 Experimental procedure and results

3.3.1 Characterisation of the pull-down transistors and nMOS inverters

Measurement of the I_G - V_G and I_{DS} - V_{DS} characteristics of the pull-down transistors using the HP4145B parametric analyzer: The gate current is only a noise in the order of nA magnitude and the drain current characteristics are shown in Fig.3.

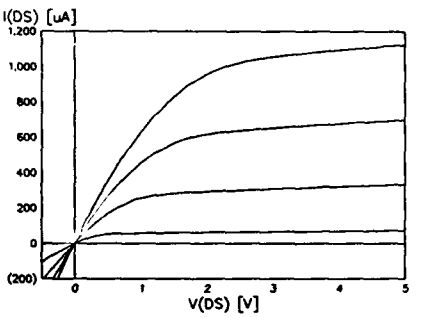


Figure 3: I_{DS} - V_{DS} characteristics of the pull-down transistor.

Measurement of the DC transfer curve of the inverter: Both the output voltage and the supply current were recorded, and are shown in Fig.4 and Fig.5 (solid lines).

Measurement of the transient characteristics of the inverter: A 300KHz square waveform from a THANSAR TG102 pulse generator was applied to the input of the inverter and a HP54111D storage oscilloscope was configured to capture the output signal. The result is shown in Fig.6 (solid line).

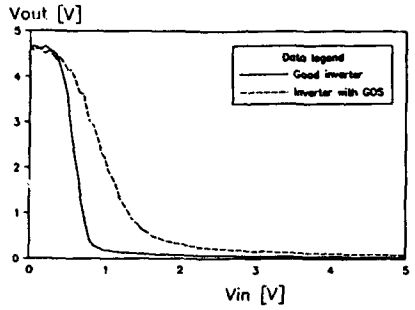


Figure 4: DC transfer curve of the nMOS inverter.

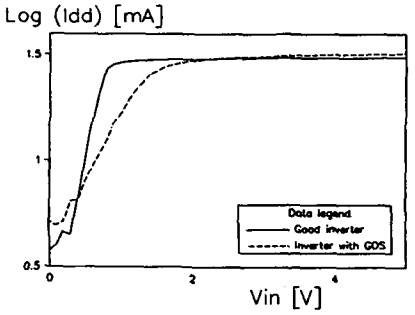


Figure 5: I_{DD} - V_{IN} characteristics of the nMOS inverter.

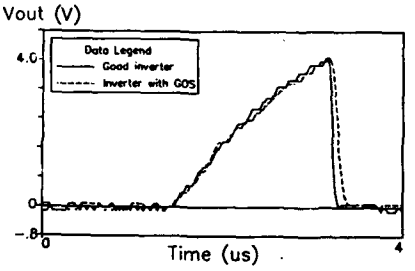


Figure 6: Transient response of the inverter to a 300KHz square waveform.

3.3.2 Oxide breakdown of the pull-down transistors

In order to introduce the gate oxide short, a constant voltage stress (-44V) from the HP4145B parametric analyzer was applied to the gate pad of the pull-down transistor. The stress was removed immediately after the oxide breakdown.

3.3.3 Re-characterisation of the pull-down transistors and the nMOS inverters

The oxide breakdown was confirmed by the gate current characteristics shown in Fig.7.

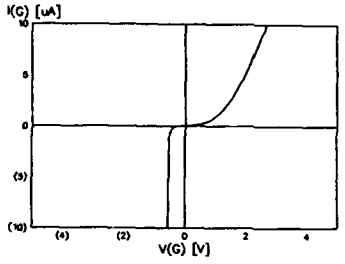


Figure 7: I_G - V_G characteristics of the pull-down transistor after the gate oxide breakdown.

After the measurement of gate current, all the other characteristics of the pull-down transistors and nMOS inverters were re-measured and are shown in Fig.4, Fig.5 and Fig.6 with dashed lines.

3.3.4 Experimental conclusion

We still consider the two-cascaded nMOS inverters circuit: in the case of GOS-free circuit, the static power current is 32.43+4.26 = 36.69 (μA). While with one of the pull-down transistor with GOS, the static power current is 32.1+5.197 = 37.297 (μA). Thus, no conclusion can be made on the static current results. Although there is a noticeable difference in the output 'low' logic level, it will be lost through the next gate. However, from the transient response of the square waveform, the difference in the discharge time is clearly shown in Fig.6, also this will not disappear through successive gates and will ultimately reach the external output.

4. SIMULATION RESULTS FROM PSPICE

Syrzycki's model (Ref.3) as shown in Fig.8 for

MOSFET with GOS was used in the simulations.

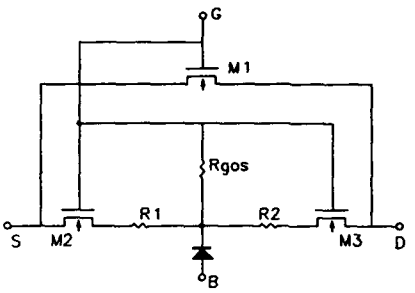


Figure 8: Syrzycki's model for nMOSFET with GOS.

The parameters of the transistors and resistors in Fig.8 were chosen to fit the experimental results of the gate current and drain current characteristics of the pull-down transistor after the inducing of GOS (Fig.5 & 6). Then the simulations for the inverter were performed based on the same parameters. The simulation results are shown in Fig.9-11 and in which the solid lines represent the case where all transistors are good devices, while the broken lines represent the case where the pull-down transistor has a GOS-fault.

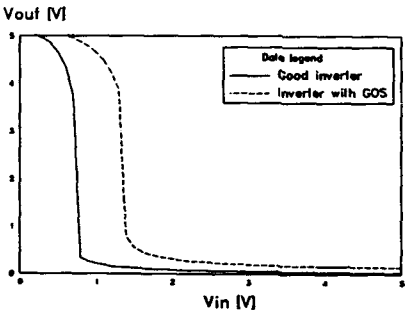


Figure 9: DC transfer curve of the nMOS inverter from PSPICE simulation.

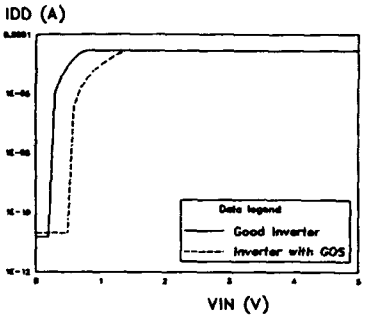


Figure 10: I_{DD} - V_{IN} curve of the nMOS inverter from PSPICE simulation.

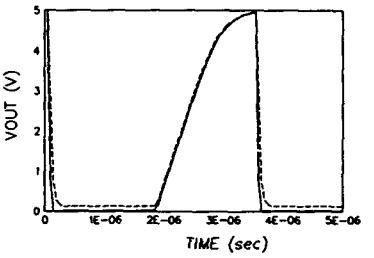


Figure 11: Transient response from simulation.

The simulation results agree very well with the experimental ones. This indicates that Syrzycki's model can not only model the DC effect of GOS in circuits, but also the transient parametric drift of the circuits.

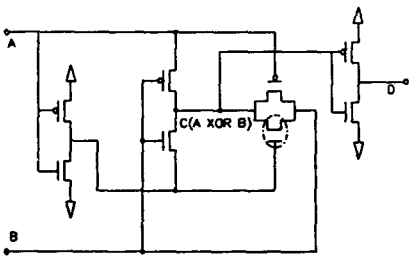


Figure 12: Transmission gate intensive NXOR logic gate.

Following the successful results above, simulations for a more complicated logic circuit as shown in Fig.12 were performed. The GOS-fault transistor is circled with broken line. The results are shown in Fig.13 and Fig.14 and the solid lines represent GOS-free circuit and broken lines represent the circled transistor is with GOS.

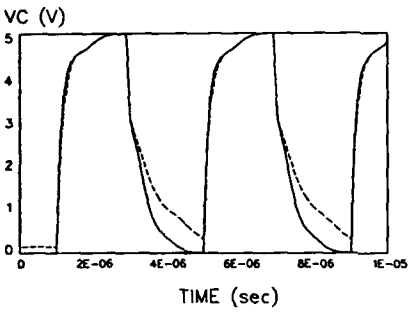


Figure 13: Output waveform at node C where the logic output is A XOR B.

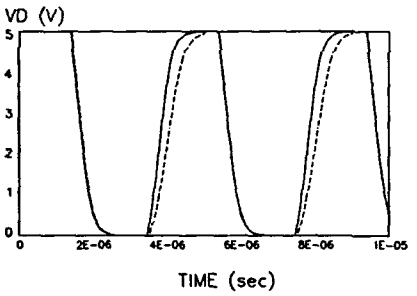


Figure 14: Output waveform at node D.

The simulation results clearly show that although the difference of output 'low' level at node C has lost at node D, the time delay passes the next gate and will reach the primary output at last.

5. CONCLUSIONS

Whilst the measurement of static current from power supply is a simple method of testing MOS circuits for gate oxide shorts, it does not work for all cases. However, a GOS also introduces an extra delay in signal transmission from input to output and consequently the delay test in those cases could be an effective alternative.

6. REFERENCES

1. *J.M.Soden and C.F.Hawkins*, "*Test* Considerations for Gate Oxide Shorts in CMOS ICs," IEEE Design and Test, August, 1986, pp.56-64.

2. W.D.Greason, Z.Kucerovsky and K.W.K.Chum, "Latent Effects Due to ESD in CMOS Integrated Circuits: Review and Experiments," IEEE Transaction on Industry Applications, No.29, 1993, pp.88-97.

3. M.Syrzycki, "Modelling of Gate Oxide Shorts in MOS Transistors," IEEE Transaction on Computer-Aided Design, No.8, 1989, pp.193-202.

ESTIMATION OF SYSTEM RELIABILITY USING A NON-CONSTANT FAILURE RATE MODEL

J.A.Hayes, J.A.Jones and M.Zahid

International Electronics Reliability Institute
Department of Electronic and Electrical Engineering
University of Technology, Loughborough
Leicestershire, LE11 3TU, United Kingdom

Tel: 0509-222849 Fax: 0509-222854

1. ABSTRACT

One of the most controversial techniques used at present in the field of reliability is the use of reliability prediction methodologies, based on component constant failure rate data for the estimation of system failure rates. This paper investigates a new reliability estimation method that does not depend upon constant failure rates.

A number of boards were selected from the International Electronics Reliability Institute's (IERI) field reliability database, and their reliability was estimated using a failure intensity based methodology, and was compared with the actual failure intensity observed in the field.

The predicted failure intensity is seen to be in close agreement with the observed value for the majority of a system's operating lifetime. This means that the failure intensity method should be investigated further to discover if it can be used to estimate the reliability of a system throughout its lifetime. Hence this may provide a more realistic picture of the way in which electronic systems behave in the field.

2. INTRODUCTION

IERI have been collecting a large amount of failure information over many years from leading British and Danish electronic manufacturing companies (Ref.1). The data stored in this database is of such high quality that IERI are able to perform reliability analysis and estimate the expected field behaviour and compare that with the observed performance.

The concept of electronic failure prediction methodology (EFPM) often affects major decisions in system design. EFPM is based on the assumption that systems fail as a result of failures of component parts, and those parts fail partly as a result of exposure to application stress (Ref.2). This means that by some consideration of the structure of such a piece of equipment and by further consideration of its usage, it is possible to obtain an estimate of the systems reliability in that particular application.

There are two traditional approaches to EFPM involving differing amounts of information about the system. The first is known as parts count analysis and requires comparatively little information. This takes the parts list for a particular design and bases the reliability estimate on the number of components used in it without any reference to operating conditions of these components. This method is generally used early in the design phase to obtain a simple estimate of the system reliability. The second method is known as parts stress analysis and involves knowledge of a wealth of information about the system, but is assumed to provide a more realistic estimate of the reliability. This second method tends to be used towards the end of the design cycle when actual circuit parameters have been established (Ref.3).

Reliability prediction methods are widely accepted throughout the electronics industry and this enables it to be used as a general yardstick which allows comparison between different equipments to be made. However, many manufacturers have commented that the models can be wildly inaccurate when compared with the performance in the field, particularly in the case of modern microelectronic devices, and their use can lead to increased costs and complexity (Ref.2).

Recent studies (Ref.4) have shown that the standard models for reliability prediction do not perform well in all circumstances, and reliability prediction performed using these models could lead to a misrepresentation of an equipment's reliability.

It has been demonstrated (Ref.5) that the reliability of components is a time dependent parameter and therefore traditional EFPM is unable to give an accurate picture of a system's reliability as it is based upon a constant failure rate assumption.

This paper investigates an alternative method of performing reliability assessment of equipments. The result is estimated as a function of time based upon the failure intensities of constituent components.

3. FAILURE INTENSITY ANALYSIS

Failure intensity is defined in equation (1)

$$FI(t) = \frac{n}{N\Delta t} \qquad (1)$$

where n is the observed number of failures in a given time period (Δt) and N is the population at risk during this period. For the purpose of this investigation the Δt interval has been chosen to be 1,000 hours.

To illustrate further the means of computation of the failure intensity Figure 1 and Figure 2 show the number of failures of MOS digital integrated circuits with between 10^3 and 10^4 gates in 1,000 hour intervals, and the corresponding number of components at risk throughout the same time intervals respectively.

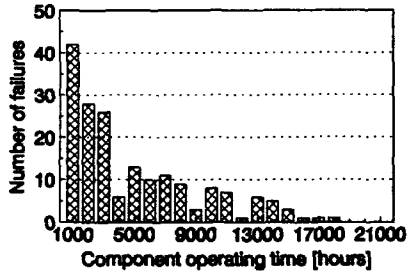


Figure 1:Number of failures of MOS digital ICs with between 10^3 to 10^4 gates versus time

The shape of Figure 2 reflects the fact that as the number of systems going into the field increases with time, the populations of the various component types is also increasing.

Figure 3 shows the failure intensity curve for CMOS digital devices with between 10^3 and 10^4 gates formed by using the information shown in Figure 1 and Figure 2. The failure intensity is shown as the central solid line, and the dotted lines above and below are the 95% χ^2 confidence limits. These limits are dependent on the number of observed failures and the population at risk. Wide limits are indicative of a lack of failure data. This situation will improve with time as more data on the various component types becomes available. It should be noted that when the failure intensity drops to zero on the time axis, this means no failures have been observed in that particular 1,000 hour period. Initially, the failure intensity value is high but it decreases to a lower level over a 4,000 hour period. The value seems to oscillate around the 5×10^{-7} failure intensity value thereafter.

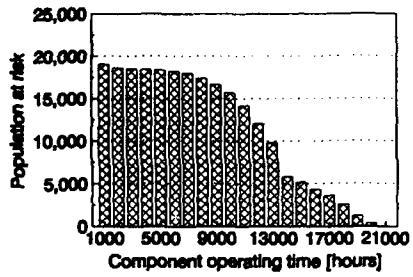


Figure 2:Number of MOS digital ICs with between 10^3 and 10^4 gates at risk in the field.

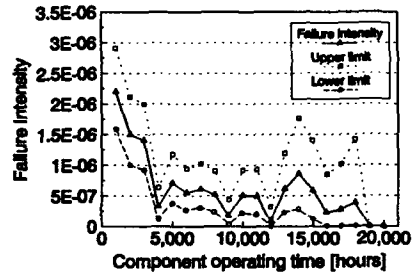


Figure 3:Failure intensity curve for digital MOS ICs with between 10^3 and 10^4 gates

The shape of the graph may imply that there are early component failures occurring in the field. These early failures may be due to members of the so called 'weak population' that have not been removed by any component or system screening performed by the component or system manufacturer.

4. RELIABILITY ESTIMATION

An electronic system can be considered to be a network of components all interconnected to one another in various complex ways. This real life model is unsuitable for reliability analysis since it is far too complex. In order to study the reliability of systems a number of assumptions need to be made.

i) Any component failure causes a system failure.

This is normally termed a series configuration (or chain structure). This is the simplest of the many models available and so is the most widely used for reliability modelling of systems.

A series configuration of n items will have a reliability function defined by (2)

$$R(t)=P(x_1,x_2,\ldots,x_n) \\ =P(x_1)P(x_2|x_1)\ldots P(x_n|x_1x_2\ldots x_{n-1}) \quad (2)$$

ii) The components that make up the system must be independent, this means that a failure by a single component must not affect other components in the system.

If the n items $x_1, x_2, \ldots, x_n$ are independent, then

$$R(t)=P(x_1)P(x_2)\ldots P(x_n) \\ =\prod_{i=1}^{n} P(x_i) \quad (3)$$

Assuming that no component failures affect any others, then the reliability of a system can be calculated by multiplying together the probability of failure for each component in the system. However, in a general system this can be difficult since each component's probability of failure could be a complex function. If, however, simple functions are used, then it is possible to proceed further.

iii) The component failure behaviour must be governed by a constant-hazard model

This last assumptions means that if the component reliability model is exponential in form, as in equation (4),

$$P(t) = e^{-\lambda t} \quad (4)$$

then equation (3) can be rewritten as equation (5)

$$R(t)=\prod_{i=1}^{n} e^{-\lambda_i t} \\ =e^{(-\sum_{i=1}^{n}\lambda_i t)} \quad (5)$$

Equation (5) is the most commonly used and the most elementary system reliability formula. However, it has the fundamental problem that it assumes a constant failure rate throughout the life of the components. If the instantaneous failure rate or failure intensity at any time is considered to be constant over a short interval, Δt, then equation (5) can be adapted to use failure intensity instead of failure rate and hence the requirements for constant failure rates can be removed.

Equation (6) gives the instantaneous failure rate for a system based upon the failure intensities of the constituent components.

$$I(t)=\prod_{i=1}^{n} e^{-I_i(t)} \\ =e^{(-\sum_{i=1}^{n} I_i(t))} \quad (6)$$

where I(t) is the failure intensity of the system in the interval Δt, and $I_i(t)$ is the failure intensity of the i'th device that comprises the system in the same interval. It then proves possible to calculate the failure intensity curve for a system when the failure intensities for all the components contained in that system are known.

5. FAILURE INTENSITY PREDICTION

Two board types were selected from the IERI database where their field performance was well known and estimation of their behaviour was made using equation (6).

Figure 4 shows the predicted and field failure intensity curves for the first board type. The solid line is the failure intensity observed in the field and the error bars give 95% χ^2 confidence limits. The dotted line represents the failure intensity predicted for this board type using equation (6).

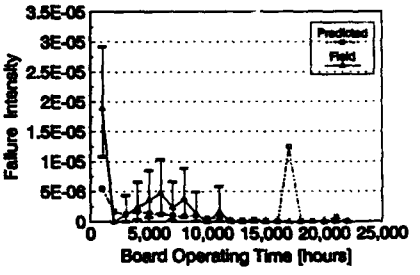


Figure 4: Predicted and field failure intensity curves for the first board type

The figure shows that the first board type

experiences early life failures up to 11,000 hours after which its operation becomes failure free. The predicted behaviour of this board type shows a decreasing failure intensity until about 10,000 hours, after which a failure free life is predicted except for an irregularity at 17,000 hours. This is due to rogue failure occurrences in the raw data. These rogue failures come from one data source and are not representative of the remainder of the stored data. It is significant to observe that the predicted line lies within the 95% confidence limits of the field behaviour throughout the boards operating life and hence would provide a fairly good estimation of board's behaviour in the field.

Figure 5 shows the predicted and field failure intensity curves for the second board type. In this case the field behaviour is more irregular, with failures occurring up to 20,000 hours. The predicted line gives a very good fit after 7,000 hours but before that it is pessimistic, probably due to early life failures observed in one or more data sources.

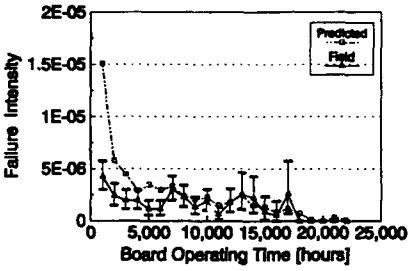


Figure 5: Predicted and field failure intensity curves for the second board type

6. CONCLUSIONS

The failure intensity methodology lends itself very easily to system reliability prediction. It appears to give a more realistic estimate of the reliability of a system throughout the operating lifetime of the equipment and does not make assumptions, such as constant failure rate, which can be detrimental to the validity of the estimate.

The predictions seem, on present evidence, to track the observed behaviour well, given the uncertainties that are evident in the field. Fine tuning of the source data is however necessary to smooth out irregularities that are due to few sources and are not representative of the majority of the data.

7. REFERENCES

1. D.S. Campbell and J.A. Hayes, "The organization of a study of the field failure of electronic components", Quality and Reliability Engineering International, Vol. 3, 1987, pp 251-258.

2. C.T.Leonard and M.Pecht, "How failure prediction methodology affects electronic equipment design", Quality and Reliability Engineering International, Vol. 6, 1990, pp 243-249.

3. Michael Pecht and Wen-Cheng Kang, "A critique of MIL-Hdbk-217E reliability prediction methods", IEEE Transactions on Reliability, Vol. 37, No. 5, 1988 December.

4. M.Zahid, J.A.Jones, and J.A.Hayes, "Evaluation of reliability prediction methodologies", Fourth European Symposium on Reliability of Electron Devices, Bordeaux, October 1993, pp 59-64.

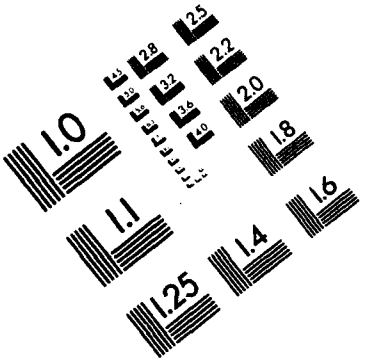
5. D. S. Campbell, J. A. Hayes, J. A. Jones and A. P. Schwarzenberger, "Reliability behaviour of electronic components as a function of time", Quality and Reliability Engineering International Vol. 8, 1992, pp 161-166.

AD-A286 614 EUROPEAN SYMPOSIUM ON RELIABILITY OF ELECTRON DEVICES FAILURE PHYSICS AND..(U) GAREL ASSOCIATES NORTHAMPTON (UNITED KINGDOM) C M BRYDON 7 OCT 94 XB-ONREUR 7/7

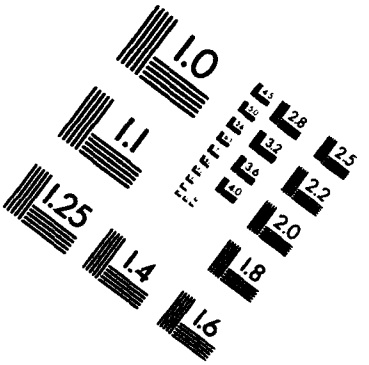
UNCLASSIFIED NL

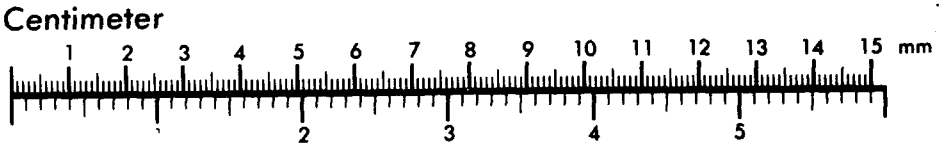
[redacted]

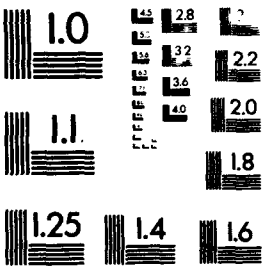
END
FILMED
DTIC

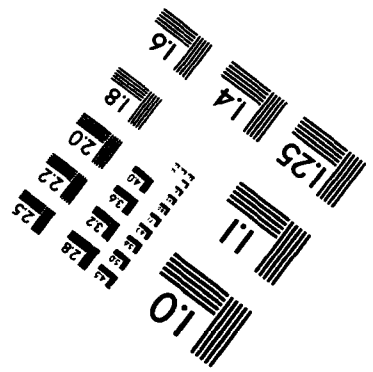


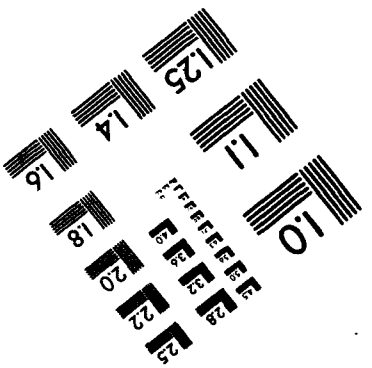
Association for Information and Image Management
1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202











MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

BIST & DIAGNOSIS IN SAFETY CRITICAL MICROSYSTEMS USING RELIABILITY INDICATORS

Th Olbrich, A M D Richardson, and D A Bradley
Dept. of Engineering, Lancaster University, Lancaster LA1 4YR, U.K.,
FAX: (++44) 524 594207, email: T.Olbrich@Lancaster.ac.uk

Abstract
Sensors and actuators with built-in local intelligence are often described as microsystems. The incorporation of processing electronics at the sensor and actuator level enables the distribution of processing tasks such as calibration and filtering as well as test and diagnostic functions from upper system hierarchies to lower levels. This paper describes *Built-In-Self-Test (BIST)* and diagnostic strategies for *Safety Critical Microsystems*. It compares different approaches and shows the importance of utilising *Reliability Indicators (RIs)* for on-chip monitoring and diagnosis. The close relationship between *Design for Testability (DfT)* for post-production tests and the BIST strategies for on-line monitoring is outlined. A multichip design strategy is described for an example microsystem.

1. INTRODUCTION

The escalating complexity of engineering systems and reliability requirements from consumers and industry are generating increasing demands for microsystems incorporating diagnostic and self-test features especially for applications which are classified as safety critical.
Microsystems are self-contained, small, stand-alone devices incorporating sensor(s) and/or actuator(s) plus electronics for signal conditioning, processing, control and communication. This paper investigates the need to also integrate Built-in-Self-Test (BIST) into microsystems.
The concept of microsystems is reviewed in section 2 with a smart sensor example and the importance of BIST and diagnosis for microsystems is outlined in section 3. Strategies suitable for on-chip diagnosis and BIST are discussed in section 4. An approach to the output of the diagnostic information and a microsystem design schematic explain how self-test can be incorporated into microsystems and its diagnostic information further processed. Conclusions and a section on future work end the paper.

2. SMART SENSORS - A MECHATRONIC MICROSYSTEM

A smart sensor comprises one (or more) sensing elements plus some associated electronics which enables delegation of processing tasks from the main controller to the sensor level and is used here as an example for a complex single-device mechatronic microsystem. Measurement of input measurands also includes signal processing (i.e. filtering), correction of offset drifts and compensation of quantity influences such as cross sensitivity to temperature. In addition, data validation functions may produce additional measurement information such as the accuracy of the measurement or the "health"-status of the device (Ref.1). A flexible configuration of a smart sensor allows adaptation to different measurement tasks. A bi-directional digital communication interface allows the microsystem to communicate directly with other systems.
Single chip smart sensors are readily available (Ref.2) but technological problems such as process incompatibility between the sensing element and the circuit technology, low yield and packaging problems, have so far prevented the anticipated high volume breakthrough. Hybrid realisations of smart sensors (Ref.3) using one or more additional integrated circuits are feasible for many applications and an implementation is proposed in section 6.

Table 1: Advantages of integrated microsystems over distributed implementations.

- Easier configuration of systems through digital interface.
- Reduced systems' installation effort through distributed processing tasks.
- Increased performance through built-in processing capabilities.
- Increased application flexibility through on-line reconfiguration.
- Higher reliability and testability through internal self-test.
- Increased process reliability through evaluated output information.
- Miniaturisation allows new applications.

Some important advantages of microsystems over their non-intelligent counterparts are listed in Table 1 and explain their increasing demand. Microsystems such as smart sensors find their applications within the process, automotive and aerospace industries and in medical applications (Ref.4). A high volume break-through is expected during the next few years which will reduce costs

dramatically so that microsystems will find their way into everyday consumer applications (Refs.5,6).

3. BIST & DIAGNOSIS DEMANDS FOR MICROSYSTEMS

Three important reasons are identified which necessitate the incorporation of BIST and diagnosis into microsystems.

3.1 Increasing the system reliability with BIST

Applications and systems are defined as safety critical if malfunction or failures can result in risk and damage to humans or the environment. Automotive (Ref.7), aerospace, medical and nuclear energy applications are safety critical and therefore need high reliability components, systems and processes. It is essential to ensure the correct function of the safety critical system during normal operation. This means *self-test* and diagnosis (preferably on-line) needs to be performed on all function blocks, including the diagnostic module itself (Figure 1). In addition, further actions leading to reconfiguration or fail-safe (graceful degradation) modes may be required following fault detection and localisation.

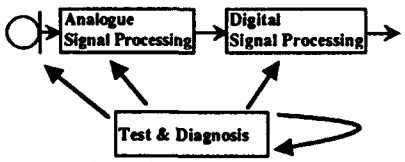


Figure 1: Smart Sensor as an example for a microsystem with Test and Diagnosis functions.

There is an increasing need within microsystems for tests to measure ageing and degradation to ensure detection of evolving or developing faults. Such early warning capabilities increase the system reliability during its planned mission time and the systems' maintenance intervals. The use of Reliability Indicators (RIs) therefore becomes increasingly important for on-chip implementation.

3.2 Increasing testability through DfT and BIST

The encapsulation of several microelectronic components plus sensing elements and actuators into one device and the use of multi-chip modules as well as feature sizes in the submicron range make it virtually impossible to access embedded function blocks directly. Therefore microsystem design needs to focus on DfT, which includes BIST and on-chip RI tests to enable exhaustive post production tests. It is not sufficient to test the various function modules, such as the sensing element and the microelectronic components on their own. Only the successful interaction of all function blocks will generate test results of acceptable credibility.

In addition, the influence of packaging effects on the microsystem performance, i.e. stress damage due to packaging or the effects of the "physical window" between the sensing element and the outside world on sensor readings prevent a meaningful system function evaluation of unpackaged microsystem components.

DfT and BIST are therefore essential for exhaustive testing of microsystems.

3.3 Reduced testing time through BIST

Testing time for post-production tests is a major issue, both technologically and financially, especially for complex application specific mixed mode integrated circuits. BIST and DfT features may help to reduce testing time drastically. This gives market advantages through decreased device costs.

BIST and DfT are essential for competitive, high reliable microsystems.

4. STRATEGIES FOR SELF-TEST AND DIAGNOSIS IN MICROSYSTEMS

This section describes different approaches to BIST and diagnosis in microsystems which are considered to be implementable on ASICs and microcontrollers for microsystems.

Figure 2 gives an overview of different diagnostic approaches that have been implemented in microsystems or are currently under development. This table is not complete and it is assumed that the number of entries will increase during the next few years. Many testing methods used for mixed signal integrated circuits (ICs) (Ref.8) can directly be applied to mechatronic microsystems.

4.1 Reliability Indicators

Richardson (Ref.9) lists a number of reliability indicator measurements including IDDQ tests, transient current tests, performance tests, offset drift measurement, low voltage tests, noise tests and phase jitter tests. Not all of these measurements are applicable to on-chip diagnostics because of the limitations in terms of chip-area and processing complexity.

Extensive research is being carried out into supply current measurement (*IDDx testing*) for digital and mixed-mode circuits (Ref.10) and promising results in the detection of catastrophic as well as parametric faults have been presented. However, the technique still has problems when applied to analogue circuits (Ref.11). Although the QTAG initiative for off-chip IDDQ/ISSQ monitoring (Ref.12) aims at post-production tests, it may well

be feasible to integrate these monitors into (multichip) microsystems as an (on-line) BIST facility.

The frequency response and transient response of mechatronic systems has been used to monitor the health of the system (Ref.13). Information in the frequency range above the one used for normal operation (i.e. noise (Ref.14)) may contain important information which is often discarded through filtering in the signal path. Limitations imposed by the processing power available on-chip may restrict such tests to comparatively simple evaluation calculations mainly during the power-up or off-line states.

Current sensing for open and/or short circuits in measurement bridges of silicon sensors account for the majority of defects in such devices. Current sensing in the bridge branches or voltage measurement at the bridge output nodes against limits or thresholds indicate instantaneously catastrophic faults in such devices. Limit checking and threshold checking are simple ways to compare the system reading against allowable boundaries.

Temperature measurements on microelectronic components or in sensing elements give information on the achievable accuracy or the reliability of a measurement. A measurement bridge in silicon gives a temperature dependant voltage if driven with a current source (Ref.15). This technique of temperature measurement is virtually free of additional components and indicates directly the temperature in the bridge resistors.

Built-in overstress fuses for example indicate whether a maximum allowable temperature had been exceeded. A means to indicate a damaging physical shock to accelerometers is also very desirable. Overstress on silicon components does not necessarily cause an immediate failure but often leads to a shorter life time, increased sensitivity to noise or other influences and may also cause unwanted failures later in the field.

4.2 Redundancy

Time redundancy is a helpful strategy to prevent intermittent measurement errors due to electromagnetic or electrostatic interference. It can for example be performed in this way:

1. Measure value.
2. Exchange input with reference element, measure and compare with expected result.
3. Measure input value again and compare with first measurement.

Hardware redundancy is widely used in safety critical applications in the aerospace and nuclear power industries where the cost factor is not as highly weighted as for example in consumer applications.

Information redundancy is widespread in software applications and communications protocols. Error correction codes in digital hardware implementations have been shown to be an interesting alternative where hardware redundancy in the form of triplicated systems etc. were not feasible but a reliable form of concurrent error checking, including correction of single intermittent faults, was required (Ref.16).

4.3 Hardware BIST

Built-in-Self-Test (BIST) is in general a design technique in which parts of the circuit (the system) are used to test parts of the system itself (Ref.17). Pseudo-Random Binary-Sequence (PRBS) test pattern generators for example apply input vectors

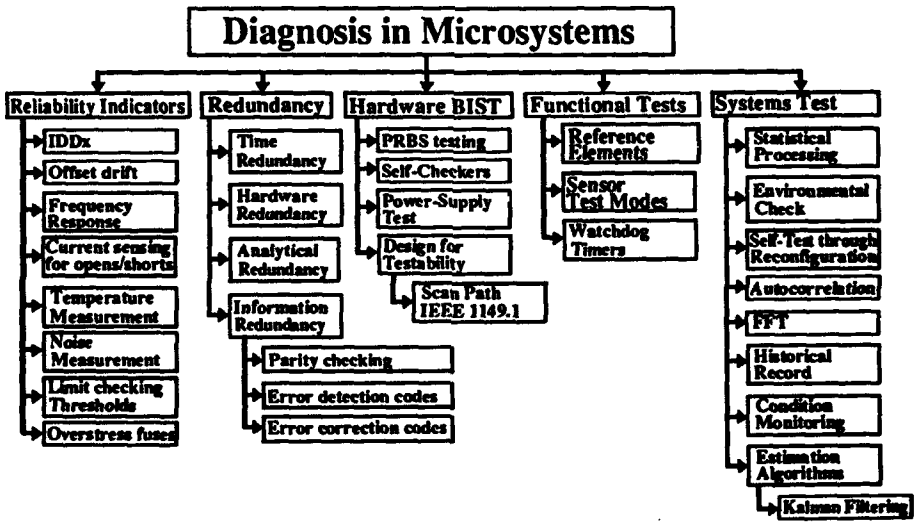


Figure 2: Methods for Test & Diagnosis in Microsystems

to digital or analogue (Ref.18) modules. Self-Checking circuit designs provide on-line test for digital (Ref.19) as well as for analogue circuits (Ref.20). The obtained output response is then compared with a correct one. Design for Testability (DfT) is one strategy to reduce the effort to test an integrated circuit and to allow faster and more exhaustive testing. The boundary scan standard IEEE 1149.1 allows the application of test pattern (or control and set-up information) to modules deeply embedded into a circuit board or integrated circuit (Ref.21). This standard is available in most ASIC design libraries and in most new microprocessors and may be useful for communication between ICs in microsystems.

4.4 Functional Tests

Abramovici et al suggested that "The objective of functional testing is to validate the correct operation of a system with respect to its functional specifications" (Ref.22). Defined excitation of sensing elements or actuators using for example reference elements attempts to exercise the functions of the system as closely as possible (and necessary) to its usual application. Self-test modes for piezoresistive accelerometers (Refs.23,24) in which a test pulse simulates a quasi-acceleration by deflecting the seismic mass derive a defined, near full-scale output signal that can be used to test the function of the sensing element and the signal path. Although these test approaches do not reflect on the detailed structural model of the system under test and are therefore not exhaustive, they deliver an accountable measure of the system function with minimal circuit overhead for providing the test function.

4.5 Systems Test

Various methods which deal with the evaluation of data groups can be exploited. Historical Records including actual and recent measurement data are evaluated using Statistical Processing methods. Environmental Checks may be performed using input variables which are not directly related to the primarily device function but are of use to further enhance the measurement characteristics. Additional information on the process and the environment in which the microsystem functions may be input via the Fieldbus interface from other systems. Most of these functions are preferably implemented in software and require a powerful embedded processor.

Condition monitoring (Ref.13) or estimation algorithms (Ref.25) to provide analytical redundancy for fault detection and isolation (FDI) are considered to be implementable in microsystems in more simplistic forms as for large machines or plants. Analytical Redundancy (Ref.26) uses mathematical models of the process and the device (Ref.27) and compares the obtained results with calculated ones.

System level tests rely on the computing capacity available in microsystems. Rapid improvements in terms of processing power can be expected during the next few years and it is up to the test designers and reliability engineers to exploit these sources for diagnosis and self-test

5. DIAGNOSTIC INFORMATION

The results obtained from BIST need to be evaluated on their implication on the measurement result and the device availability. A self validating (SEVA) measurement system is described by Henry (Ref.1) and may indicate the quality of output information required from future complex microsystems. It outputs device status information plus three measurement parameters: the (validated) measurement value, an uncertainty value and measurement status value as in figure 3.

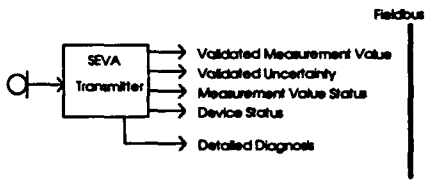


Figure 3 : SEVA Standardised Interface showing the different output parameters of a measurement (Ref.1)

The Device Status is a "simple, generic health indicator" generated by the maintenance / fault detection functions (BIST functions, results of IDDQ tests, checksums etc.).

The uncertainty value indicates the accuracy of the measurement, taking *all* factors affecting the data quality (measurement technology, environment conditions, noise, calibration, maintenance and defect status) into account.

The measurement value status indicates device specific faults. Henry (Ref.1) defines four possible values: CLEAR, DAZZLED, BLURRED and BLIND, "which indicate measurement behaviour ranging from normal (CLEAR) down to a complete absence of transducer data (BLIND)". The MV status will indicate bad measurement conditions such as high temperature and increased humidity which influence the measurement uncertainty.

The advantage of complex microsystems over their "non-intelligent" counterparts is in the provision of evaluated diagnostic information during operation in the field. Such enhanced output information can only be gained with the utilisation of RIs for BIST in microsystems.

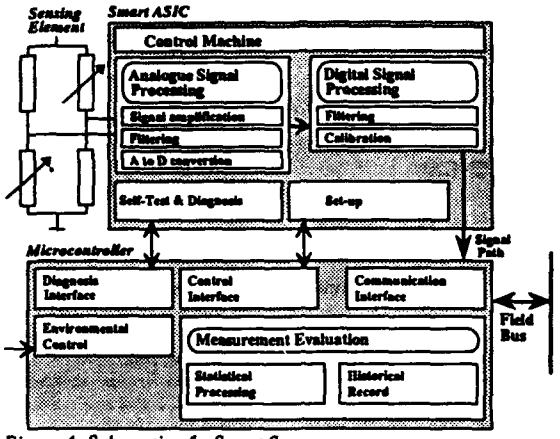


Figure 4: Schematic of a Smart Sensor

6. MULTI-CHIP IMPLEMENTATION OF A MICROSYSTEM

Important questions relating to the integration of BIST and self-test into a microsystem are "How to implement it?" or "Where to allocate the functions?". Complex microsystems to date consist of several modules which are integrated to one device. Such an approach for a smart sensor is described in figure 4. This microsystem consists of three submodules: The sensing element, an application specific integrated circuit (ASIC) as an analogue front-end and a microcontroller to provide processing power for control, set-up and a fieldbus interface (Ref.28). The ASIC is customised to the sensing element and the specific applicational demands for an optimum conditioning of the sensor signal whereas the microcontroller is an off-the-shelf device with large on-chip memory to allow storage of set-up data, filter and calibration coefficients and a record of measurement data.

Self-test and diagnostic features in microsystems are restricted by the processing capabilities available in the device. The usage of RIs and other diagnostic strategies in microsystems implies careful consideration of these limitations.

7. CONCLUSIONS

This paper describes the importance of self-test and diagnosis in microsystems. The utilisation of RIs for on-chip diagnosis is evaluated together with other self-test approaches. A proposal for a multi-chip microsystem shows the requirements, possibilities and limitations for self-tests. This paper outlines a close relationship between DfT for post-production tests and BIST during normal operation.

Microsystems are complex mechatronic systems consisting of mechanical interfaces plus analogue and digital electronics which result in immense difficulties for fast and exhaustive tests after production, during operation or maintenance. BIST and diagnosis are important features of microsystems to increase testability, decrease testing time and improve its reliability during operation.

Reliability Indicators for on-chip diagnosis are an important test strategy for cost effective and more exhaustive testing in high reliability mechatronic microsystems.

8. FUTURE WORK

Reliability Indicators for post-production or maintenance tests using additional test equipment are already used in industrial applications. On-chip diagnosis for mechatronic microsystems using RIs is considered to be a new field of testing and therefore not yet mature. Simple techniques such as on-chip current monitoring have been demonstrated and promise to become useful tools for BIST. Other RI measurements suitable for on-chip implementation need to be developed for analogue and digital circuits as well as for the peripheral interfaces, i.e. sensing elements and actuators.

The microelectronics group at Lancaster University investigates in self-testable, robust analogue designs for CMOS, suitable for microsystem implementation.

Efficient CAD tools and libraries for microsystem design and simulation featuring self-test and diagnosis functions, based on RIs and other techniques, are required as an intrinsic property of their generic modules for cost-effective design and prototyping of future generations of microsystems.

9. ACKNOWLEDGEMENTS

We would like to thank Dr A Bratt and Dr R Harvey of the Microelectronics Research Group, Lancaster University, for their valuable discussions and input.

10. REFERENCES

1. Henry M.P. and Clarke D.W. "A Standard Interface for Self-Validating Sensors", IFAC Fault Detection, Supervision and Safety for Technical Processes, Int. Conference, Baden-Baden, Germany, 1991, pp. 57-64

2. Ormond T. "Smart Sensors get tough", Electronics Weekly, November 10th 1993

3. Huijsing J.H. "Integrated smart sensors", Sensors and Actuators A, 30(1992), pp.167-174

4. Mokwa W. "Application of the Silicon Technology for Microsystems in the Medical Techniques", Micro System Technologies 91, 2nd Int. Conf., Berlin, 1991, pp.19-26

5. Tschulena G. "Micromechanics Business Opportunities", Micro System Technologies 92, 3rd Int. Conference, Berlin, 21-23.10.92,pp.51-57

6. Knieling M. "Technolgies and Market Developments for Microsystem Technologies", Micro System Technologies 91, 2nd Int. Conf., Berlin, 29.10.-1.11.1991, pp.544-551

7. Bergfried D. , Nitschke W. , Rutz M. "Airbag Control Modules - Performance and Reliability", SAE 92C054, pp. 155-162

8. Harvey R.J.A., Bratt A.H. and Dorey A.P. "A review of testing methods for mixed-signal ICs", Microelectronics Journal 24 (1993), pp. 663-674

9. Richardson A.M.D. and Dorey A.P. "Reliability Indicators", 3rd ESREF - European Symposium on reliability of electron devices, failure physics and analysis, Schwäbisch Gmünd, Germany, 5-8 Oct 92, pp. 277-283

10. Richardson A.M.D. and Dorey A.P. "Supply current monitoring in CMOS circuits for relabilty prediction and test", Quality and Reliability Engineering International, Vol.8 p.
543 - 548 (1992)

11. Bratt A H, et al "Aspects of current reference generation and distribution for IDDx pass/fail current threshold determination", IEE colloquium on "Mixed Signal VLSI Test", Digest No: 1993/240, 13.Dec. 1993, London

12. Baker K. "Philips Draft Proposal for QTAG off-chip IDDQ/ISSQ Monitor Standard", version 1.0 QTAG 1993

13. Darling J. "A Study of Acoustic Emission for Condition Monitoring in Machines", IMechE colloquium on "Intelligent Sensors", 25th Nov. 1993, S184, London

14. Higham E.H. "A route to better process measurements", I Chem E Symp Ser, Vol 115, 1989, pp 321-333

15. "Silicon Sensors and Micro-Structures", NOVA SENSOR, 1055 Mission Court, Fremont, CA 94539, USA

16. Russell G. and Elliott I.D. "Design of highly reliable VLSI Processors incorporating concurrent error detection and correction", Euro-ASIC '91, pp. 316-320

17. Savir J. "Built-In Self-Test: Milestones and Challenges", VLSI design 1993, Vol. 1, pp.23-44

18. Al-Qutayri M.A. and. Sheperd R. "Go/no-go testing of analogue macros", IEE proceedings G, Vol.139, No 4, Aug. 1992

19. Lala P.K "Fault tolerant & fault testable hardware design", Prentice Hall, 1985

20. Kolarik V, Lubaszewski M, Courtois B "Towards Self-Checking Mixed-Signal Integrated Circuits", Proc. European Solid State Circuits Conf., pp.202-5, Seville 1993

21. Langford T. "Utilizing boundary scan to implement BIST", International Test Conference 17.-21. Oct. 1993, pp. 167-173

22. Abramovici M., Breuer M., Friedman A.D. "Digital systems testing & testable design", Computer Science Press, New York, 1990

23. Allen H.v. et. al. "Accelerometer Systems with Self-testable Features", Sensors and Actuators A, 20 (1989), pp. 153 - 161

24. Bryzek J. et al., "New Technologies for Silicon Accelerometers enable Automotive Applications", Lucas Nova Sensors, 920474, pp. 25-32

25. Isermann R. "Fault Diagnosis of Machines via Parameter Estimation and Knowledge Processing", IFAC Fault Detection, Supervision and Safety for Technical Processes, Baden-Baden, Germany, 1991, pp.43-55

26. Gertler J. "Analytical Redundancy Methods in Fault Detection and Isolation, Survey and Synthesis", IFAC Fault Detection, Supervision and Safety for Technical Processes, Baden-Baden, 1991, pp.9-21

27. Gertler J. et al. "Model-Based On-Board Fault Detection and Diagnosis for Automotive Engines", IFAC Fault Detection, Supervision and Safety for Technical Processes, Baden-Baden, 1991, pp. 503-508

28. Gare P. "CAN Bus for Local Control", Embedded System Engineering, Feb/March 1994, pp.26-28

TRANSPORT AND NOISE QUALITY INDICATORS FOR PN JUNCTION DIODES

P. Hruska, B. Koktavy, P. Vasina, H. Navarova, V. Jurankova and J. Sikula,
Department of Physics, Technical University of Brno
Zizkova 17, 60200 Brno, Czech Republic, phone-fax: ++425 41211125

1. GENERAL

Quality indicators based on the shape of static *I-U* curves of *PN* junctions are commonly referred to as Transport Indicators. The definition of a meaningfull set of transport indicators should be based on such points of the *I-U* curves, that can indicate imperfections in technology and thus possible degradation of the junction in future. Several hundreds of GaAsP LED diodes, Si Zener diodes, power diodes, etc, were thoroughly examined for a period of several thousands hours. During the examinations a vast collection of data was obtained. Based on the data collection, three basic quality indicators were introduced and examined as follows:

-the forward voltage U_{FI} for a defined current I_{FI},
-the reverse current I_{RI} for a defined bias U_{RI},
-the reverse voltage U_{RB} for a defined current I_{RB},
the quantities U_{RB}, I_{RB} being very close to the junction breakdown.

The indicators description, mainly of U_{FI}, and reasons for their choice are presented in the paper. Beyond the transport indicators, the noise quality indicators were applied. They are described elsewhere (Refs.1,2,3). The maximum noise spectral density in forward direction S_{UM} is used in this paper to discuss the results obtained by application of transport indicators U_{FI}, U_{BR}.

2. CARRIER TRANSPORT IN P-N JUNCTIONS

The PN junction operation is based on well-known transport mechanisms: diffusion mechanism of Shockley, high injection mechanism (in forward polarization), generation-recombination mechanism (in both forward and reverse junction polarization) and Zener and avalanche mechanisms (in reverse polarization). *I-U* curves of diodes with those transport mechanisms are referred to as ideal curves.

The diode degradation and operation failure result generally from uncontrolled technology

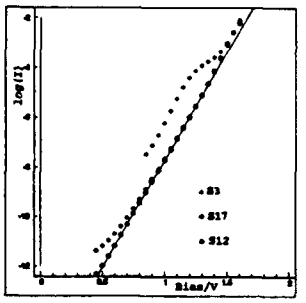


Fig.1 Deviation of I-U curve caused by tunnel current (sample S3) and by volume leakage current (sample S17). Sample S12 exhibits ideal I-U curve

procedures or steps in semiconductor crystal growth, junction formation, surface treatment and packaging. They can produce imperfections such as additional energy states in the forbidden gap, impurity distribution irregularities, local overdoped degenerated microregions in the junction, clusters of impurity atoms, dislocations, deviations from desired junction geometry, etc. Some of the imperfections bring about additional charge carriers transports that can be observed on *I-U* curves as deviations from the curves ideality. The nature of the carrier transports is discussed in the paper.

The additional charge carrier transports were detected in several diodes. To demonstrate their effect on *I-U* curves, Fig. 1 is provided. In it, curves of three GaAsP LED diodes are plotted. Brief inspection shows that regions of diffusion and high injection currents overlap at bias greater than

appr. 1.5V. Solid line in Fig. 1 indicates function *exp(eU/2kT)*, which governs the generation-recombination transport mechanism. Experimental curve of diode S12 coincides with the function within the bias range 0.5V to 1.45V. This curve is the ideal curve. Deviation from ideality can be seen for diode S17 in the range from 0.4V through to .8V and for diode S3 below 1.45V.

2.1 Tunneling in overdoped microregions

In the case of diode S3 one can assume that the current mechanism is of tunneling nature. Electrons tunnel from local overdoped degenerate microregions into unoccupied states in the depletion region. The tunneling is direct and it is accompanied by nonradiative recombination. The asumption is based on the temperature variation analysis of the diodes, which was performed in temperature range 18°C to 75°C. It shows that the current is independent of temperature. The probable origin of local overdoped microregions are dislocations, grown from the junction boundary during the epitaxy process. The dislocations enable access into the transition region to a large number of impurity atoms, thus forming overdoped P^+N microjunctions.

2.2 Volume leakage current

Sample S17 exhibits a deviation of *I-U* curve at lower bias. Corresponding current was investigated by temperature variation analysis. It follows, that the current is proportional to U^{γ}, $\gamma>1$. The nature of the mechanism involved could be explained by a model of two parallel circuit branches. One of them consists of PN junction itself and the other of a PN junction (or metal-semiconductor contact) and an ohmic resistance in series. The ohmic contact is nearly temperature independent.

Microcracs or other crystal imperfections could be responsible for the ohmic resistance and, consequently, for the volume leakage current. On the other hand, formation of microscopic Schottky contacts, resulting from metal atoms transport along dislocations to the transition region, could be accompanied with relatively high spreading resistance. The spreading resistance, which is virtually independent of temperature, represents series ohmic resistance.

2.3 Reverse polarity

High impurity distribution irregularities, that accompany certain technology techniques, affect the transition region width, which, in turn, determines the reverse generation-recombination current and, to a certain degree, the avalanche breakdown voltage. The surface and volume leakage currents cause additional deviation of reverse *I-U* curves from ideality. The quality indicators I_{RI} and U_{R9} can control effect of such flaws.

3. REAL *I-U* CURVES

The summary of individual current components, associated with mechanisms discussed above, is schematically indicated in Fig. 2. Tunnel current is indicated by curve 4, diffusion current by curve 3, generation-recombination current by curve 2 and volume leakage current by curve 1. It is clear from the picture, that the occurence of the tunnel

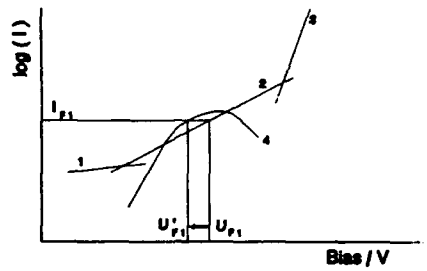


Fig.2 Effect of tunneling current on quality indicator U_{FI}. Volume leak current 1, generation-recombination current 2, diffusion current 3, tunneling current 4.

curent results in a bias drop for a preselected current I_{FI}. Corresponding bias on curve 2 (or 3) can be taken for the transport quality indicator. The value of I_{FI} should not exceed 100 μA (as applied in this paper). It could be even lower, generally within the range 0.1 μA to 100 μA.

In a similar manner, the occurence of the volume leakage current can be controlled by au

additional transport quality indicator, possibly I_{F2}. Its value should be low, below 1 nA. Diode examination according to I_{F2} was not included in this paper.

4. RESULTS. CONCLUSIONS

For comparison of the sensitivity and efficiency of the indicators a set of 20 GaAsP LED samples was examined. The set was aged for a period of 6000 hours. The results are presented in Fig. 3,4,5, where only data for a group of typical samples, same in all figures, are plotted.

4.1 Transport indicators

The time variation of the U_{F1} indicator is in Fig. 3.

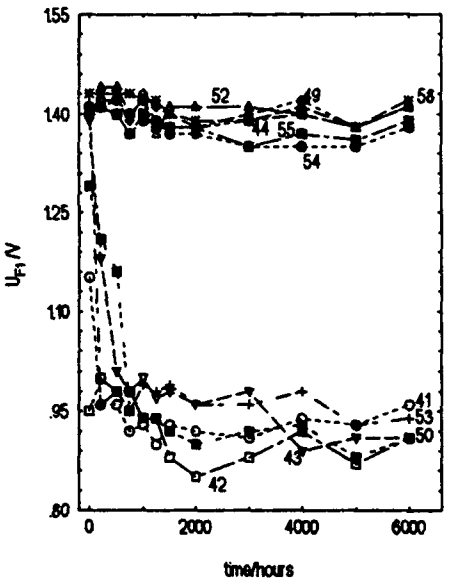


Fig.3 U_{F1} indicator for a group of typical samples of GaAsP LEDs at I_{F1} = 100 μA

It follows,that three samples of the set exhibited considerably lower values of U_{F1}, ranging from .85 to 1.3 V (No 41, 42, 50) at the beginning of the aging process. Seventeen samples had the value U_{F1} very close to the ideal (theoretical) value

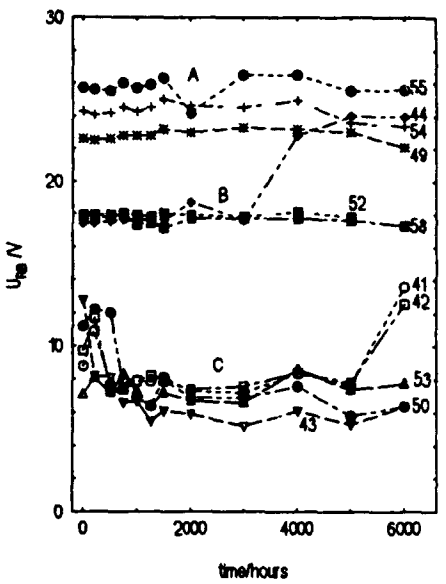


Fig.4 U_{RB} indicator for a group of samples of GaAsP LEDs at I_{RB} = 100 μA

of 1.40 V at I_{F1}=100 μA. After 200 hours of aging the value of samples 43, 53 dropped and these samples joined the three former samples with low U_{F1}.

The spread of values U_{F1} in good samples is caused by non-uniformity in doping and, consequently, by the depletion layer width. The time evolution of this ensemble of samples shows an interesting feature: average values quality indicators for high quality samples vary only slightly over the whole aging period, whereas those for the other samples decreased considerably.

The time variation of the U_{RB} indicator for I_{RB}=100 μA is given in Fig. 4, where three groups of samples can be distinguished. Groups A, B proved to be stable during the aging process. Samples of group C exhibit soft breakdown and their U_{RB} values are much lower than in groups A, B. During the aging process the U_{RB} indicator decreased for certain samples. Samples, as No 53, were good at the beginning of aging from the poin

of view of the indicator U_{FI}, but not acceptable from the point of view of the U_{RB} indicator. These samples exhibit excess noise in forward direction, as mentioned below.

4.2 Noise measurement

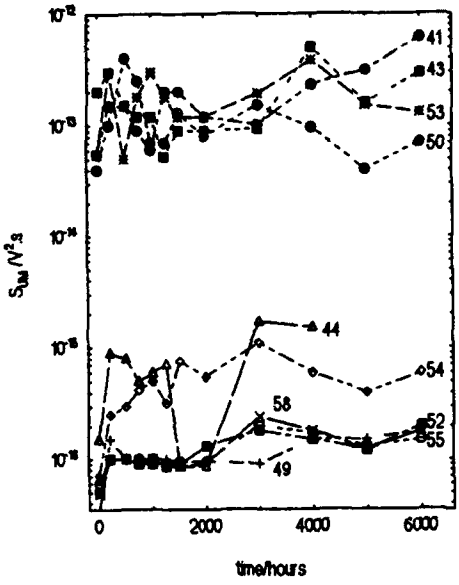


Fig.5 S_{UM} indicator for a group of samples of GaAsP LEDs

In the course of the aging process the maximum spectral density in the forward direction S_{UM} was measured. In Fig. 5 the time evolution of S_{UM}, measured across a load resistor of 10 kΩ, is indicated. Two groups of samples can be distinguished. They are separated by a gap of at least two decadic orders of magnitude. Inspection of the sample numbers shows that results of transport and noise measurements are closely related and there is a strong correlation among indicators U_{FI}, U_{RB} and S_{UM}. Samples exhibiting low value of U_{FI} have low value of U_{RB} and high value of S_{UM}. Good samples exhibit high value of U_{FI} and U_{RB}, while S_{UM} is rather low. From the point of view of transport there are only three poor samples in the set. (No 41, 42, 50) at the very beginning of the aging. After 200 hours of aging samples No 43 and 53 also deteriorated. It is worthwhile noting that these two samples had high values of of S_{UM} at the beginning. Application of noise indicators seem to be more sensitive to defects and degradation process and, thus, more efficient in quality predictions.

6.REFERENCES

1. Vašina P. "Noise, Degradation and Reliability of Luminescent Diodes", Journal of Electrical Ingeneering, Vol.45, 1994, No 3, pp 90-94.

2. Šikula J., Koktavý B., Vašina P. and Čermáková A.: 1/f Noise, Degradation and Reliability of GaAsP LEDs", Procc. of 10th Int. Conf. of Noise in Physical Systems, Budapest, Hungary, 1989, pp 479.

3. Jones B.K., "Electrical Noise as a Measure of Reliability in Electronic Devices", Advances in Electronics and Electron Physics, 1994, Vol.67, pp 201.

APPLICATION OF THE RELIABILITY DEFECT DENSITY CONCEPT FOR IC FAILURE RATE ASSESSMENT

P. De Pauw, G. Gallopyn, and L. Vandenbroucke
Alcatel Mietec, Westerring 15, B-9700 Oudenaarde.

ABSTRACT

In order to predict IC failure rate during IC lifetime, one has to take into account the defected population. Methods in order to quantify this defected population are difficult to apply to high reliable integrated circuits and their manufacturing processes. The usefulness of the parameter "reliability defect density" is explained for IC reliability prediction and assessment. As opposed to conventional methods only one parameter for each failure mechanism is necessary in order to predict IC reliability.

As an example the methodology has been worked out for the reliability assessment of gate oxides. It is shown how the "reliability defect density" can be obtained from a simple constant voltage stress test or from available ramp voltage breakdown measurement data. The proposed method is featured by a short test duration, a low overstress, and a moderately large amount of large area test samples. Experimental ramp breakdown data for a practical technology are presented and analysed. Using the measured reliability defect density, the oxide related part of the IC failure rate is predicted for four different use condions.

1. INTRODUCTION

Extensive field return analysis programs (Refs. 1-2) have investigated the IC failure modes occuring during operation in the field. Although dependent on environmental and operational conditions, a general picture is apparent : roughly half of the confirmed failures can be attributed to electrical overstresses (EOS/ESD) while the other half is waferfab or assembly process related. The waferfab and assembly related failures can be subdivided into two main categories : "maverick" or "freak" wafers due to operator errors or equipment malfunctioning (e.g. metal thickness below specification, bad wire bond parameters applied, contamination, ...), and small manufacturing defects. These defects must be viewed in a broad perspective : any localized physical material unhomogenity which weakens a component. Defects can be caused by particles, but can also originate from material or processing anomalies. Examples are local gate oxide thinning or metal voids, but also scratches during wafer handling, probing and transport, occasional high ohmic vias, occasional weak wire bonds, surface asperities below thin oxides etc.

As the technology margins become smaller and IC die areas increase during the ongoing scaling evolution, IC's become more vulnerable to manufacturing defects. On the other hand, the reliability targets on component level become tougher. A failure rate as low as a few ppm / year is becoming a realisitic requirement in high reliability markets. These trends necessitate the implementation of efficient (early) failure rate assessment methods.

Conventional failure mechanism assessment methods, using a low number of test structures with relatively small area, focus on intrinsic phenomena, i.e. wear-out of the IC. Although both useful and necessary for the characterization of new processes during technology development, these methods fail in detecting present-day low defect densities. Reliability requirements in some applications have risen considerably. Since field return programs suffer from high cost and excessive feedback times, methods are required in order to assess the required defect levels in the wafer fab.

2. FIELD FAILURE RATE ASSESSMENT

2.1 IC failure rate model

The set of failure mechanisms possibly affecting the reliability of integrated circuits, can be subdivided in time dependent and stress event related failure mechanisms. The cumulative amount of failures of an integrated circuit throughout its lifetime, $F_{IC}(t)$, can be approximated as the sum of the cumulative amount of failures of these individual failure mechanisms, as these fractions are small in virtually all practical cases :

$$F_{IC}(t) = F_{EM}(t) + F_{TDDB}(t) + F_{COR}(t) + \ldots + F_{ESD} + F_{THCYCL} + \ldots \quad (1)$$

For a time dependent failure mechanism i, the cumulative amount of failures $F_i(t)$ can be written as:

$$F_i(t_{use}) = P_{d,i} \cdot F_{d,i}(t_{use}) + (1-P_{d,i}) \cdot F_{m,i}(t_{use}) \quad (2)$$

With :

- $F_{d,i}(t)$: the cumulative distribution function for the defected population (early failure fraction) for failure mechanism i
- $F_{m,i}(t)$: the cumulative distribution function for the intrinsic population (wear-out failure fraction) for failure mechanism i
- t_{use} : the effective operational time : $t_{use} = t_{test} \cdot Acc_i / \eta_{DF}$
- Acc_i : the acceleration factor between test and use conditions for failure mechanism i
- η_{DF} : the operational duty cycle factor of the application : $\eta_{DF} = t_{ON} / (t_{ON}+t_{OFF})$
- $P_{d,i}$: the fraction of IC's containing a defect for failure mechanism i :

For IC's, manufactured in well designed technologies with capable processes, the intrinsic failure fraction will be negligible during typical component lifetimes (<25 years) at typical operating conditions (T_j < 125°C). The field failures will consist predominantly of components of the defect related populatio:

$$F_{IC}(t_{use}) \approx \sum_i P_{d,i} \cdot F_{d,i}(t_{use}) \quad (3)$$

Note that not necessarily all of the components with an internal defect will fail during operational life.

2.2 Global defect fraction method.

Conventionally, assessment of the IC failure rate is done by performing an accelerated life test. In order to investigate one failure mechanism, dedicated test structures are used. After the test, all parameters of the model expressed by eq. (3) are determined by a bimodal fit from the lifetest data: the global defected fraction $P_{d,i}$, and the shape and location parameters of the failure distribution function $F_{d,i}(t)$. For small defect densities, the sample sizes required are impractically large.

2.3 Reliability defect density method

The IC failure model described above requires for each failure mechanism the knowledge of the full distribution function. However in practice the devices which appear as field failures constitute only a fraction of the total defected population, dependent on the operational environmental and electrical conditions. Consequently it must be possible to assess field failure rates with only part of the information needed in the general model.

Defining a parameter $P_{r,i(IC)}$ as :

$$P_{r,i(IC)} = P_{d,i} \cdot F_{d,i}(t_{use}) \quad (4)$$

allows to rewrite eq. (3) as :

$$F_{IC}(t_{use}) \approx \sum_i P_{r,i(IC)} \quad (5)$$

The parameter $P_{r,i(IC)}$ can be described using the concept of reliability defect density. The term "reliability defect density", as opposed to the well known "yield defect density", has been introduced a few years ago (Ref. 3) and this concept is also applied in recent other studies (Ref. 4). While yield defects cause integrated circuits to fail at the first electrical tests, reliability defects will cause failure later on during operational life. Smaller defects not causing a reliability fail are not included in this figure. While yield defect density is relatively easily measurable, efficient measurement methods in order to assess reliability defect density are lacking.
We define reliability defect density as the number of defects per unit area which would cause a failure for a specified failure mechanism during operation for a prescribed period of time under well defined environmental and electrical conditions. This parameter is not only dependent on defect density, but also on the operational environment and the studied period ; defects that would cause failures in harsh environments might survive under mild operational conditions.

Using the concept of reliability defect density, the field failure rate assessment for a time dependent failure mechanism then proceeds as follows :

- calculation of test conditions : applied stress S_{test} and T_{test} and test time t_{test} . These accelerated test conditions should be equivalent to the component use conditions. They are determined using theoretical or empirical acceleration models. Overstress is kept as low as practically possible in order to minimize errors due to the lack of an established law which allows extrapolation of failure rates found at high stress conditions during test to low stress conditions during operation. The effective operational time t_{use} is to be derived from the required lifetime t_{life} and the operational duty factor η_{DF} :

$$t_{use} = t_{life} \cdot \eta_{DF} \quad (6)$$

- Test structures: In order to assess low defect densities, either an extremely large number of conventional test structures, either a moderately large number of large area test structures is needed. The preferred approach is to use test structure areas as large as practically possible, in casu existing yield defect density structures.

- After application of the stress test, the number of failing test structures is counted, and $P_{r,i(test)}$ is determined as the fraction of failing test structures devided by the sample size.

- The reliability defect density under operational conditions is determined, according to the well-known yield equation $Y = i - P = 1 / (1 + D.A)$:

$$D_{r,i} = 1 / A_{crit,i(test)} \cdot P_{r,i(test)} / (1 - P_{r,i(test)}) \quad (7)$$

- Finally, for any given IC and for failure mechanism i, the failing fraction during lifetime, under the specified conditions, is determined as follows:

$$P_{r,i(IC)} = (D_{r,i} \cdot A_{crit,i(IC)}) / (1 + D_{r,i} \cdot A_{crit,i(IC)}) \quad (8)$$

In most cases $P_{r,i(test)}$ is small, and eq. (8) can be approximated by :

$$P_{r,i(IC)} \approx (A_{crit,i(IC)}) / A_{crit,i(test)}) \cdot P_{r,i(test)} \quad (9)$$

As an example of this methodology, we explore the use of oxide reliability defect density in the following section. The technique can also be applied to assess the field reliability of metal interconnect (Ref. 5).

3. EXAMPLE: OXIDE RELIABILITY DEFECT DENSITY

Available techniques in order to predict oxide reliability are constant voltage and ramped voltage. When the defected population of the gate oxide is small, assessment of gate oxide reliability is impractical using conventional constant voltage methods, while ramped voltage so far is not used to predict gate oxide reliability. As gate oxide reliability steadily improves, while IC reliability requirements steadily increase, it becomes necessary to develop methods suited for high reliability assessment. Therefore we adapted both methods in order to make them suitable for high reliability assessment.

3.1 Theoretical background

As acceleration model we selected the reciprocal field model combined with the "effective oxide thinning" concept (Ref. 8), which accounts for the presence of defects. Defects are modeled as localized oxide thinning. The relation between the breakdown time and the thickness of the remaining oxide X_{eff} is expressed as follows (Ref. 8) :

$$t_{BD} = \tau_0 \cdot \exp[\, G.X_{eff} / V_{ox} \,] \quad (10)$$

with : t_{BD} : time to breakdown
X_{eff} : effective oxide thickness
V_{ox} : stress voltage
τ_0 : intrinsic breakdown time
G : field acceleration parameter

Determination of G and τ_o for each process is impractical. As both parameters do not vary much between similar processes, it is possible to rely on published data. Note that G and τ_o are both temperature dependent. Typical values at room temperature are $\tau_o = 1.0 \times 10^{-11}$ s and $G = 350$ MV/cm (Ref. 9). Values at other temperatures can be calculated from formulas developed in Ref. 10. At 55°C these parameters have the values $\tau_0 = 3.0 \times 10^{-11}$ s and G = 325 MV/cm and at 125°C, $\tau_0 = 1.7 \times 10^{-10}$ s and G = 290 MV/cm.

The smallest value of X_{eff} leading to no IC failures during operational life, X_{eff}^*, can be determined from eq. (10) as follows:

$$X_{eff}^* = \frac{V_{use}}{G(T_{use})} \cdot \ln[\, t_{use} / \tau_0(T_{use}) \,] \quad (11)$$

3.1.1 Constant voltage method

In this test, the stress conditions, V_{test} and T_{test}, are applied, during a fixed time period t_{test}. At the end of the stress test, the failing fraction $P_{r,TDDB(test)}$ is recorded. This failing fraction is the same as would occur during lifetime for an IC with the same gate oxide area, stressed with V_{use} and T_{use}, during a time period t_{use}. The reliability defect density for the failure mechanism TDDB, $D_{r,TDDB}$ can be determined using eq. (7).

Applying equation (10) both to the stress test conditions and to the use conditions allows to express t_{test} as follows :

$$\ln t_{test} = \ln \tau_0(T_{test}) + \frac{G(T_{test}).V_{use}}{G(T_{use}).V_{test}} \cdot \ln[t_{use}/\tau_0(T_{use})] \quad (12)$$

This expression allows to determine t_{test}, for the given use conditions, and the selected stress test conditions V_{test} and T_{test}.

During constant voltage stress tests, we prefer to limit the electrical field to 6 MV/cm as there may be different physical mechanisms active at high and low field stress (Ref. 7).

3.1.2 Ramped voltage method

In this test, a ramped voltage is applied at the temperature T_{test}. The breakdown voltage of each test structure is recorded. The breakdown voltage, V_{BD}*, corresponding to an effective gate oxide thickness X_{eff}* is determined. The breakdown voltage interval [V_{use} , V_{BD}*] corresponds to the lifetime period, and is called the reliability voltage interval. The failing fraction $P_{f,TDDB(test)}$ is determined as the number of fails occuring in the reliability voltage interval [Vuse, V_{BD}*], divided by the sample size.

Ramped voltage measurements data are available from measurements on yield structures on large amounts of samples, since this test type is frequently performed during technology development and as a yield monitoring test. Therefore it is logical to use these data for reliability assessment purposes.

A good approximation for the relation between the ramp breakdown voltage V_{BD} at a given ramp rate R, and the effective oxide thickness X_{eff} is given by (Ref. 10):

$$X_{eff} \approx \frac{V_{BD}}{G(T_{test})} \cdot \ln\left[\frac{V^2_{BD}}{R \cdot \tau_0(T_{test}) \cdot G(T_{test}) \cdot X_{eff}}\right] \qquad (13)$$

For a given value of Xeff = Xeff*, determined using eq. (11), V_{BD}* can be solved from eq. (13).

For oxides with excellent reliability, for small sample sizes, or for test structures of small area, the number of fails in the interval [Vuse , V_{BD}*] is small, or can be even zero. In this case $P_{f,TDDB(test)}$ can be determined with a better accuracy by interpolation. The validity of this interpolation is shown below.

For constant voltage test results, it is known that, when the cumulative failing fraction versus t_{BD} is plotted on log-normal or on log-Weibull paper, the defect related and the intrinsic population fit a straight line (Ref.6).

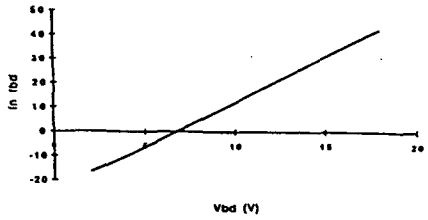


Figure 1 : Relation between ln t_{BD} and V_{BD} calculated according to eq. (14).

From Ref. 10, we find the following relation between t_{BD} and V_{BD} :

$$\ln t_{BD} = C_1 \cdot V_{BD} \cdot [\ln V_{BD} + C_2] + C_3 \qquad (14)$$

with t_{BD} the time to breakdown under constant voltage stress, V_{BD} the ramp breakdown voltage, and C_1, C_2, and C_3 constants. Eq. (14) is illustrated in figure 1 : a nearly straight line is obtained over the voltage interval [2V, 18V]. Therefore a linear relationship between ln t_{BD} and V_{BD} can be assumed and eq. (14) can be approximated by :

$$\ln t_{BD} = C_4 \cdot V_{BD} + C_5 \qquad (15)$$

		Operational conditions @ 5.5V			Test definitions @ Troom	
		T_j (°C)	Period (years)	ηDF (%)	t_{BD} @10.5V (ms)	V_{BD} @5V/s (V)
Early Life	A	125	[0 - 1]	100	782	11.4
	B	125	[0 - 1]	5	118	10.6
	C	55	[0 - 1]	100	128	10.6
	D	55	[0 - 1]	5	24	9.9
Total Life	A	125	[0 - 25]	100	5960	12.3
	B	125	[0 - 25]	5	900	11.5
	C	55	[0 - 10]	100	464	11.2
	D	55	[0 - 10]	5	86	10.5

Table I : Test definition examples for constant field stress tests (t_{BD}) and ramp voltage tests (V_{BD})

		Operational conditions @ 5.5V			N well	P well	IC
		T_j (°C)	Period (years)	ηDF (%)	D_r (cm^{-2})	D_r (cm^{-2})	$F_{IC}(t)$ (ppm)
Early Life	A	125	[0 - 1]	100	0.80	0.36	101
	B	125	[0 - 1]	5	0.63	0.28	79
	C	55	[0 - 1]	100	0.63	0.28	79
	D	55	[0 - 1]	5	0.50	0.22	63
Total Life	A	125	[0 - 25]	100	1.04	0.45	129
	B	125	[0 - 25]	5	0.83	0.36	103
	C	55	[0 - 10]	100	0.76	0.33	95
	D	55	[0 - 10]	5	0.61	0.27	77

Table II : Reliability defect densities for N and P well gate oxide capacitors and cumulative number of failures for an IC with a total gate oxide area of 0.2 mm^2

Consequently, when the cumulative failing fraction versus V_{BD} is plotted on lin-normal or on lin-Weibull paper, the defect related and the intrinsic population fit on a straight line. Therefore, in order to determine $P_{r,TDDB(test)}$ from a ramped voltage plot, linear interpolation with the data points outside the reliability voltage interval is possible on lin-normal or lin-Weibull paper.

3.1.3 Comparison between the Constant voltage and the Ramped voltage method

The ramped voltage method has a big advantage over the constant voltage method in that the measurement data are readily available in many cases.
On the other hand, the accuracy of the ramped voltage method entirely depends on the accuracy of eq. (13). This accuracy is not easy to assess.

For a gate oxide with a thickness of 17.5 nm and for four typical use conditions, denoted as A, B, C, and D, the corresponding stress test conditions were calculated for a constant voltage test at 10.5 V (corresponding with E_{max} = 6 MV/cm) and a ramped voltage test at 5V/s ramp rate. The results are shown in Table I.

3.2 Experimental data and discussion

A set of CMOS test structure capacitors was used as part of a yield monitor structure. Gate oxide thickness is 17.5 nm and the area of each capacitor is 1.0 mm^2. The gate oxides are grown on N and P-wells. They were processed according to a standard process flow. Data was collected during a period of about 2 months. About 10 000 capacitors were subjected to a ramped voltage test during this time period.

The breakdown data are shown in figures 2 and 3 for the N and P-well capacitors respectively. In these graphs the population failing below 5 V has been

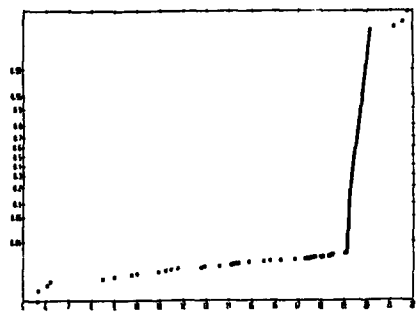


Figure 2 : Ramp voltage breakdown data for N- well gate oxide capacitors (failure fraction vs. voltage)

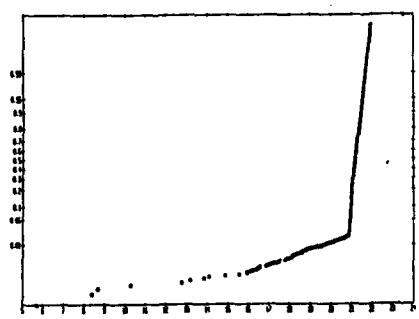


Figure 3 : Ramp voltage breakdown data for P-well gate oxide capacitors (failure fraction vs. voltage)

censored as it is yield loss and will not reach the field. Also the part of the population which survived at voltage compliance, due to probing problems etc., was removed from the data. The data are presented in a lin-normal chart. The defect related part in these experimental breakdown curves follows a straight line, in accordance to the theory developed above.

As long as a few failures in the voltage range 5V - 15V are present, a reasonable fit within the reliability voltage interval is possible. In our case results of acceptable accuracy could also have been obtained with a considerably lower sample size.

For an integrated circuit with a gate oxide area of 0.2 mm^2, IC failure rates were determined, for the four selected use conditions A, B, C, and D. The results are shown in Table II. Comparing the early failure results for condition B, which is equivalent to 450 Hrs continuous operating time, to the long term failure results for condition A, corresponding to 225 kHrs operating time, it is apparent that most oxide related field failures will occur in the first weeks of operation for components processed in the evaluated technology. This also explains why the influence of the operational duty factor on the failure fractions is small : even in applications with low duty cycle, most of the oxide related failures will occur shortly after installation.

Another observation from table II is that the influence of operational temperature on the field drop-out is relatively small. The higher temperature causes an even higher concentration of failures immediately after voltage application.

These data can be used for the following purposes:
- process qualifications
- assessment of IC failure rates
- determination of the effectiveness of voltage screens, applied on products at electrical test in order to improve reliability for high reliability products. Because the occurance of circuit degradation problems such as hot carrier degradation and punch through, a voltage screen, which would eliminate all first year failures, is not always possible.

From these observations it can be concluded that for the particular oxide evaluated most oxide related field failures will occur in the first operational year for a broad range of temperatures and duty cycles.

3.3 Further work.

Currently work is ongoing in order to compare the results from the ramped voltage method with the constant voltage method.

4. CONCLUSION

The usefulness of the parameter "reliability defect density" is explained for reliability assessment of integrated circuits. As opposed to conventional methods only one parameter per failure mechanism is necessary in order to predict IC reliability.

As an example the methodology has been worked out for the reliability assessment of gate oxides. It is shown how the "reliability defect density" can be obtained from a simple constant voltage stress test or from available ramp voltage breakdown measurement data. The proposed method is featured by a short test duration, a low overstress, and a moderately large amount of large area test samples.

Experimental ramp breakdown data for a practical technology were presented and analysed. Using the measured reliability defect density, the oxide related part of the IC failure rate was predicted for four different use condions.

5. REFERENCES

1. Stephens C.E. "The Relation between Failures at Qualification and in Field Service", Proc. ESREF, 1992, pp 17-25.
2. Sinnadurai N., Kuppuswamy T.S., Chandramouli R. and Rao B.K.N. "Environmental Testing and Component Reliability Observations in Telecommunications Equipment Operated in Indian Climatic Conditions", Proc. ESREF, 1991, pp 55-63.
3. Huston H.H. and Clarke C.P. "Reliability Defect Detection and Screening during Processing - Theory and Implementation", Proc. IRPS, 1992, pp 268-275.
4. Bruls E.M.J.G. "Reliability Aspects of Defect Analysis", to be published.
5. De Pauw P., Gallopyn G. and Vandenbroucke L. "Reliability Defect Density : Measurement and Use", VDE ITG workshop, May 10-11 1994, Murnau, Germany.
6. Vollertsen R.-P. "Pragmatic Procedure for the Estimation of Thin Dielectric Reliability", Quality and Reliability Engineering International, vol. 8, 1992, pp 557-564.
7. Boyko K.C. and Gerlach D.L. "Time dependent Dielectric Breakdown of 210Å Oxides", Proc. IRPS, 1989, pp 1-8.
8. Lee J., Chen I.-C. and Hu C. "Statistical Modeling of Silicon Dioxide Reliability", Proc. IRPS, 1988, pp 131-138.
9. Moazzami R., Lee J. and Hu C. "Temperature Acceleration of Time dependent Dielectric Breakdown", IEEE Trans. Electron Devices, vol. 36, no. 11, Nov. 1989, pp 2462.
10. Moazzami R. and Hu C. "Projecting Gate Oxide Reliability and Optimizing Reliability Screens", IEEE Trans. Electron Devices, vol. 37, no. 7, Jul. 1990, pp 1643-1650.

END

FILMED

DATE: 1-95

DTIC