

DTIC COPY

1

AD-A220 721

DTIC
ELECTE
APR 23 1990
S D D

ANNUAL REPORT

ULTRA HIGH SPEED COMPOUND SEMICONDUCTORS
AND
REAL TIME SIGNAL PROCESSING

MAY 1, 1989 - APRIL 30, 1990

CONTRACT #F49620-87-C-0044

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

04 17 05

ANNUAL REPORT

Joint Services Electronics Program
Contract # F49620-87-C-0044
May 1, 1989 - April 30, 1990

ULTRA HIGH SPEED COMPOUND SEMICONDUCTORS AND REAL TIME SIGNAL PROCESSING

March 1, 1990

J. P. Krusius
Principal Investigator

Cornell University,
School of Electrical Engineering
Ithaca, New York 14853

Accession For	
NTIS - NAB	☒
DTIC - TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution /	
Availability Codes	
Dist	Availability or Special
A-1	

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT Approval for public release. Distribution unlimited.		
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Cornell University		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research		
6c. ADDRESS (City, State, and ZIP Code) 119 Phillips Hall Ithaca, NY 14853-5401			7b. ADDRESS (City, State, and ZIP Code) Building 410, Bolling Air Force Base Washington, DC 20332-6448		
8a. NAME OF FUNDING / SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-87-C-0044		
8c. ADDRESS (City, State, and ZIP Code)			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.
11. TITLE (Include Security Classification) Ultra High Speed Compound Semiconductors and Real Time Signal Processing					
12. PERSONAL AUTHOR(S) J. Peter Krusius					
13a. TYPE OF REPORT Annual Report		13b. TIME COVERED FROM 5/1/89 TO 4/30/90		14. DATE OF REPORT (Year, Month, Day) 90/03/01	
15. PAGE COUNT					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Compound semiconductor, organometallic vapor phase epitaxy, femtosecond laser probing, Monte Carlo simulation, VLSI, fault tolerance, computer architecture		
FIELD	GROUP	SUB-GROUP			
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This report is the annual report on research conducted under the auspices of the Joint Services Electronics Program at Cornell University. The research is grouped under two themes: (a) ultra high speed compound semiconductors, and (b) real time signal processing. Results on OMVPE materials growth, femtosecond laser probing of hot carriers, and ensemble Monte Carlo simulations are reported on under the first theme. Accomplishments on VLSI algorithms, fault tolerant architectures, and architectures with multiple functional units for signal processing are given under the second theme.					
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL J. Peter Krusius			22b. TELEPHONE (Include Area Code) (607) 255-3401		22c. OFFICE SYMBOL

TABLE OF CONTENTS

	<u>Page</u>
A. Director's Overview	1
B. Description of Special Accomplishments and Technology Transition	2
B.1 Femtosecond Carrier Processes in Compound Semiconductors	2
B.2 Real Time Signal Processing.....	3
C. Description of Individual Work Units.....	4
C.1 Femtosecond Carrier Processes in Compound Semiconductors	
Task #1 OMVPE Growth of III-V Alloys and Structures for New High Speed Electron Devices.....	5
Task #2 Femtosecond Laser Studies of Ultrafast Processes in Compound Semiconductors.....	12
Task #3 Ultrafast Studies of Device Structures and of Hot Carriers in Narrow Bandgap Semiconductors.....	16
Task #4 Physics of High Speed Meso Scale Compound Semiconductor Devices	22
C.2 Real Time Signal Processing	
Task #5 Novel VLSI Algorithms and Architectures for High Data Rate Digital Filtering.....	30
Task #6 Fault Tolerant Signal Processing Architectures	35
Task #7 Real-Time Signal Processing Systems with Multiple Functional Units.....	40

A. DIRECTOR'S OVERVIEW

This document is the second year annual report of the Cornell Joint Services Electronics Program for the period from May 1, 1989 to April 30, 1990. The Cornell program was broadened from an exclusive focus on compound semiconductor materials and devices into a two theme approach at the beginning of this two year period. One of these themes continued the compound semiconductor research concentrating on more fundamental phenomena, femtosecond transport and optical phenomena in heterostructures, while the other theme added research in a new area, real time digital signal processing, to the program. The new objectives brought four new faculty (C. Pollock, G. Bilardi, F. Luk and H. Torng) to a total of seven principal investigators now participating in the program.

A major optoelectronics proposal was prepared to DARPA with C. Tang of Cornell University, one of the JSEP work unit leaders, as the principal investigator largely leveraging research interactions and expertise of past and current JSEP research at Cornell. The proposal titled "National Optoelectronic Materials Center" involved a team of universities (Cornell University and University of California Santa Barbara (UCSB) as the main institutions with contributions from Rensselaer Polytechnique Institute and Syracuse University). According to Congressional press releases and DARPA, the highest award went to a team lead by the University of Southern California but the Cornell/UCSB team was awarded \$6M to \$7M for a two year period. Contract negotiations between DARPA and the Cornell/UCSB team are now under way. At the time of this writing, it appears that research proposed by all other JSEP task investigators of the compound semiconductor theme (R. Shealy, C. Pollock, and J.P. Krusius) will also receive funding under the DARPA optoelectronics program.

Efforts to establish the new compound semiconductor growth facility at Cornell have continued to be a collaborative issue for the JSEP faculty during the current program period. The new facility is expected to require continued attention well into future years of the program. It also serves as one of the cornerstones to the DARPA program. The old organometallic vapor phase epitaxy (OMVPE) operation on the fourth floor of Phillips Hall had become substandard because of ever tightening hazard gas safety regulations. The new growth laboratory, a shared facility operated under the technical direction of R. Shealy, a JSEP work unit leader, and an oversight committee consisting of users and administration, is being established in an existing building off-campus. The facility will comply with the most stringent hazard gas safety regulations in the country ("California code"). Construction of this new facility in an existing building calls for the installation of three independent OMVPE systems to be used for specialized growth tasks. One of these reactors will be the rebuilt reactor moved from Phillips Hall, the second a fully operational reactor donated by General Electric Company, and the third a reactor currently under construction. The facility is scheduled to be operational in the fall of 1990. Much of the effort in R. Shealy's JSEP task has been devoted to the

establishment of this new OMVPE facility, which will provide truly unique capabilities for compound semiconductor heterostructure growth. Compound semiconductor JSEP faculty in addition to R. Shealy have contributed to the planning, fund raising, and extensive discussions with university administration because of the importance of this facility to JSEP and related research. The valuable and extensive experience acquired during the planning and construction of this facility will be made available to any interested DoD laboratory on request.

The three task investigators in the present program contributing to the real time signal processing theme, G. Bilardi, F. Luk and H. Torng, have held regular meetings, identified overlapping research areas, and defined unifying research issues. Significant results have been accomplished by this group in the second year of the JSEP program.

B. DESCRIPTION OF SPECIAL ACCOMPLISHMENTS AND TECHNOLOGY TRANSITION

B.1. Femtosecond Carrier Processes in Compound Semiconductors

Significant accomplishments both in facilities and research have been achieved. The construction of the new off-campus OMVPE facility has started under the direction of task investigator R. Shealy. The first of the three reactors is scheduled to become operational in the fall of 1990. The JSEP solid state faculty was instrumental in the planning, design, and fund raising activities. This facility is likely to provide the most modern and versatile OMVPE growth capability for a variety of compound semiconductor materials anywhere.

The work on new unique tunable femtosecond laser sources has been continued. The first one, a high repetition rate UV femtosecond source is based on intracavity frequency doubling in a BaB_2O_4 crystal. The second broadly tunable red to mid IR femtosecond laser source employs resonant parametric oscillation and a KTiOPO_4 . The third source is a high power color center laser tunable in the 0.7 to 0.85 eV photon energy range. The JSEP investigators have already started to exploit these femtosecond laser sources for the study of the dynamics of carrier processes in a variety of compound semiconductor materials, heterostructures, and devices. A dual carrier ensemble Monte Carlo transport simulation method has been developed to simulate and analyze these femtosecond experiments. It is expected that significant advances, understanding of femtosecond carrier processes, and perhaps real breakthroughs, will accrue from the close cooperation of experimental and theoretical tasks.

Two of the JSEP investigators have spent their sabbatical leaves at research laboratories on problems related to JSEP research. C. Pollock worked for six months at the NRL in Washington, D.C. on lasers and P. Krusius was for the entire academic year 1988/89 at IBM Research at Yorktown Heights studying on hot carrier transport in the strain layer $\text{Ge}_x\text{Si}_{1-x}/\text{Si}$ material system.

B.2 Real Time Signal Processing

The task investigators in this theme started JSEP research just two years ago. Despite of this, significant accomplishments have already been achieved. The Naval Ocean Systems Center (San Diego, CA) is building a linear algebra parallel processor based on the work of F. Luk. The Boston based company, Computational Engineering, Inc., has constructed a transputer-based systolic array, also based on F. Luk's work, for real time analysis of airplane wing flutter for the Army. A patent for the dispatch stack, a new method for speeding up RISC processors, was issued. H. Torng is continuing his work on the dispatch stack for real time computing systems with multiple functional units. H. Torng is organizing the second meeting of "Project 2000", an interactive partnership between academia and industrial researcher working on high speed computers for the future. It will be held on the Cornell Campus, June 4 and 5, 1990. Close research interactions with Intel and AMD on superscalar computers have been established.

C. DESCRIPTION OF INDIVIDUAL WORK UNITS

Page

C.1 Femtosecond Carrier Processes in Compound Semiconductors

Task #1	OMVPE Growth of III-V Alloys and Structures for New High Speed Electron Devices.....	5
Task #2	Femtosecond Laser Studies of Ultrafast Processes in Compound Semiconductors.....	12
Task #3	Ultrafast Studies of Device Structures and of Hot Carriers in Narrow Bandgap Semiconductors.....	16
Task #4	Physics of High Speed Meso Scale Compound Semi- conductor Devices.....	22

C.2 Real Time Signal Processing

Task #5	Novel VLSI Algorithms and Architectures for High Data Rate Digital Filtering.....	30
Task #6	Fault Tolerant Signal Processing Architectures	35
Task #7	Real-Time Signal Processing Systems with Multiple Functional Units	40

OMVPE GROWTH OF III-V ALLOYS AND STRUCTURES FOR NEW HIGH SPEED ELECTRON DEVICES

TASK #: 1

TASK PRINCIPAL INVESTIGATOR: J. Richard Shealy
(607) 255-4657

OBJECTIVE

The materials task for the JSEP program has several objectives. The first and major goal is to extend the crystal growth technology which has been developed in this program to allow new semiconductor structures to be prepared for use in high speed electron devices. This will require, in most cases, the pioneering of new epitaxial structures often using novel modifications of the OMVPE process. The second objective is to prepare more standard materials, such as lattice matched and pseudomorphic systems on InP and GaAs substrates, for other characterization and device fabrication studies in this task as well as the tasks of Professors Tang, Pollock and Krusius. The final objective is to develop an optical probing technique to characterize the properties of bulk and 2 dimensional electron systems.

DISCUSSION OF STATE-OF-THE-ART

Recently, the optical absorption spectrum of the commonly used organometallic species for AlGaAs growth has been determined [1]. The absorption spectra of TMG, TMA, and AsH₃ are such that little or no absorption will occur at wavelengths longer than 220 nm. As a result, a potentially very important innovation in OMVPE growth of III-V alloys is the incorporation of a deep UV laser excitation during growth to allow selective growth with a high degree of spatial resolution. This previous study utilized a broad band ArF excimer laser operating at 193 nm with 40 mJ/cm² of pulsed energy density. The pulse repetition rate was varied to control the growth rate. It was found that at substrate temperatures in excess of 500°C, good quality GaAs films could be grown in a selective fashion. The growth rate could be doubled by the absorption of this modest laser power in the gas phase. With the deep UV source, the excitation directly excites absorbed surface species (TMG-AsH₃ adducts, for example) and stimulates growth only where the laser light resides. If the diffusion of the stimulated surface species can be maintained in the dark to less than several 100 Å, a condition which would be expected at low substrate temperatures, then interference holography can be incorporated in the UV stimulation process. In contrast, the use of a visible excitation source [2] results in the absorption of light that occurs in the substrate bulk. The resultant thermal broadening or diffusion of injected carriers into the semiconductor due to the laser (either process has been proposed to explain selective growth

behavior with an argon laser source) limits the achievable line width to greater than several μm .

The conventional OMVPE process has produced many of the III-V materials and structures which find applications in high speed electron devices. The vast majority of published literature involves the AlGaAs materials system. Newly developed reactor geometries have improved the deposition uniformity to $\pm 1\%$ with good quality interfaces as demonstrated by a high mobility modulation doped heterostructure [3]. The use of the multi-chamber reaction cell [4] has demonstrated an AlGaAs/GaAs interface abruptness which approaches, and in the case of high temperature growth, surpasses that of MBE. This conclusion is based on interpretation of Raman spectra of confined optic phonon vibrations present in short period superlattices [5]. This method is subject to less interpretation than other commonly used techniques such as quantum well PL or TEM lattice images.

Wide bandgap III-V alloys, mainly the AlGaInP/GaAs system, have been studied and improved by the use of ethyl organometallics [6]. The first observation of crystal ordering was recently made in the quaternary, AlGaInP, where the ordering is similar to that observed in GaInP alloys [7]. However, the Al and Ga atoms are arranged randomly (lacking order) on a (111) plane followed by a (111) plane containing predominantly In. The ethyl sources used for the growth of AlGaInP have been shown to be advantageous for improving the impurity doping efficiency of these widegap materials. Finally, the first report of the successful use of GaInP layers as the electron confinement layer in modulation doped FETs has been reported [8]. It was shown that larger 2DEG sheet concentrations can be achieved with the GaInP/GaAs interface for the same mobility as that of the AlGaAs/GaAs case. This supports the idea that the GaInP alloy without the presence of deep donor species is a better electron supply and confinement layer to GaAs.

There have been many studies recently reported in the OMVPE growth of materials and device structures which relates to the proposed research, in fact, too many to discuss in this document. Some of the main technical advances have been highlighted. Additional reference material is collected in the most recent OMVPE conference proceedings (Hakone, Japan) which appears in 93 volume, nos. 1-4 of the *Journal of Crystal Growth*.

PROGRESS

During the present reporting period, progress on the optical probing of semiconductor materials with Raman Spectroscopy has been made; studies on the preparation of enhanced Schottky Barriers on InP have continued; and finally, the thermal stability and selective disordering of AlInAs/GaInAs has been studied. In addition, in the last year Cornell's new compound semiconductor materials laboratory has entered the construction phase and will be completed early in the summer of 1990. A novel approach to the safe handling of hydrides has been incorporated in this facility and will be briefly discussed.

i) Optical Probing of 2 Dimensional Semiconductor Structures with Raman Scattering

The first use of Raman scattering to study GaAs/AlGaAs graded index-separate confinement heterostructure (GRIN-SCH) quantum well lasers is reported. A forward scattering geometry in which the waveguide is endfired is used, and the light emerging from the opposite end facet is collected and spectrally analyzed. The probe is confined by the waveguide and thus interacts with the entire laser cavity. Because Raman scattering occurs in all regions of the heterostructure to which the optical mode is confined, this technique is a useful indication of the mode profile in waveguide heterostructures. Inhomogeneously broadened longitudinal and transverse optical phonons in the graded region are observed as well as the vibrational modes of the single 100 Å quantum well active region. Previous attempts to study single quantum wells by Raman scattering required resonant excitation at liquid He temperatures¹¹. The immediate applications of this work include characterization of single pseudomorphic thin films, which are currently of great interest for electronic and optical applications, and the study of fundamental processes in semiconductor lasers. Recently there has been some indication that photoexcited LO phonons may participate in stimulated emission in GaAs/AlGaAs quantum well heterostructures^{12,13}. Such nonequilibrium phonon distributions should be readily observable with Raman spectroscopy.

ii) Enhanced Schottky Barrier on InP ?

Recently, the UV enhancement process has been refined to obtain reproducible results. This technique is much different than that reported in the literature. First, this process occurs in an O₂ ambient. Second, the previous studies indicate that ozone is a catalysis in the enhancement process. However, these results indicate that the near UV range (300 nm and above) produces the necessary surface reactions for enhancement. The ozone producing range of 220-260 nm is not critical. Third, the growth at room temperature did not occur, instead, better results are observed around the congruent sublimation temperature of 350°C. Under these conditions the best enhancement achieved without annealing was 0.7 V. However, upon annealing and the subsequent characterization of this device, MOS like characteristics appeared. The series resistance jumped from 1kΩ to 9 MΩ. The forward and reverse biased IV data demonstrated space charge limited current flow. This insulating behavior disappeared after three days of storage in air. This stability problem will be addressed by encapsulating the device with Si₃N₄.

The insulating behavior of the most recent devices brings into question Iladis's claim of enhanced Schottky Barriers⁹. The published data on the electrical behavior of his devices has a number of shortcomings. First, he only

illustrates forward biased IV data with no mention of the behavior of the series resistance as the barrier height is increased. Second, no C-V data is provided. Accumulation of charge in the forward biased region would indicate whether his devices are MOS like in character. If P_2O_5 is producing the enhancement, the long term survival of these devices is questionable. Without encapsulation, P_2O_5 is known to be hygroscopic.

iii) Thermal Stability and Selective Disordering of AlInAs/GaInAs

Substantial blue shifts in the transition energies of GaInAs/AlInAs single quantum wells were observed due to localized SiO_2 capping and rapid thermal annealing at temperatures between 750 and 900°C. In contrast to previously reported results¹⁴, regions capped with SiO_2 exhibited blue shifts up to 74 meV while regions with no SiO_2 showed minimal shifting. With this bandgap change, a lateral index change of approximately -0.6% is anticipated making this process suitable for index guided lasers. Samples also exhibited up to 15-fold increases in PL efficiencies due to the annealing process. The dependence of energy shifts and PL efficiencies are studied by measuring room-temperature and low-temperature photoluminescence. These Results show that this materials system has comparable thermal stability as the AlGaAs/GaAs system.

iv) Secondary Containment of Hazardous Gases Used In OMVPE

Cornell's new OMVPE laboratory is the first facility to embody a secondary containment system for the hazardous gases in the growth of III-V alloys, arsine and phosphine. Currently, most all facilities operating in the U.S. incorporate flow limiting orifices in these cylinders to limit the spill hazard. However, since the orifices are located down stream of the cylinder valve, a catastrophic cylinder failure would likely result in personnel injury and perhaps a fatality as some 300,000 CFM's of exhaust is required to remove such as spill (beyond the limit of practicality). This exhaust flow will dilute the volume of arsine gas escaping from the cylinder to the 1/2 IDLH value prior to its escaping into the environment. Concentration in excess of this value will result in symptoms after seconds of exposure. Such an event would likely cause the shut down of many DoD sponsored research programs.

The secondary containment dry box enclosure will capture the hazardous gas release during a worst case scenario and introduce the gas into an incinerator capable of removing very large concentrations of arsine and phosphine which appear at its inlet. Both the personnel and the environment are protected. Such an installation could be examined by DoD, and if found acceptable, it could be phased into other research and manufacturing operations. This will represent a move to significantly lower the risk which OMVPE processing pose, thereby, accelerating its use in the manufacture of advanced semiconductor devices.

POTENTIAL SCIENTIFIC IMPACT OF RESEARCH

The potential impact of the proposed research is twofold. First, through collaborations with task (2-4), the discovery of new semiconductor structures for improved performance of high speed electron devices will result. This will include the use of wide bandgap electron confinement structures and novel structures on InP to take advantage of its intrinsic electron transport properties. Furthermore, with the successful completion of submicron selective OMVPE growth of III-V alloys, the first practical technology for producing quantum wire devices will emerge. Secondly, pioneering a new scattering geometry for electronic Raman spectroscopy, and the subsequent examination of 2 dimensional electron systems, will allow a non-destructive optical probe to overlap electron channels in devices under operating conditions. The insight gained from such measurements will likely lead to new epitaxial structures and device geometries for improved 2 dimensional electron gas transport properties. Presently, this technique has been used to observe for the first time, the room temperature vibrational spectra of single quantum wells with a passive probe.

DEGREES

None.

REFERENCES

- [1] "UV Laser Stimulated MOVPE of GaAs," P. Balk, Third Biennial OMVPE Workshop Abstracts, Brewster, MA (1987).
- [2] "Low Temperature Selective Epitaxy of III-V Compounds by Laser Assisted Chemical Vapor Deposition," N. H. Karam, H. Liu, I. Yoshida, B.-L. Jiang, and S. M. Bedair, *Journal of Crystal Growth*, 93, 254-258 (1988).
- [3] "A New Versatile, Large Size MOVPE Reactor," P. M. Frijlink, *Journal of Crystal Growth*, 93, 207-215, (1988).
- [4] "Characterization of AlGaAs/GaAs Heterostructures Grown in a Multi-chamber OMVPE Apparatus," J. R. Shealy, *Journal of Crystal Growth*, 87, 350-356 (1988).
- [5] "Raman Scattering Characterization of Interface Broadening in GaAs/ AlAs Short Period Superlattices Grown by Molecular Beam Epitaxy," B. Jusserand, F. Alexandre, D. Paquet, and G. Le Roux, *Applied Physics Letters*, 47 (3) (August 1985).

- [6] "MOCVD Growth of AlGaInP at Atmospheric Pressure Using Triethylmetals and Phosphine," M. Ikeda, K. Nakano, Y. Mori, K. Kaneko, and N. Watanabe, *Journal of Crystal Growth*, 77, 380-385 (1986).
- [7] "Growth Temperature Dependent Atomic Arrangements and their Role on Band-Gap of InGaAlP Alloys Grown by MOCVD," C. Nozaki, Y. Ohba, H. Sugawara, S. Yasuami, and T. Nakanisi, *Journal of Crystal Growth*, 93, 406-411 (1988).
- [8] "Uniform and Abrupt InGaP/GaAs Selectively Doped Heterostructures Grown by MOVPE for HEMT ICs," T. Ohori, M. Takechi, M. Suzuki, M. Takikawa, and J. Komeno, *Journal of Crystal Growth*, 93, 905-910 (1988).
- [9] "Schottky Barrier Height Enhancement in n-InP for MESFET Applications," A.A. Iliadis, *Inst. Phys. Conf. Ser. No. 96*: Chapter 6, 413-416 (1988).
- [10] "Schottky Diode and Field-Effect Transistor on InP," S. Loualiche, A. Ginoudi, H. L'Haridon, M. Salvi, A. Le Corre, D. Lecrosnier, and P. N. Favenec, *Appl. Phys. Lett.*, 54 (13), 1238-1240, 27 March 1989.
- [11] "Confined Optical Phonons in a GaAs Single Quantum Well in a GaAs/Al_xGa_{1-x}As Heterostructure," A. Arora, E. K. Suh, A. K. Ramdas, F. A. Chambers and A. L. Moretti, *Phys. Rev. B*, 36, 6142 (1987).
- [12] "Observation of Phonon-Assisted Laser Operation of Al_xGa_{1-x}As-GaAs Quantum Well Heterostructures," N. Holonyak, Jr., D. W. Nam, W. E. Plano, E. J. Vesely and K. C. Hsieh, *Appl. Phys. Lett.*, 54, 1022 (1989).
- [13] "Phonon-Assisted Recombination in GaAs/AlGaAs Multiple-Quantum-Well Structures," H. H. Dai, M. S. Choi, and M. A. Gunderson, H. C. Lee, P. D. Dapkus and C.W. Myles, *Appl. Phys. Lett.*, 66, 2538 (1989).
- [14] "Modification of Exciton Emission Energies in InGaAs/AlInAs Quantum Wells by Rapid Thermal Annealing," J. Y. Chi, E. S. Koteles, and R.P. Holmstrom, *Appl. Phys. Lett.*, 53, 2185-87 (1988).

IJSEP PUBLICATIONS

- 1. "Characterization of GaAs/AlGaAs GRIN-SCH Lasers by Raman Scattering," J. Bradshaw and J. R. Shealy, *J. Appl. Phys.* (May 1989).
- 2. "Plasmon-LO Phonon Interaction in n-type (Al_{0.3}Ga_{0.3}As)_n(In_{0.3}Ga_{0.3}As)_m Short Period Strained Layer Superlattices," Ostergaard, O. Leistiko, X. Song, J. R. Shealy, W. J. Schaff, L. F. Eastman and K. Yamasaki, submitted to *J. Appl. Phys.* (Feb. 1990).

3. "Effects of Rapid Thermal Annealing and SiO₂ Encapsulation on GaInAs/AlInAs Heterostructures," S. O'Brien, J. R. Shealy, D. P. Bour, L. Elbaum and J. Y. Chi, *Appl. Phys. Lett.* (April 1990).

FEMTOSECOND LASER STUDIES OF ULTRAFAST PROCESSES IN COMPOUND SEMICONDUCTORS

TASK #: 2

TASK PRINCIPAL INVESTIGATOR: C. L. Tang
(607) 255-5120

OBJECTIVE

The basic objective of this task is to study the dynamics of nonequilibrium electrons and holes in compound semiconductors and related structures using recently developed femtosecond laser sources and measurement techniques. A major breakthrough achieved in the current year that will have a significant impact on the study of the ultrafast dynamics of hot carriers in semiconductors is the development of a broadly tunable femtosecond laser source and its improvements over the initial results. This source and the uv femtosecond laser source developed in our laboratory earlier will be used to study the dynamics of, for example, capture of nonequilibrium carriers into quantum wells, tunneling between quantum wells, coherent electron wavepacket excitation and relaxation, and hole relaxation in compound semiconductors.

DISCUSSION OF STATE-OF-THE ART

Femtosecond laser sources and measurement techniques have made it possible to study the relaxation dynamics of nonequilibrium carriers in compound semiconductors directly in the time domain for the first time. Until recently, the accessible wavelength range was, however, limited to basically the operating range of the Rh6G/DODCI dye laser, or approximately 630 nm. Nonetheless, a great deal of useful information on the relaxation dynamics of III-V compounds has been obtained for the first time using such a laser. Much remains to be done, however. To make further progress, the accessible wavelength range must be extended.

The usual approach to extending the available femtosecond wavelength range has been either to search for new dye combinations or through femtosecond continuum generation. Despite extensive efforts at many laboratories, few dye combinations adequate for femtosecond laser applications have been found. Dye femtosecond laser sources, most of them not nearly as good as the Rh6G/DODCI laser, are available only in a few very narrow wavelength ranges in the visible. In the case of continuum generation, because of the need to amplify the initial femtosecond laser pulses, the repetition rate of the continuum generated is generally more than five orders of magnitude down from that usually available in Rh6G dye lasers and the time resolution is typically also degraded from 25 femtoseconds down to several hundred femtoseconds.

Recent developments in Professor Pollock's laboratory and in our (Tang's) laboratory have led to the first truly broadly tunable femtosecond laser sources in the infrared and the first extension of the femtosecond sources into the uv. Pollock's source is based on color-center lasers and is tunable from 1.4 to 1.8 μ . Our source is based upon the optical parametric oscillator in the ir and is tunable from 700 nm to 4.5 μ at 10^8 Hz rate, but the power level is in the mW range and lower than the color center lasers. The uv source is based upon the intra-cavity second harmonic generation technique using the new nonlinear optical crystal, b-barium borate, grown and fabricated in our laboratory. This led to nearly 100% conversion of the 630 nm fundamental light to 315 nm at the same pulse repetition rate. Thus, a uv femtosecond source with comparable characteristics as those of the Rh6G laser is now available for the first time. Combined with other dye lasers, this technique should extend the accessible wavelength range in the uv down to approximately 240 nm. Together with the femtosecond optical parametric oscillators, tunable femtosecond sources are now available for studying ultrafast processes from 240 nm to 4.5 μ . This vastly extends the accessible wavelength range for studying ultrafast dynamic processes.

There has been extensive work on the relaxation dynamics of electrons in GaAs and related materials and structures. The dynamics of the holes are still far from understood, however. There have been some recent studies addressing this issue, although the picture is still far from clear. The main difficulty is that the hole relaxation process is expected to be even faster. With the femtosecond lasers restricted to near 2 eV, the holes generated are very near the top of the valence band and relax quickly. To study the hole dynamics optically, the holes must be created far from the zone center. This means femtosecond pulses at shorter wavelengths than 630 nm are needed. With the new uv source we have developed, this will be possible for the first time. This is one area that we plan to investigate.

In addition to bulk GaAs, the relaxation dynamics in other important materials such as GaInP, GaInAsP, etc. can all now be studied for the first time with the new ir femtosecond laser sources developed at Cornell. In fact, IV-IV compounds such as Si can also be studied with our new uv source. In terms of ultrafast dynamics in these materials, very little is known. We expect to look at some of these in the next grant period. We will probably begin with GaInP, since samples of this material are available from J. R. Shealy's group.

More important than the bulk materials are structures such as quantum wells, superlattices, and tunneling structures. Although some preliminary results have been obtained on simple GaAs/AlGaAs quantum wells and superlattices, because of the limitations of the available source wavelengths, these structures have hardly been explored and much needs to be done. With the new ir femtosecond sources down to 4.5 μ , we will have an opportunity to study for the first time optical transitions between the quantum well states on a picosecond time scale. Also, the effects of applied electric field on such structures can be studied with the help of such sources.

There have also been some recent studies on the question of tunneling time using picosecond lasers. The results are very crude and nonquantitative. With wavelength tunability and less than 100 fs time resolution, we should be able to make substantially better measurements of the tunneling time in different materials and structures than heretofore possible. This is another area that we expect to be able to make a unique contribution during the next grant period. Availability of suitable tunneling structures is important, however.

PROGRESS

The major break-through this year is the development of the first truly broadly tunable femtosecond laser source from the deep red to mid-ir: a singly resonant optical parametric oscillator based on a thin crystal of KTiOPO_4 is pumped by intracavity femtosecond pulses at 620 nm from a visible femtosecond laser. Oscillation results in stable continuous outputs of femtosecond pulses at 10^8 Hz repetition rate and milliwatt average power levels in both signal and idler beams. Tuning from 820 - 920 nm and 1.90 - 2.54 μM with a sing set of mirrors has been demonstrated. With multiple sets of mirrors, continuously tunable outputs from ~ 0.72 to ~ 4.5 m μM should be possible. The pulse width obtained initially was around 200 fs in the visible. With the addition of an intracavity 4-prism sequence to compensate for group dispersion in the nonlinear KTP crystal, the pulse width is now reduced down to around 100 fs at 840 nm. To the device to work the first time was extremely difficult and time consuming. Fortunately, we are now over this initial stage and the device is now working routinely in the laboratory and is being used for experimental studies.

Preliminary data on bulk GaInAs have been obtained using the tunable femtosecond optical parametric oscillator for the first time. We have also obtained preliminary data on the capture of hot carriers by quantum wells and their subsequent relaxation using the time-resolved hot-luminescence spectroscopic technique. Comprehensive studies of these problems are being carried out.

POTENTIAL SCIENTIFIC IMPACT OF RESEARCH

A clear understanding of the dynamics of highly excited nonequilibrium carriers in semiconductors are of basic importance to solid-state physics and high-speed electronic and optoelectronic devices. The proposed program is aimed at providing the needed information through optical studies based upon femtosecond lasers and measurement techniques. The program is expected to provide not only basic material parameters important for designing and understanding the behavior of high-speed devices and but also new femtosecond laser sources and techniques that might used for a wide range of material and device studies.

DEGREES

1. F. W. Wise
Ph.D., Applied & Engineering Physics, August 1989
"Femtosecond Studies of Hot-Electron Relaxation in Semiconductors"
2. D. Edelstein
Ph.D., Applied & Engineering Physics, October 1989
"New Sources and Techniques for Ultrafast Laser Spectroscopy"

REFERENCES

None.

ISEP PUBLICATIONS

1. "Theory of Quantum Beats in Optical Transmission-Correlation and Pump-Probe Experiments for a General Raman Configuration," I. A. Walmsley, M. Mitsunaga, and C. L. Tang, *Phys. Rev. A*, 4681-4689 (November 1988).
2. "Femtosecond Laser Studies of the Relaxation Dynamics of Semiconductors and Large Molecules," C. L. Tang, F. W. Wise, M. J. Rosker, and I. A. Walmsley, invited review, *IBM Journal of Research and Development*, 33, 447-454 (July 1989).
3. "Subpicosecond Luminescence Study of Hot-Electron Relaxation in GaAs Quantum Wells," F. W. Wise and C. L. Tang, *Sol. State Comm.* 69, 821-826 (1989).
4. "Broadly Tunable High Repetition Rate Femtosecond Optical Parametric Oscillator," D. D. Edelstein, E. Wachman, and C. L. Tang, *Appl. Phys. Lett.* 54, 1728-1730 (May 1, 1989); also, Post Deadline paper, CLEO '89, April 27, 1989, Baltimore, MD.
5. "Ultrashort Optical Pulses," invited review, *Encyclopedia of Phys.*, G. L. Trigg, ed. (VCH Publishers, Inc., NY).
6. "Continuous-Wave Mode-Locked and Dispersion-Compensated Femtosecond Optical Parametric Oscillator," E. S. Wachman, D. C. Edelstein, and C. L. Tang, *Optics Letters*, 15, 136-138 (January 15, 1990).

ULTRAFAST STUDIES OF DEVICE STRUCTURES AND OF HOT CARRIERS IN NARROW BANDGAP SEMICONDUCTORS

TASK #: 3

TASK PRINCIPAL INVESTIGATOR: C. R. Pollock
(607) 255-5032

OBJECTIVE

The temporal relaxation of hot carriers in narrow bandgap bulk semiconductors is being studied using infrared pulses of 60-200 fsec duration. The femtosecond pulses are obtained from a color center laser, tunable from 0.7 to 0.85 eV. This wavelength range is directly useful for GaInAs based materials. Tunability with femtosecond resolution provides the unique ability to measure relaxation rates of carriers lying between the bottom of the conduction band up to approximately the L and X valleys in certain semiconductors. Measurements of both energy and momentum relaxation are being done in bulk GaInAs as a function of probe energy.

DISCUSSION OF STATE-OF-THE-ART

There has been widespread application of femtosecond pulses to the study of ultrafast phenomena in semiconductors. However most femtosecond spectroscopy has been done using fixed energy 2 eV photons on materials with relatively large bandgaps ($E_g > 1.4$ eV). Extensive work has been done in the measurement of the scattering rates of carriers in GaAs, AlGaAs, and quantum well structures using these materials. Tang (Cornell)¹, Ippen (MIT)², and Knox (Bell Labs)³ have all pioneered techniques, instrumentation, and measurements on these samples. The current ability of optical probing is well established for determining both the rate at which carriers relax from the initial state, and the mechanisms (such as intervalley scattering) responsible for the short-lived carrier distributions in GaAs/AlGaAs. There has not been as much work on direct measurement of specific devices or structures with the aim of improving the mobility of the carriers in the material; most work to date has primarily focussed on determining the dynamics of the carrier relaxation within a given structure.

To date, because of the lack of femtosecond sources in the 0.8 eV photon energy range, little work has been done on the GaInAs system. Chemla (Bell Labs)⁴ has recently concluded a study of exciton absorption in GaInAs quantum wells using 200 fsec resolution pulses. His work concentrated only on exciton interactions with the lattice and free carriers, and confirms earlier theory about the scaling of binding energy with bandgap and well dimension. There was no discussion of hot carrier dynamics.

Using magnetotransport measurements, Barlow et. al. (University of Essex, UK)⁵ measured the rate at which electrons cool down to the lattice temperature, and found picosecond times for this process. No temporal measurements were made directly. In a similar study, Kash et. al (Bell Labs)⁶ used luminescence to measure the lifetime of hot electrons in GaInAs as they also cooled to the lattice temperature over a 10 psec period. To date, there are no published reports of femtosecond studies of the initial carrier scattering in GaInAs.

PROGRESS

We have been working for almost two years under JSEP support. To date, we have developed our NaCl laser source to the point where it can deliver 75 fsec pulses over the tuning range from 1.47 to 1.75 μm , which is an ideal range for studying GaInAs⁷; we have frequency doubled this source to provide tunable 70 fsec pulses in the 750-850 nm range⁸ for use in collaborative studies with other JSEP members; and we have begun an extensive set of measurements on several samples of GaInAs⁹. The direct output of the NaCl femtosecond laser has been used to study the temporal relaxation rates of hot carriers in GaInAs bulk material. In our first measurements, we probed the excited carrier lifetimes in GaInAs/InP bulk material (the GaInAs was approximately 4 μm thick) provided by the Cornell Semiconductor Material Growth facility. The probe pulses were tuned from 1.675 μm (which corresponds to photons with barely enough energy to promote an electron across the bandgap) to 1.53 μm , which corresponds to photon energy that is about 70 meV above the band gap (70 meV is roughly equal to the energy in two LO phonons). Results have not yet been fully analyzed or published, but the data shows clear trends. Fig. 1 at the end of this section shows the raw data from the two-pulse correlation experiments. The lower trace shows the nonlinear transmission of a sample excited with 1.53 μm photons (sufficient energy to place the carriers about 70 meV above the conduction band minimum). The sample transmission recovers within about 200 fsec, indicating the time it takes for excited carriers to scatter out of their initial excited states. The upper trace shows the transmission of the sample when pumped by photons with energy near the bandgap energy. The relaxation time becomes noticeably longer, displaying long-lived tails. This response is consistent with a model of carrier scattering due to LO phonons: at low excitation energy there is not enough excess energy for the carriers to relax by LO phonon emission, so they cannot rapidly cool. The laser and instrumentation to the point where our signal-to-noise ratio in the wings of the relaxation signal is on the order of 60 dB.

The data we obtained has being shared with Prof. Krusius' group, who are using it to compare with numeric simulations of the carrier dynamics for this material. Unfortunately, the sample thickness that we chose for our first experiments was too thick. The 5 μm layer attenuated the measured transmitted signal by about 10 dB, leading to a strong spatially dependent

excitation population, and also a spatially varying nonlinear response to the probe light. This has complicated the numerical simulation efforts of Prof. Krusius' group. A spatially inhomogeneous formulation had to be developed (see task 4).

Dr. Bill Schaff recently provided us with a new thin sample of GaInAs (0.5 μm thick) on an InP substrate, and we are presently making a series of measurements on this new sample. We anticipate that by the end of April, 1990, we will have essentially characterized the relaxation dynamics of bulk GaInAs as a function of carrier energy near the conduction band minimum as a function of carrier concentration and as a function of temperature. Portions of this work will be presented in March, 1990 at the San Diego SPIE conference SPIE meeting: Advances in Semiconductors and Superconductors: Physics Toward Device Applications (Ultrafast Processes) in San Diego, CA⁹.

To date we have used the frequency doubled source in collaboration with Dr. Paul Tasker of Prof. Eastman's group to study the response speed of phototransistors, and with Dr. Bill Grande of Prof. Tang's group to study the switching speed of his novel OEIC switches. In the former work, we used direct femtosecond pulses to excite a GaAs MODFET phototransistor, we were able to directly determine the switching speed of the device (turn-on time was 50 psec, turn-off time was 100 psec) without any problem of deconvolving the measured response with comparably long electronic or optical pulses. The experiment provided clean and exact information about the device. The work with Tang's group was not successful, because the wavelength of the pulses we could generate did not overlap the gain bandwidth of the devices under test.

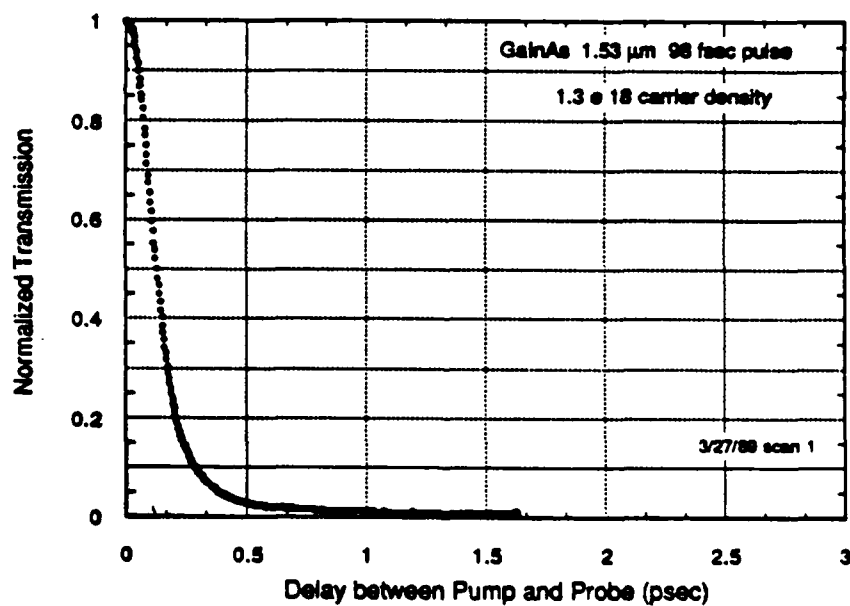
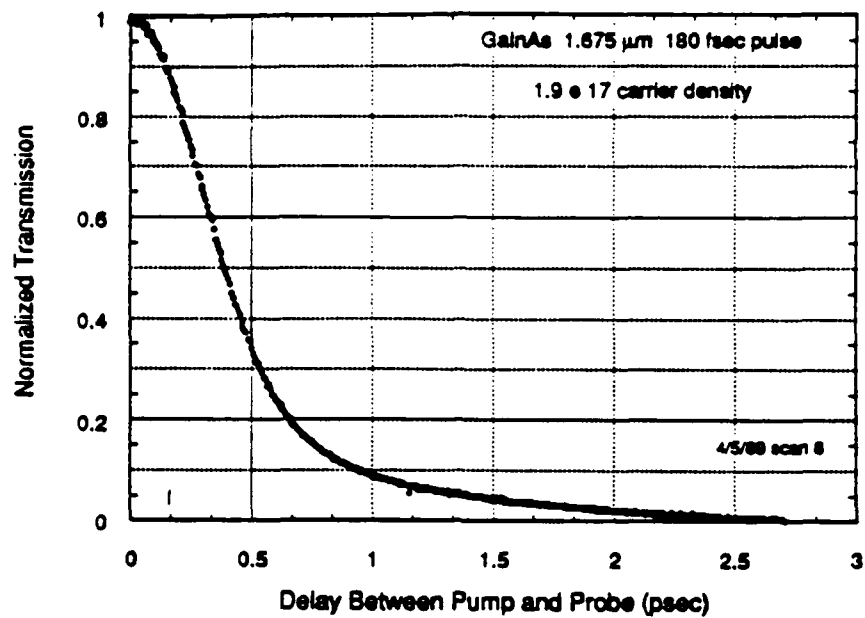


Figure 1. Normalized optical transmission as a function of delay between pump and probe pulses. Upper and lower figures for carrier densities of $1.9 \times 10^{17} \text{ cm}^{-3}$ and $1.3 \times 10^{18} \text{ cm}^{-3}$.

POTENTIAL SCIENTIFIC IMPACT OF RESEARCH

The major goal of this research is to gain an understanding of the carrier relaxation rate in InGaAs and InGaAs/InP quantum wells. Since this work is being done on an essentially new material from the femtosecond spectroscopy point-of-view, we expect at the bare minimum that an improved understanding of the physics of alloyed III-V semiconductors will be developed. Hopefully, these measurements will lead to faster electronic devices, and faster optical sources and modulators. GaInAs is a relatively new material which has a relatively high electron mobility. The bandgap of GaInAs alloys is ideally suited for present optical communication systems. Direct measurements of the ultrafast behaviour of material properties and devices based on this material should have a strong impact on the engineering and design of future optoelectronic and electronic devices.

DEGREES

1. Christopher P. Yakymyshyn,
Ph.D, Electrical Engineering, 1989
"Development and Application of the Additive Pulse Modelocked NaCl Laser".
2. Efstratious T. Georgiou
Ph.D, Physics, 1989
"Formation and Optical Properties of New Laser-Active Color Center Lasers in Sodium and Potassium Chloride".
3. Charles F. Culpepper
PhD., Electrical Engineering, 1989
"Cryogenic and Room Temperature Operation of a Pulsed NaCl Color Center Laser".

REFERENCES

- [1] C. L. Tang, F. W. Wise, and I. A. Walmsley, *Solid State Electronics*, 31, 439-442, (1988).
- [2] W-Z Lin, R. W. Schoenlein, J. G. Fujimoto, and E. P. Ippen, *IEEE JQE* 24, 267-275 (1988).
- [3] W. H. Knox, D. S. Chemla, G. Livescu, J. E. Cunningham, and J. E. Henry, *Phys. Rev. Lett.*, 61, 1290-1293 (1988).

- [4] "Femtosecond Dynamics of Exciton Absorption in the Infrared InGaAs Quantum Wells," M. Wegener, I. Bar-Joseph, G. Sucha, M. N. Islam, N. Sauer, T. Y. Chang, and D. S. Chemla, submitted for publication.
- [5] M. J. Barlow, B. K. Ridley, M. J. Kane, and S. J. Bass, *Solid State Electronics* 31, 501-505 (1988).
- [6] K. Kash and J. Shah, *Jour. Luminescence*, 30, 333-339 (1985).
- [7] C. P. Yakymyshyn, J. F. Pinto, and C. R. Pollock, *Optics Letters*, 14, 621-623 (1989).
- [8] C. P. Yakymyshyn, and C. R. Pollock, *Optics Letters*, 14, 791-793 (1989).
- [9] SPIE Conference "Advances in Semiconductors and Superconductors: Physics Toward Device Applications," 17-21 March 1990, San Diego, California (Paper 1282-15)

ISEP PUBLICATIONS

- 1. "Frequency Doubled Additive Pulse Modelocked NaCl Laser," C. P. Yakymyshyn, and C. R. Pollock, *Optics Letters*, 14, 791-793 (1989)
- 2. "Dispersion Effects in Coupled Cavity Lasers," B. J. Zook, C. P. Yakymyshyn, and C. R. Pollock, submitted to *Journal of Opt. Soc. Am. B*, (February 1990)

PHYSICS OF HIGH SPEED MESO SCALE COMPOUND SEMICONDUCTOR DEVICES

TASK #: 4

TASK PRINCIPAL INVESTIGATOR: J. P. Krusius
(607) 255-3401

OBJECTIVE

The objective in this task is to explore the physics of hot carrier transport in small inhomogeneous ultra high speed compound semiconductor heterostructures. Interactions with thermodynamically open boundaries, graded material composition with imbedded heterojunctions, two-dimensional space charge phenomena, optical fields, and steady state and transient conditions are considered. Specific transport issues to be pursued are: ballistic carrier transport across heterojunctions, two carrier transport under high density and recombination conditions, and two carrier processes under optical interactions with femtosecond laser probes.

DISCUSSION OF STATE-OF-THE-ART

Transport and optical processes of carriers in compound semiconductor materials have been studied intensively for the past two decades. Carrier processes for both electrons and holes in the non-interacting quasi-equilibrium limit are well understood within the framework of linear response, and extensive work on hot non-equilibrium carriers is being performed. Hot carrier research has primarily been focused on hot electrons because of their importance for high speed semiconductor devices. A variety of transport and optical methods have been used to probe physical processes influencing hot electron behavior. These include electron interactions with phonons, impurities, defects, photons, device boundaries, and external electro-magnetic fields. Transport studies in bulk materials and small device structures, and in particular recent picosecond and femtosecond optical probing techniques, have helped to quantify the physical processes determining hot electron behavior. It is fair to state that the understanding of hot electron behavior on all but the shortest time and the smallest spatial scales is approaching maturity.

A significant part of the hot carrier processes, namely those involving both electrons and holes, and their interactions under non-equilibrium conditions, has received little attention. Recent experimental and theoretical indicators are pointing to the importance of electron-hole interactions and hole processes in the behavior of hot carriers. A few examples illustrate this statement. Optically generated hot electron and hole distributions have been shown to thermalize on drastically different time scales with the distribution function itself influencing thermalization dynamics [1]. Transport of minority

carriers in dense semiconductor plasmas has been demonstrated to be so strongly affected by the electron-hole interaction that negative minority carrier mobilities have been measured (electron-hole drag) [2]. Minority carrier velocity-field characteristics for minority electrons in p-type doped $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ have not shown evidence of the transferred electron effect, which reduces the average drift velocities of electrons in all III-V compound semiconductors for higher electric fields [3]. In a recent theoretical study the electron-hole interaction was found to become one of the primary energy loss mechanisms for carrier thermalization in GaAs at high carrier concentrations [4]. While experiments on hot electron thermalization using the pump/probe technique developed at Cornell by Tang [5] have in the past been analyzed neglecting the contribution of holes [6], such approximations do not seem to be justified in the general case. Much longer relaxation times were recently measured by an IBM group in an energy dependent cw luminescence study [7]. These latter time constants appear incompatible with previous pump and probe measurements and their subsequent theoretical analysis [5,6]. This discrepancy has to date remained unresolved.

Dual carrier processes also determine the characteristics of a number of important electronic and optoelectronic devices, such as: heterostructure bipolar transistor, photodiode, phototransistor, and semiconductor laser. Heterostructure bipolar devices have shown considerable potential for high speed gate array type applications, where their superior current drive capability can be exploited. However, the analysis of their operation is presently limited either to a hydrodynamic model for both electrons and holes [8], or hot carrier particle formulation for electrons and a hydrodynamic model for holes [9,10] without the consideration of important two-dimensional phenomena and associated space charges. Optoelectronic devices have recently become increasingly more important because of their applications in long distance fiber optic communication, mixed electronic and opto-electronic systems, and potentially also in optical computing. While the recent literature on these devices and their applications is too voluminous to be quoted here, their operation, design, and limitations cannot be fully understood until non-equilibrium dual carrier transport and optical interactions are explored in inhomogeneous device structures on a femtosecond time scale.

PROGRESS

Research during the second year of this two year program has focused on three non-equilibrium carrier problems: (1) carrier transport across graded heterojunctions, (2) femtosecond carrier thermalization processes in narrow band gap heterostructures, and (3) dual carrier transport.

1. Non-equilibrium Carrier Transport Across Graded Heterojunction

Hot electron injection across graded and abrupt III-V compound semiconductor heterostructures has been explored using a self-consistent time-dependent ensemble Monte Carlo transport formulation. Electron bands are

described within a position dependent $k.p$ framework in combination with the virtual crystal approximation to account for pseudobinary alloy effects. Scattering processes include intra and inter valley phonons (optical and acoustic), ionized impurities (Ridley screening), and the alloy effect (chemical disorder, Harrison and Hauser formulation). Scattering rates were calculated within the $k.p$ theory including all overlap integrals. This transport formulation, described in detail in an earlier JSEP publications (see Ref. [11, 12]), has been implemented in a two-dimensional time-dependent computer code for non-equilibrium electron injection studies. This code, 2D-TCMC, allows one to simulate the behavior of non-equilibrium electrons in a rectangular domain comprised of several compound semiconductor regions including compositional grading and imbedded abrupt or graded heterojunctions. Ohmic and Schottky contacts can be placed anywhere on the periphery of the rectangular domain and with a small amount of rework also into the interior. Ohmic contacts are described via an interaction with external thermal reservoirs using microscopic injection statistics. Particle conservation is not explicitly enforced. Two-dimensional space charges are fully included. Charge assignment is performed using the cloud-in-cell method and Poisson's equation for the ensemble is solved using Hockney's Fast Fourier Transform based technique.

Hot electron transport processes across laterally uniform one-dimensional and laterally non-uniform two-dimensional graded heterojunctions have been explored. During the previous reporting period it was found that ensemble phenomena dominate the carrier injection process across the heterojunctions influencing distribution functions, drift velocities, and ballistic carrier fractions. Space charges and current continuity play a crucial role. In a study of hot electron injection in the $Al_xGa_{1-x}As/GaAs$ materials system across one-dimensional heterojunctions it was established that drift velocities downstream from the heterojunction can vary by a factor of four depending on the state of the space charge at the junction. The injection efficiency can be influenced via grading, doping density, temperature, and applied voltage. Flat band conditions desirable for high current drive device applications are not attainable for any applied voltages, if the injecting junction has not been correctly designed, an important finding for high speed device design. Lateral electrodes can be used to shape the space charge at the heterojunction in two-dimensional device structures. Lateral control electrode placement can influence the device current almost by an order of magnitude for a similar current modulation capability. Two-dimensional space charge phenomena have been studied with the vertical FET (VFET) taken as the generic example. Both steady state and transient ballistic carrier transport across heterojunctions has been shown to be controlled by lateral space charges (JSEP publication 1). Steady state and transient distribution functions have been calculated and analyzed (JSEP publication 2). Measurable high frequency device parameters have been determined and correlated with experiment (JSEP publication 3). Current continuity and associated space charges were observed to

have a profound effect both on steady state and transient switching characteristics.

Based on these results it is now possible to understand hot electron injection in a variety of multi-dimensional compound semiconductor devices. Optimization issues for a class of heterostructure unipolar devices, including the VFET, the permeable base transistor (PBT) and the planar doped barrier transistor (PDBT) in progress. Extensive simulations are performed both on engineering workstations and the IBM 3090-600E supercomputer (JSEP publications 4 and 5). A cpu time speed ratio of about 5x is observed with non-optimized scalar code compared to an HP 370 engineering workstation. No efforts have been made to date to utilize the vector capability of the 3090.

2. Femtosecond Carrier Thermalization

Femtosecond carrier thermalization is being explored in collaboration with Pollock's research group focusing on the narrow gap $\text{Ga}_x\text{In}_{1-x}\text{As}/\text{InP}$ heterostructure system. Femtosecond pump/probe experiments with the unique tunable color center laser are designed jointly. Pollock's group is performing the femtosecond measurements, while theory and data analysis is performed within this task. This collaboration allows experiment and theory to interact during all stages of the research and thus maximize the yield of results. Initial experiments on $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{InP}$ films are currently in progress (see task #3).

The physics of hot carrier thermalization in these thin films is described with a dual carrier self-consistent ensemble Monte Carlo formulation including electrons and heavy and light holes. Initially only homogeneous films without spatial inhomogeneities were considered. However, experiments are often performed on films thicker than the absorption length, which results in significant variation of the optical intensity across the film (see task #3). After some initial correlations we found that neither the homogeneous film, nor a multi-layer homogeneous slab model, could explain measured absorption curves. Currently a fully self-consistent one-dimensional formulation is employed. It facilitates the true description of the spatial dependence of the optical fields and carrier space charges with associated plasma effects. Electron and hole bands are again calculated within the $\mathbf{k}\cdot\mathbf{p}$ formulation with corrections for higher bands included through second order perturbation theory. Due to the multiplicity and warping these expression are quite involved. Both acoustic and optical phonons are included inelastically through the long range (polar, piezoelectric) and short range (deformation potential) interactions. Electron-electron, hole-hole and electron-hole scattering are included, but no scattering phenomena leading to direct transitions from the valence bands to conduction bands, or vice versa are considered. Scattering between valence bands is fully included. All interactions are statically screened. For the $\text{Ga}_x\text{In}_{1-x}\text{As}/\text{InP}$ and many other pseudobinary systems it is necessary to include both chemical and structural disorder, when treating alloy scattering because of local anion site related bond length and angle distortions. This can be accomplished using the

molecular coherent potential approximation [13]. Optical interactions are included within first order time dependent perturbation theory with full inclusion of photon polarization effects. The inhomogeneous one-dimensional self-consistent formulation has been completed and implemented in computer code for the simulation of the tunable pump/probe experiment including photon polarization effects. The development of this code leveraged significant parts from the two-dimensional non-equilibrium electron transport code 2D-TCMC (see part 1).

Simulations of the basic pump-probe experiment are in progress. Preliminary results correlate well with measured data even on thick samples. From our initial results it appears that both holes and plasma effects play a significant role in these experiments (JSEP publication 6). We now feel that the condensation of the carrier thermalization dynamics into a set of effective exponential time constants, an approach used to extract scattering rates from measured data earlier by a number of researchers, is questionable. The synergistic experimental and theoretical approach pursued here is needed to correctly interpret the dual carrier thermalization dynamics in pump/probe studies. Efforts are also under way to apply the formulation and code developed here to other materials systems.

3. Dual Carrier Transport

A two-dimensional dual carrier transport formulation with electrons and heavy and light holes has been completed by leveraging past work on unipolar transport and femtosecond carrier thermalization. A computer code implementing this formulation is in progress. The previously used Poisson solver based on fast Fourier transform techniques is being replaced in order to relax the restrictive conditions on the mesh and the boundary conditions. Although preliminary feasibility simulations have been performed, this code is currently in an extensive testing phase. It is expected that the first HBT device simulations can be performed in the June/July timeframe.

POTENTIAL SCIENTIFIC IMPACT OF RESEARCH

In order to fully understand femtosecond electronic and optical processes in compound semiconductors it is necessary to examine non-equilibrium electrons and holes simultaneously. The dual carrier ensemble particle methods developed in this task will allow us to analyze these processes in full detail in realistic inhomogeneous heterostructures. Although these methods will be extremely complex, and take considerable amount of time to develop, they are necessary for the unambiguous interpretation of the femtosecond laser measurement of carrier processes in compound semiconductor heterostructures. Once these methods have been developed, correlations with optical and transport measurements performed, and their validity established,

they can be applied to the analysis, design and optimization of a large number of ultrafast electronic and optoelectronic devices.

DEGREES

None.

REFERENCES

- [1] "Femtosecond Carrier Thermalization in Dense Fermi Seas," W. H. Knox, D. S. Chemla, G. Livescu, J. E. Cunningham, and J. E. Henry, *Physical Review Letters*, 61, (11), 1290 (1988).
- [2] "Electron-Hole Drag in Semiconductors," R. A. Hopfel, *Solid State Electronics*, 31, (3), 4,643 (1988). "Negative Photoconductivity due to Carrier Drag in GaAs and GaAlAs Quantum Wells," S. Juen, R. A. Hopfel, A. C. Gossard, *Applied Physics Letters*, 54, (21), 2097 (1989).
- [3] "Velocity Field Characteristics of Minority Carriers (Electrons) in p-In_{0.53}Ga_{0.47}As," J. Degani, R. F. Leheny, R. E. Nahory, and J. P. Heritage, *Applied Physics Letters*, 39, (7), 569 (1981).
- [4] "Monte Carlo Investigation of Electron-Hole-Interactions Effects on the Ultra Fast Relaxation of Hot Photoexcited Carriers in GaAs," M. A. Osman and D. K. Ferry, *Physical Review Letters*, 36, (11), 6018 (1987).
- [5] "Femtosecond Spectroscopy of Semiconductors and Large Molecules," C. L. Tang, F. W. Wise, and I. A. Walmsley, *Journal of Modern Optics*, 35, (12), 1939 (1988) and references therein.
- [6] "Monte Carlo Modeling of Femtosecond Relaxation Processes in AlGaAs/GaAs Quantum Wells," C. J. Stanton, D. W. Bailey, and K. Hess, *IEEE Journal of Quantum Electronics*, 24, (8), 1614 (1988).
- [7] "Hot-Electron Recombination at Neutral Acceptors in GaAs: A CW Probe of Femtosecond Intervalley Scattering," R. G. Ulbrich, J. A. Kash, and J. C. Tsang, *Physical Review Letters*, 62, (8), 949 (1989).
- [8] "Energy Transport Numerical Simulation of Graded AlGaAs/GaAs Heterojunction Bipolar Transistors," E. M. Azoff, *IEEE Transactions on Electron Devices*, 36, (#4), 609 (1989).
- [9] "A Self-Consistent Particle Simulation for (AlGa)As/GaAs HBTs with Improved Base-Collector Structures," R. Katoh, M. Kurata, and J. Yoshida, *Proceedings of the International Electron Devices Meeting*, p. 248 (1987).

- [10] "Monte Carlo Simulation of AlGaAs/GaAs Heterojunction Bipolar Transistors," K. Tomizawa, Y. Awano, H. Hashizume, *IEEE Electron Devices Letters*, 5, (#9), 362 (1984).
- [11] "Self-Consistent Monte Carlo Study of High Field Transport in Graded Heterostructures," A. Al-Omar and J. P. Krusius, *Journal of Applied Physics*, 62, (9), 3825 (1987).
- [12] "Microscopic High Field Transport in Graded Heterostructures," A. Al-Omar and J. P. Krusius, *Solid State Electronics*, 31, (#3/4), 329 (1988).
- [13] "Molecular Coherent Potential Approximation for Zinc-Blende Pseudobinary Alloys," R. J. Lempert, K. C. Hass, and H. Ehrenreich, *Physical Review*, B36, (2), 1111 (1987).

ISEP PUBLICATIONS

- 1. "Investigation of Lateral Effects on Electron Launching in Graded AlGaAs-GaAs Heterostructure Devices Using Two-Dimensional Self-Consistent Monte Carlo Method," S. Weinzierl and J. P. Krusius, presented at the NASECODE VI Conference, Dublin, Ireland, July 11-14, 1989. *Proceedings of the Sixth International NASECODE Conference*, Boole Press, Ireland, edited by J.J.H. Miller, pp. 434-440.
- 2. "Lateral Space-Charge Effects on Ballistic Electron Transport Across Graded Heterojunctions," S. Weinzierl and J. P. Krusius, presented at the 6th International Conference on Hot Carriers, Phoenix, Arizona, July 1989. *Solid State Electronics*, 32, (#12), 1557-1561 (1989).
- 3. "Optimization of the Design of the Heterojunction Vertical Field Effect Transistor Using a Two-Dimensional Self-Consistent Monte Carlo Method," S. Weinzierl and J. P. Krusius, presented at the 1989 IEEE/Cornell Conference on High Speed Devices and Circuits, Ithaca, NY, August 1989. *Proceedings of the IEEE/ Cornell Conference on Advanced Concepts in High Speed Semiconductor Devices and Circuits*, IEEE, Cat. No. 89CH2790-4, pp. 237-245 (1989).
- 4. "Conditions for Suppression of Ballistic Injection at AlGaAs/GaAs Heterojunctions due to Lateral Space Charges," S. Weinzierl and J. P. Krusius, manuscript to be submitted for publication (March 1990).
- 5. "Optimization of Steady State and Transient Ballistic Injection in Vertical Heterojunction FET," S. Weinzierl and J. P. Krusius, manuscript in progress, to be submitted for publication (April 1990).

6. "Simulation of Ultrafast Carrier Processes in Pulse/Probe and Dual Pulse Correlation Probing of InGaAs Type Narrow Gap Semiconductors," J. E. Bair and J. P. Krusius, to be presented at the SPIE Technical Conference on Advances in Semiconductors and Superconductors: Physics Toward Device Applications: Ultrafast Laser Probe Phenomena in Bulk and Microstructure Semiconductors III, Conference 1282, San Diego, March 17-21, 1990 (invited paper); to be submitted for publication in proceedings.

NOVEL VLSI ALGORITHMS AND ARCHITECTURES FOR HIGH DATA RATE DIGITAL FILTERING

TASK #: 5

TASK PRINCIPAL INVESTIGATOR: Gianfranco Bilardi
(607) 255-9212

OBJECTIVE

The goal of this research is the design of algorithm, architecture, and layout of special-purpose VLSI systems for very fast signal processing. The objective is to obtain circuits that make optimal use of silicon, achieving the maximum data rate possible for a given amount of silicon area. It is important to identify the factors that limit the performance and to obtain quantitative expressions for such limitations.

DISCUSSION OF STATE-OF-THE-ART

About a decade ago, a VLSI model of computation was proposed [1,2] to capture the essential features of VLSI as a computing medium and to allow for mathematical analysis of chip design. The performance of a VLSI circuit has generally been measured in terms of the chip area A and the computation time T . The tradeoff between these two measures has been investigated for many computational problems (see [3] for some examples). In the process, considerable knowledge has been gained on algorithmic, architectural, and layout issues arising in the design of VLSI systems.

In the field of signal processing, the area-time tradeoff of basic operations such as *convolution* [4], *discrete Fourier transform* (see [5,6]) has been studied extensively. An investigation of the VLSI complexity of *digital filtering* was initiated in [7], although much remains to be done in this direction.

The design of a special-purpose VLSI system typically exploits the properties of the particular operation to be performed. In spite of many case studies, few general design principles have emerged. Progress in this area would be very desirable since it could simplify the design process considerably.

In the next section, we shall report about the progress made in the past year both on the study of specific signal processing operations, and on the study of general questions which relate to the discipline of VLSI design.

PROGRESS

We summarize below some of the main findings and directions of our research. A more detailed account can be found in the references listed in the section JSEP Publications.

Filtering and Prefix Computations.

One of the main targets of our work is a complete characterization of the area/data-rate tradeoff of digital filtering. An early study [7] indicated that the *twisted-reflected-tree* (TRT) is very efficient for the execution of some high data-rate algorithms for digital filtering. The TRT is also the architecture of choice for a general class of problems known as *prefix computations*.

The desire to obtain a deeper understanding of the relationship between filtering and prefix computation has motivated an extensive investigation of the latter. Various resource tradeoffs have been completely characterized in terms of algebraic properties of the semigroup underlying the prefix computation (Ref. [8], JSEP publication [3]). These results have independent interest since variants of the prefix operation (e.g., fetch-and-add on the Ultracomputer, scan on the Connection Machine, and multiprefix on the Fluent Machine) play an important role in parallel computing.

One interesting finding has been that the TRT is not an optimal network for all prefix computations. In fact, there is a large class of semigroups whose prefix problem can be solved on a more compact binary-tree network. An algebraic characterization of this class has been developed.

Recently, we have been able to establish a connection between filtering and prefix computation, by introducing the notion of *universal filter*, a circuit with two types of input: the signal to be processed and the description of the filter to be applied to that signal. Clearly, a specific filter can always be obtained as a specialization of a universal one. We have shown that, assuming infinite precision, a universal filter can be viewed as performing a prefix computation over a certain semigroup. We are currently exploring extensions of these results to finite-precision computation.

Multidimensional Signal Processing.

VLSI signal processing can be extended to multidimensional signals. In JSEP publication [1] we have taken a step in this direction for the *multidimensional discrete Fourier transform*. The cases of complex arithmetic and modular arithmetic have both been investigated. Area-time optimal designs have been developed for a wide range of computation times. Previously published lower bounds on the area-time performance were based on fallacious arguments, and completely new arguments have been developed to establish performance lower bounds.

The results published in JSEP publication [1], in common with almost all the studies on DFTs, make specific assumptions on the factorability of the size of the transform. In recent unpublished work we have succeeded in constructing optimal circuits in the general case. We are also investigating extensions of the lower bounds to the complex field.

Distributed Implementation of Shared Memory.

The design of a special-purpose VLSI system typically involves the choice of a suitable parallel algorithm and architecture for the desired task. The architecture should support well the execution of that algorithm, and have the

smallest layout compatible with this requirement. It is often the case that parallel algorithms have already been proposed in the literature. However, most algorithms are developed for a shared-memory model of computation such as the *parallel random access machine* (PRAM). It would be of great interest if one could automatically transform a PRAM algorithm into a VLSI algorithm.

As a step in this direction, together with Kieran Herley (a graduate student supported by JSEP, who has recently completed a Ph.D.), we have studied the problem of simulating a PRAM on a *bounded-degree network*, a model more suited to VLSI implementation. Optimal simulation schemes have been obtained [10] for the case in which the memory size grows at least as a polynomial function in the number of processing elements. Further results have been obtained recently by Herley in JSEP publication [4] in the case of smaller memory. Other interesting findings include an optimal algorithm for a generalized version of message routing on bounded-degree networks (JSEP publication [5]), and a novel distributed memory map with compact representation (JSEP publication [6]).

Lower Bounds.

Communication is often the critical factor limiting the speed of parallel algorithms. In JSEP publication [2], the constraints on the computation time posed by propagation of information are analyzed for a specific class of functions. The goal is to gain a better understanding on what properties of a function constrain its parallel computation time. A general technique is developed to obtain lower bound on the parallel computation time of *monotone* boolean functions in terms of the length of their largest prime implicant or prime clause.

Universal Networks.

Most architectural designs for special-purpose VLSI systems are on a case by case basis. It is interesting to investigate the existence of "universal" networks, which should be almost as good as any other network of the same cost (area). With Ph.D. student, Paul Bay, we have recently discovered a network of one $O(A)$ that can be programmed to simulate any VLSI circuit of area A , with a logarithmic penalty in time (JSEP publication [7]). This result is encouraging, and we hope to extend it in several ways.

SCIENTIFIC IMPACT OF RESEARCH

The work on prefix computation has the potential for important applications, not only to special-purpose VLSI structures, but also to general-purpose parallel computers. Indeed, already a number of parallel programming languages provide prefix as a primitive operation (often under a different name), and some machines support the operation in hardware.

The relationship between prefix and filtering that we have established is likely to lead to a new perspective on the filtering problem. We are already exploring this perspective which looks very promising.

Fourier transform techniques are fundamental in signal processing. Our optimal circuits for the multidimensional Fourier transform should find applications to multidimensional filtering, among others.

The problem of translating shared-memory algorithms into distributed-memory algorithms is of fundamental importance. Any implementation of a large memory is bound by technological constraints to be a distributed one. Thus, the work reported in JSEP publications [4] and [5] has consequences for VLSI implementation of shared-memory algorithms.

Universal networks may provide the basis for considerable progress in silicon compilation.

DEGREES

1. Kieran Herley
Ph.D., Computer Science, December 1989
"Deterministic Simulation of Shared Memory on Bounded Degree Networks".

REFERENCES

- [1] "A Complexity Theory for VLSI," C. D. Thompson, Ph.D. Thesis, Department of Computer Science, Carnegie-Mellon University (Aug. 1980).
- [2] "The Chip Complexity of Binary Arithmetic," R. P. Brent and H. T. Kung, *Journal of ACM*, 28 (3) 521-534 (July 1981).
- [3] J.D. Ullman, *Computational Aspects of VLSI*, Computer Science Press, Rockville, MD. 1983.
- [4] "Area-Time Optimal VLSI Circuits for Convolution," G. M. Baudet, F. P. Preparata and J. E. Vuillemin, *IEEE Trans. on Computers*, C-32 (7) 684-688 (July 1983).
- [5] "Fourier Transforms in VLSI," C. D. Thompson, *IEEE Transactions on Computer*, 32 (1983), 1047-1057.
- [6] "Optimal VLSI Circuits for Discrete Fourier Transform," G. Bilardi and M. Sarrafzadeh, in *Advances in Computing Research*, 4, 87-101 (F. P. Preparata, Editor), JAI Press, Greenwich, CT.
- [7] "Digital Filtering VLSI," G. Bilardi and F.P. Preparata (invited paper), *Proc. Aegean Workshop on Computing: VLSI Algorithms and Architectures*, Loutraki, Greece, 1-11, Springer Verlag (July 1986).

- [8] "Size-Time Complexity of Boolean Networks for Prefix Computation," G. Bilardi and F. P. Preparata, *Journal of ACM*, 36 (2) 362-282, April 1989.
- [9] "Parallel Prefix Computation," R. E. Ladner and M. J. Fisher, *Journal of ACM*, 27 (4) 831-838 (Oct. 1980).
- [10] "Deterministic Simulations of PRAMs on Bounded-Degree Networks," K. Herley and G. Bilardi, *Proceedings of the 26th Annual Allerton Conference on Communication, Control, and Computing*, 1084-1093, Monticello, Illinois, September 1988. An extended version appears in the Technical Report 88-951, Department of Computer Science, Cornell University, November 1988.

ISEP PUBLICATIONS

- 1. "Optimal VLSI Architectures for Multidimensional DFT," G. Bilardi, S. Hornick, and M. Sarrafzadeh, *1989 ACM Symposium on Parallel Algorithms and Architectures*, Santa Fe, New Mexico, June 1989, 265-272.
- 2. "Time Lower Bounds for CREW-PRAM Computation of Monotone Functions," G. Bilardi and A. Moitra, *International Symposium on Automata, Languages, and Programming*, Stresa, Italy, 95-107, July 1989. Also Technical Report 89-1012, Department of Computer Science, Cornell University, May 1989.
- 3. "Efficient Simulations of Small Shared Memories on Bounded Degree Networks," K. Herley, *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, 390-395, North Carolina, October 1989.
- 4. "Improved Bounds for the Token Distribution Problem," K. Herley, Technical Report 89-1051, Department of Computer Science, Cornell University, October 1989.
- 5. "Deterministic Simulation of Shared Memory on Bounded-Degree Networks," K. Herley, Ph.D. Thesis, Technical Report 90-1090, Department of Computer Science, Cornell University, February 1990.
- 6. "Deterministic On-line Routing on Area-Universal Networks," P. Bay and G. Bilardi, manuscript in preparation, February 1990.
- 7. "Characterization of Associative Operations with Prefix Circuits of Constant Depth and Linear Size," G. Bilardi and F. P. Preparata, *SIAM Journal on Computing*, 19, (2) April 1990.

FAULT TOLERANT SIGNAL PROCESSING ARCHITECTURES

TASK #: 6

TASK PRINCIPAL INVESTIGATOR: Franklin T. Luk
(607) 255-5075

OBJECTIVE

Systolic arrays have emerged as a preferred means for performing the many matrix computations central to real-time signal processing. For examples, interesting hardware projects are in progress at the Naval Ocean Systems Center in California [1], the Lincoln Laboratory in Massachusetts [2], and the Royal Signals and Radar Establishment in the United Kingdom [3]. Indeed, the Systolic Linear Algebra Parallel Processor project at NOSC is based on our theoretical design. An important task is provide a wide range of cost effective fault tolerance options to meet the varying needs of all users of systolic arrays. What are these needs? At the minimum, error detection is necessary because errors in matrix computations are essentially undetectable by a mere examination of the results. For the real-time environment, we need totally reliable arrays that do not reduce the throughput of the original array.

Existing fault tolerance methods can be divided into three categories: concurrent error detection followed by reconfiguration, error masking, and error correcting data encoding. Reconfiguration schemes are extremely powerful techniques that tolerate any pattern of errors, permanent or transient. However, the performance loss caused by concurrent error detection, reconfiguration, and rollback makes reconfiguration too debilitating for the real-time environment while the hardware redundancy and complexity make it too costly for any other. Error masking schemes are superior in that continuous system operation is provided. However, the apparently necessary tripling or quadrupling of hardware is extremely costly and error masking schemes are vulnerable to certain patterns of errors. Encoded data is exciting for its low time and hardware overheads, but its error detection and correction capabilities for multiple errors seem limited.

Which technique can provide the choices we need? We like a variety of methods, each with its own strengths and weaknesses. The techniques are algorithm-based fault tolerance, virtual redundancy, and pair and spare. These strategies, when used either individually or in combination, will provide a rich source of cost-effective fault tolerant systolic arrays.

In this task, we are primarily concerned with hard and soft processor calculation errors. We propose to examine critically both existing and new techniques for achieving fault tolerance, first in systolic arrays and then in real-time signal processing systems. Further, we propose to formulate and evaluate schemes for fault-tolerance in such arrays and systems, including the digital

filtering structures being investigated by Bilardi and the configurations with multiple functional units being studied by Tornø.

DISCUSSION OF STATE-OF-THE-ART

Algorithm-based fault tolerance, proposed by Jacob Abraham and his students at the University of Illinois [4,5,6,7], is a technique specially tailored for systolic algorithms and architectures. By encoding the input data as checksums and by redesigning algorithms for the encoded data, one can detect and correct transient errors that have occurred during the computations. This approach requires a very low overhead and uses simple arithmetic; Abraham et al. apply it to basic operations such as matrix-matrix multiplication, LU decomposition and matrix inversion. For correcting a transient error in Gaussian elimination, they propose a computation rollback to the point just before the error occurred. However, it is hard to execute rollbacks on systolic arrays. In [8] we show how to avoid rollbacks by computing the correct decomposition from the erroneous one.

In [9] we extend Abraham's checksum scheme for the LU decomposition to a unified scheme for three different triangularization procedures: LU decomposition, Gaussian elimination with pairwise pivoting, and QR decomposition. We show how to represent the error as a rank-one perturbation to the data, and develop a new error model where the occurrence time of the error is not involved.

Although in exact arithmetic the checksum scheme works well (an inconsistent checksum indicates the presence of a transient error), in floating point arithmetic an undesirable growth of rounding errors may cause confusion. There is a need to establish a tolerance to decide if an inconsistent checksum were caused by a (large) transient error or by roundoffs. In [10] we analyze the effects of rounding errors on the checksum scheme and establish a tolerance for transient error detection. Furthermore, we show that the tolerance is necessarily large for the LU decomposition and for Gaussian elimination with pairwise pivoting, but is acceptably small for the QR decomposition.

The guiding principle of virtual redundancy is the same as that of any space redundancy technique: make the same calculation on different processors and compare results to detect and correct errors. However, instead of replicating the hardware, one replicates the data and takes advantage of idle processors to make the redundant calculations (cf. Kim and Reddy [11]).

In a "pair and spare" configuration, there are two pairs, say A and B, of processors. While both processors of pair A agree and both processors of pair B agree, the system uses the results of pair A. If either pair disagrees, the results of the agreeing pair are used while a signal is sent to maintenance alerting them of a critical state. We define a critical state to be a state where one more error may cause the system to produce faulty results. While either pair is being repaired, the other pair is used to run the computer. This scheme is used by Stratus Computer Corp. in their non-stop systems.

PROGRESS

In [12] we introduce a new algorithm-based fault tolerance technique specifically designed for use in recursive least squares minimization. Through monitoring a single scalar $x(n)$, we get error detection. The same quantity can also be used as an indicator for correction. This technique applies in equally effective fashion to many other systolic arrays, as it provides the same properties for the QR decomposition phase of the algorithms. Our scheme does not require the summation process during the decoding stage; it requires only the monitoring of $x(n)$. Furthermore, due to the flow of data into the array, which is in a wavefront pattern, $x(n)$ is the only reasonable quantity to examine continually. The chief attribute of our technique is its remarkable simplicity.

Error correction has proved to be a much more difficult problem to solve than error detection when using weighted checksums. In [13] we provide a theoretical basis for the correction problem. We show that for a distance $d+1$ weighted checksum scheme, if a maximum of floor $(d/2)$ errors ensue then we can determine exactly how many errors have occurred. We further demonstrate that in this case we can correct the errors and give a procedure for doing so. We also show the close relationship of our scheme to the Reed-Solomon Code.

To avoid numerical overflows, a method is proposed in [5] that uses modular arithmetic to compute weighted checksums. A new scheme has been derived by us in [14]. We compare the two methods. In [5], if the word length equals 32 then the arithmetic is performed modulo the prime 8,589,934,583. Our scheme is relatively independent of the word length; it depends on the dimension of the matrix, say n , and the number of checksum vectors, say d . For $n = 1000$ and $d=50$, large values in light of current signal processing needs and hardware quality, we use the prime $p = 1,051$.

A totally unexpected bonus of our work is coming to light. In our study of various coding schemes for fault tolerance, we have discovered an efficient algorithm for solving the Yule-Walker problem: a set of linear equations with a coefficient matrix that is Toeplitz and a right hand vector that has essentially the same elements. This problem has very important applications in signal processing.

SCIENTIFIC IMPACT OF RESEARCH

Systolic arrays have been proposed and constructed for various applications, especially in the signal processing area. An in-depth and systematic investigation of fault-tolerance techniques for systolic arrays is both timely and promising.

New VLSI signal processing systems built upon wafer scale integration call for new fault tolerance techniques that will recognize the unique constraints and opportunities offered by this emerging technology. The task of detecting and correcting transient errors is gaining in importance as transistors get smaller and as we send more computing systems into out space (alpha particles are known to change bits).

We believe that the investigations being proposed will yield approaches, more cost effective and flexible than traditional techniques such as modular redundancy and quadded logic, and may initiate new areas in the study of fault tolerance. Indeed our interdisciplinary approach has led to the discovery of new algorithms for Toeplitz problems, a very important area in numerical computing and signal processing.

DEGREES

1. L. Magnus Ewerbring
Ph.D., Electrical Engineering, August 1989
"A New Generalization of the Singular Value Decomposition:
Algorithms and Applications".
2. Cynthia J. Anfinson
M.S., Applied Mathematics, August 1989
"A Theory of Algorithm-based Fault Tolerance".

REFERENCES

- [1] "SLAPP: A Systolic Linear Algebra Parallel Processor," B. L. Drake, F. T. Luk, J. M. Speiser and J. J. Symanski, *IEEE Computer*, 20, (7) (Special Issue on Systolic Arrays), 45-49 (July 1987).
- [2] "Wafer-scale Systolic Array for Antenna Processing," C. M. Rader, *Proceedings of ICASSP 88, International Conference on Acoustics, Speech, and Signal Processing*, New York, New York, 2069-2071 (1988).
- [3] "Some Systolic Array Developments in the United Kingdom," J. V. McCanny and J. G. McWhirter, *IEEE Computer*, 20, (7) (Special Issue on Systolic Arrays), 51-63 (July 1987).
- [4] "Algorithm-Based Fault Tolerance for Matrix Operations," K-H. Huang and J. A. Abraham, *IEEE Transactions on Computers*, C-33, 518-528 (1984).
- [5] "Fault-Tolerant Matrix Arithmetic and Signal Processing on Highly Concurrent Computing Structures," J-Y. Jou and J. A. Abraham, *Proceedings IEEE*, 74, 732-741 (1986).
- [6] "Fault Tolerance Techniques for Highly Parallel Signal Processing Architectures," J. A. Abraham, *Highly Parallel Signal Processing Architectures*, K. Bromley, Ed., *Proceedings SPIE*, 614, 49-65 (1986).
- [7] "Fault-Tolerant Systems for the Computation of Eigenvalues and Singular Values," C. Y. Chen and J. A. Abraham, *Advanced Algorithms and*

- Architectures for Signal Processing I*, J. M. Speiser, Ed., *Proceedings SPIE*, 696, 228-237 (1986).
- [8] "Algorithm-Based Fault Tolerance for Parallel Matrix Equation Solvers," F. T. Luk, *Real Time Signal Processing VIII*, W.J. Miceli and K. Bromley, Eds., *Proceedings SPIE*, 564, 49-53 (1985).
 - [9] "Fault-Tolerant Matrix Triangularizations on Systolic Arrays," F. T. Luk and H. Park, *IEEE Transactions on Computers*, C-37, 1434-1438 (1988).
 - [10] "Analysis of Algorithm-Based Fault Tolerance Techniques," F. T. Luk and H. Park, *Journal of Parallel and Distributed Computing*, 5, 172-184 (1988).
 - [11] "A Fault-Tolerant Systolic Array Using TMR Method," J. H. Kim and S. M. Reddy, *Proceedings ICCD*, 769-773 (1985).
 - [12] "A Novel Fault Tolerance Technique for Recursive Least Squares Minimization," F. T. Luk, E. K. Torng, and C. J. Anfinson, *Journal of VLSI Signal Processing*, 1, 181-188 (1989).
 - [13] "A Theoretical Foundation for Weighted Checksum Methods," C. J. Anfinson, R. P. Brent and F. T. Luk, *Advanced Algorithms and Architectures for Signal Processing III*, F. T. Luk, Ed., *Proceedings SPIE*, 975, 10-18 (1988).
 - [14] "Choosing Small Weights for Multiple Error Detection," R. P. Brent, F. T. Luk and C. J. Anfinson, *High Speed Computing II*, K. Bromley, Ed., *Proceedings SPIE*, 1058, 130-136 (1989).

ISEP PUBLICATIONS

1. *Proceedings of SPIE Vol. 1152, Advanced Algorithms and Architectures for Signal Processing IV*, F.T. Luk, Editor, SPIE - The International Society for Optical Engineering, Bellingham, Washington, 1989, 43 papers/507 pages.
2. "A Novel Fault Tolerance Technique for Recursive Least Squares Minimization," F. T. Luk, E. K. Torng, and C. J. Anfinson, *Journal of VLSI Signal Processing*, 1, 181-188 (1989).
3. "Algorithm-based Fault Tolerance Techniques for MVDR Beamforming," C. J. Anfinson, A. W. Bojanczyk, F. T. Luk and E. K. Torng, *Proceedings of ICASSP 89, International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, Scotland, 2417-2420 (1989).

REAL-TIME SIGNAL PROCESSING SYSTEMS WITH MULTIPLE FUNCTIONAL UNITS

TASK #: 7

TASK PRINCIPAL INVESTIGATOR: H. C. Torng
(607) 255-5191

OBJECTIVE

This task investigates the architecture and design of superscalar processors for real-time signal processing tasks. Superscalar processors, with its complement of functional units, have the potential of implementing concurrent processing at both the instruction and the task levels to meet stringent deadlines.

Instructions from a single instruction stream are dispatched, sometimes simultaneously, to available functional units; this constitutes a form of concurrent processing at the instruction level. A task, on the other hand, presents an instruction stream to the processor. Independent or partially ordered tasks can be processed by a superscalar processor at the same time; this constitutes a form of concurrent processing at the task level.

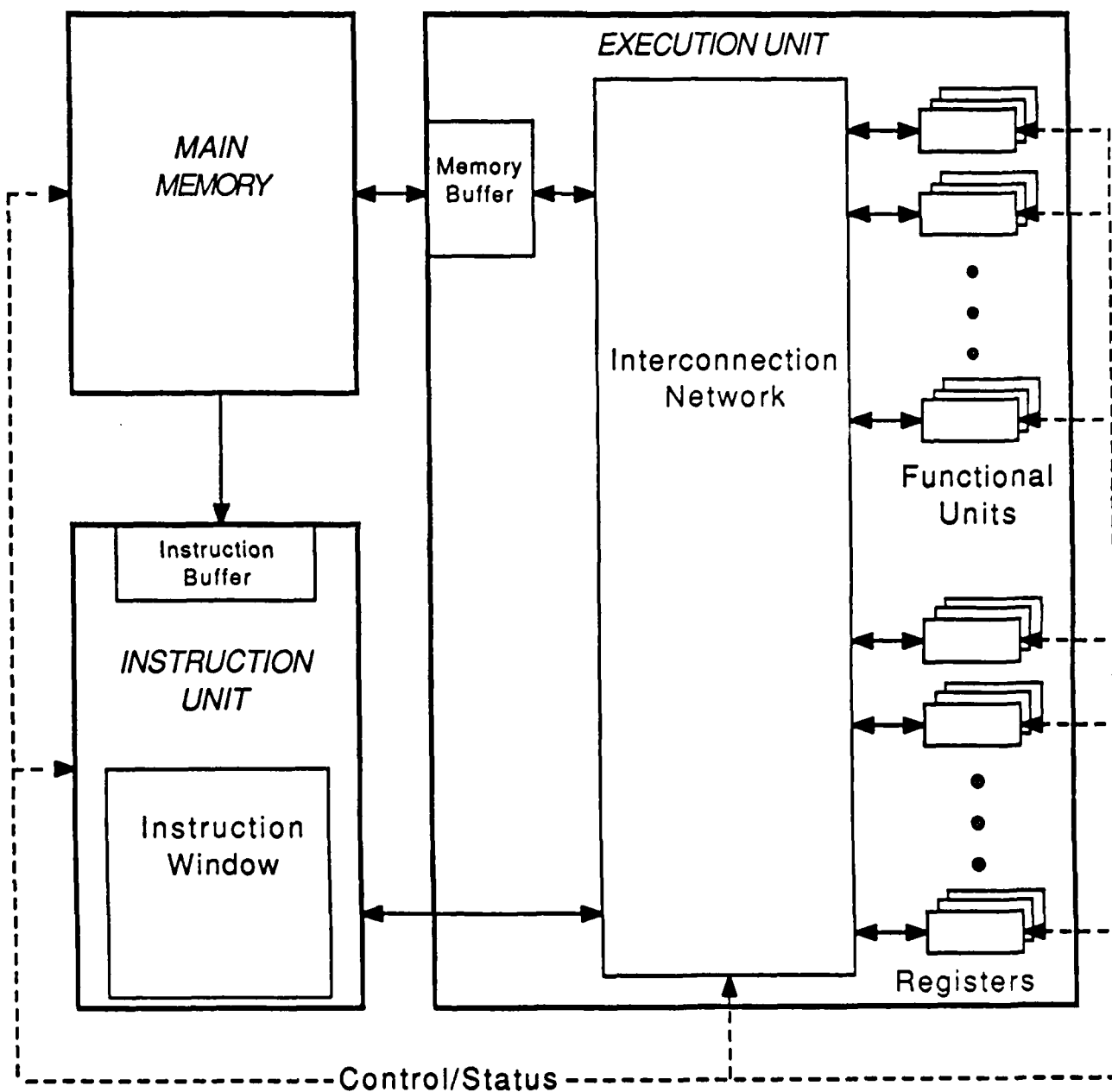
We use concurrent processing as a means for performance enhancement. Concurrent processing should complement expected continuing improvements in device and packaging technologies to handle the imposed loads by real-time signal processing and other time critical tasks, and to reduce computation latency.

In our study, we have concentrated on processor systems with multiple functional units (superscalar). The general structure of a superscalar is depicted in Figure 1.

In this period, we have been investigating the following issues that still confront the deployment of superscalars for real-time signal processing tasks:

1. The incorporation of data-flow concept in instruction issuance: A key to performance enhancement for superscalars is to identify and issue multiple instructions concurrently. The introduction of data flow concept into instruction windowing mechanisms has the potential of making the detection of dependency-free instructions for multiple and out-of-order issuances easy; this of course bring about enhanced throughput.
2. Interrupt handling: A critical requirement for real-time signal processing computation is that the computer system be able to provide *prompt* and *precise* interrupt handling capabilities. Interrupt requests have to be promptly handled because tasks that initiate these requests have to be processed as soon as possible. Responding to an interrupt request, the processor first stores its processor state; this has to be done precisely so that the interrupted process can be resumed at the point of interruption later.

**FIGURE 1: Multiple Functional Unit Processor:
General Architecture**



The presence of multiple functional units enables the concurrent execution of several instructions from the same instruction stream. Since these instructions are at various stage of execution, it is a challenging task to identify and then store a precise processor state quickly.

3. Branch handling: The presence of conditional branch instructions invariably introduces turbulences into a dynamic instruction stream. It can be safely stated that these undesirable effects are magnified in superscalars.
4. Concurrent execution of multiple instruction streams: The processing of two or more independent instruction streams on a superscalar may increase the average density of "independent" instructions in an instruction window, and "mask" turbulences due to branching and instruction execution latency. A superscalar thus has the potential of realizing a multiple instruction stream and multiple data stream (MIMD) machine.

DISCUSSION OF STATE-OF-THE-ART

Concurrent processing is an approach to meeting the "real-time challenge" in signal processing; it takes many forms: parallel systems, superscalars, and various combinations of the previous two.

The main difference between a "conventional machine" and a superscalar is the presence of multiple and "specialized" functional units in the latter. The instruction unit is charged with the task of issuing these functional units with sufficient instructions to keep them busy. At a given machine cycle, several instructions may be executed concurrently; this is one of the main reasons that superscalars produce high throughput.

The interconnection network provides paths between registers and functional units; and between registers and the main memory. The network may range from a non-blocking, fully connected crossbar switch to a set of buses. Its selection plays a critical role in determining the performance of a superscalar for a given set of tasks.

Superscalars, such as CRAY [1] and Floating Point Systems [2] machines, are designed for general purpose applications. Recently announced microprocessors such as Motorola 88000, Intel 80960, MIPS, and AMD 29000 also implement this configuration, with a small number of chips; furthermore, these processors represent the new wave of Reduced Instruction Set Computers (RISC).

A very important point in the design of superscalars for real-time signal processing is: When a specific operation is frequently performed, an optimally designed hardware structure can be designed and incorporated into the system as a functional unit.

For example, Luk and Bilardi in their tasks develop efficient structures for matrix and filtering operations; these structures can be considered as functional units in a real-time signal processing superscalar.

The problem we have been working on is that such machines generally do not make the best use of their functional units as most of these units stay idle; this is so because at most one instruction is issued per machine cycle. In other words, these precious execution resources are being starved because of an inadequate supply of instructions.

Two notable schemes have been implemented to alleviate the starvation problem: Thornton's scoreboard [3] and Tomasulo's reservation stations with Common Data Bus [4]. In both cases, instructions are still issued according to the order they appear in an instruction stream -- an instruction will not be issued until all instructions which precede it have been issued already, and at most one instruction is issued per cycle.

To remedy this situation, we formulated a "Dispatch Stack" (DS) scheme [5], which issues

1. multiple instructions per machine cycle, if possible; and
2. instructions out of sequence.

According to the DS scheme, two or more instructions may be issued concurrently as long as there are no data dependencies and there are functional units to execute them. Multiple instruction issuances per machine cycle increases the rate of instructions dispatched to the execution complex and thus enhances system performance.

Furthermore, an instruction can be issued to an available functional unit as long as it is free of data dependencies, even though some of its preceding instructions are still awaiting issuances; the issuance of instructions is thus non-sequential. In implementing such a scheme, the instruction issuance rate is enhanced as ready instructions can be issued ahead of those which precede them.

The DS is conceptually a "window", which displays and checks dependencies among instructions in an instruction stream. The dispatch stack can be either realized in software or hardware.

One task that we have undertaken is to study a more suitable format for the instruction window to facilitate dependency checking and instruction dispatching.

To use superscalars for real-time and transaction oriented applications, we also have to address the issue of how to handle interrupts promptly and efficiently.

Interrupts can be classified into three types: external interrupts, exception traps, and software traps. External interrupts are generated from or by the environment -- such as the processing of a newly arrived task. Abnormalities encountered in system processing, such as division by zero, overflow, or illegal operations, generate exception traps. Software traps are instructions which initiate interrupt requests; these traps provide a means of controlling certain software applications.

Interrupt handling mechanisms are evaluated by the following two factors:

Latency: An interrupt handling mechanism should be judged by the latency between the receipt of an interrupt request and the completion of saving the processor state. Clearly, any acceptable interrupt handling mechanism should strive to reduce this latency as much as possible. This measure is very important for real-time applications, transaction processing, and multiprocessing.

Cost: The amount of hardware and software costs incurred by the installation of an interrupt handling mechanism must be taken into account. Furthermore, we have to identify precisely the performance degradation that the interrupt handling mechanism may have inflicted on the system.

The CRAY machines [1] generally allow instructions under execution to complete before the processor state is stored; a penalty in long latency is consequently exacted. In the IBM 360/91 [4], a precise interrupt is realized by allowing all issued instructions to complete their execution; this results in considerable latency. If an imprecise interrupt is generated, the processor state of the system is lost and the system cannot be restarted precisely at the interrupted point.

Two other approaches to interrupt handling for superscalars have recently been proposed. In installing two or more additional "checkpoints" [6], the system can respond to an interrupt request by "retreating" to one of these checkpoints. Clearly, this proposed approach will degrade system performance, both in processor speed, and in the time required to restore to a consistent processor state upon receiving an interrupt request. The speed of the system will be slowed down by the movement of state information as the states change, and by the additional read instruction which must precede all instructions which alter the memory. A performance penalty has to be taken to correct the memory to a consistent state when an interrupt request is received.

Additional shift registers can also be installed in the processor to make certain that results are loaded into registers in order [7], even though they may be produced out of sequence. This approach introduces considerable degradation to system performance and also incurs additional hardware costs.

Branching is an indispensable ingredient in any meaningful program; it however injects performance damping turbulences into the instruction stream. How to handle conditional branching efficiently remains a difficult challenge for computer architects. A clear survey of possible techniques in handling conditional branches can be found in [8]. The proposed and implemented systems discussed previously do not approach this opportunity aggressively. Prefetching, small and tentative, is implemented in some. Checkpointing can again be applied to allow instruction execution on an assumed path. If the assumption made is proven incorrect, a consistent processor state can be restored through the processor state corresponding to the checkpoints implemented [6]. In most

cases, the supply of instructions is usually disrupted by the presence of conditional branch instructions.

Our investigation indicates that due to inter-instruction dependencies and branching turbulences, a single instruction stream may not be able to take the full advantage of the execution resources of a superscalar. The notion that a superscalar can concurrently execute several independent instruction streams is NEW and exciting.

PROGRESS

We have made considerable progress in the four issues enumerated in the OBJECTIVE section.

DATA FLOW DISPATCH STACK: To facilitate the detection of "independent" instructions in a DS, we have formulated a new format for each entry in the stack. Each source field contains either the operand, copied from a source register, or an instruction tag, denoting the instruction that will produce the required item. Each entry of the DS is essentially a reservation station, specified in [4].

When an instruction completes its execution, its result will be deposited into all waiting locations that have its tag. Furthermore, a new instruction can be brought in, using the same tag and the same entry in the DS.

The new structure, called the Data Flow DS, obviates the need for the DS to check data dependencies among instructions. And new instructions are brought into empty slots in the stack without having to perform time consuming compression.

Initial simulation results for this innovative structure on our networked HP 370 workstations are very encouraging. Detailed information is documented in a preliminary report.

INTERRUPT HANDLING: We have constructed a new approach to implementing efficient and prompt interrupt handling. The contents of the DS can be used to define a revised interrupt point. By saving the DS together with other conventional processor state components, we can restore the interrupted process precisely.

To reduce the wasted efforts due to the abandonment of an instruction processing vector elements, we proposed to add a "vector element number" field to the DS. In so doing, much work that has been accomplished for processing a vector instruction can be saved.

A paper describing this approach is submitted to IEEE Trans. on Computers. And a patent application has been filed in January 1990.

BRANCH PREDICTION: We have investigated the installation of a modified Branch Target Buffer[8] into the Instruction Unit to handle multiple

outstanding branch predictions. The detailed design and its performance will be presented in a paper under preparation.

MULTIPLE STREAM PROCESSING: We have formulated and evaluated a really innovative concept to boost superscalar performance: the processing of two or more independent instruction streams on a superscalar processor, creating an MIMD system. A paper has been submitted to 1990 International Conference on Computer Design.

SCIENTIFIC IMPACT OF RESEARCH

Superscalar processors represent an important architecture that can be exploited for the implementation of real-time signal processing systems. Our task has addressed several issues that computer designers will face in the next few years.

We have made considerable progress in the instruction issuing mechanism, interrupt handling, branch prediction and multi-stream processing. These features enhance significantly the performances of superscalars without raising the clock rate. And we believe that our study provides timely and much needed investigation into areas that are vital to the further development of such systems.

We have active ongoing discussions with IBM, Intel and AMD.

We have organized a special session on superscalars in the 1990 International Conference on Computer Design, with contributions from Cornell, Illinois, Wisconsin, IBM, Intel, and AMD.

DEGREES

None

REFERENCES

- [1] "The CRAY-1 Computer System," R. M. Russell, *ACM Communications*, 21, (1), 63-72 (January 1978).
- [2] "An Approach to Scientific Array Processing: The architectural design of the AP-120B/FPS 164 family," A. E. Charlesworth, *Computer*, 14, (9), 18-27 (September 1981).
- [3] DESIGN OF A COMPUTER: The Control Data 6600, J. E. Thornton, *Scott Foresman* (1970).
- [4] "An Efficient Algorithms for Exploiting Multiple Arithmetic Units," R. M. Tomasulo, *IBM Journal*, 11, 25-33 (January 1967).

- [5] "An Instruction Issuing Approach to Enhancing Performance in Multiple Functional Unit Processors," R. D. Acosta, J. Kjelstrup and H. C. Torng, *IEEE Trans. on Computers*, C-35, (9), 815-828 (September 1986).
- [6] "Checkpoint Repair for High Performance Out-of-Order Execution Machines," Hwu, Y. N. Patt, *IEEE Trans. Computer*, C-36, (12), 1496-1514 (December 1987).
- [7] "Implementing Precise Interrupts in Pipelined Processors," J. E. Smith and A. R. Pleszkun, *IEEE Trans. Computer*, C-37, (5), 562-573 (May 1988).
- [8] " Branch Prediction Strategies and Branch Target Buffer Design," J. K. F. Lee and A. J. Smith, *IEEE COMPUTER*, 17, (1), 6-22 (January 1984).

ISEP PUBLICATIONS

- 1. "A New Interrupt Handling Approach for Out-of-Order Execution Machines," Martin Day and H. C. Torng, submitted to *IEEE Transactions on Computers*, February 1990.
- 2. "The Concurrent Execution of Multiple Instruction Streams on Superscalar Processors," George E. Daddis Jr. and H. C. Torng, paper submitted to 1990 International Conference on Computer Design.
- 3. "A Taxonomy of Multiple Processor Scheduling," Esther A. Hughes and H. C. Torng, paper submitted to 1990 International Conference on Parallel Processing.
- 4. "A Data Flow Dispatch Stack with Branch Prediction," George E. Daddis Jr. and H. C. Torng, manuscript in preparation, to be submitted June 1990.