(1)

ANNUAL REPORT

FEMTOSECOND CARRIER PROCESSES IN COMPOUND SEMICONDUCTORS AND REAL TIME SIGNAL PROCESSING

MAY 1, 1992 - APRIL 30, 1993

CONTRACT #F49620-90-C-0039

ANNUAL REPORT

Joint Services Electronics Program
Contract # F49620-90-C-0039
May 1, 1992 - April 30, 1993

FEMTOSECOND CARRIER PROCESSES IN COMPOUND SEMICONDUCTORS AND REAL TIME SIGNAL PROCESSING

March 10, 1993

J. P. Krusius
Principal Investigator

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	3. DISTRIBUTION / AVAILABILITY OF REPORT Approval for public release. Distribution unlimited.
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 F49620-90-C-0039
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Femtosecond Carrier Processes in Compound Semiconductors and Real Time Signal Processing

12. PERSONAL AUTHOR(S)
J. Peter Krusius

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual Report	FROM 5/1/92 TO 4/30/93	93/03/10	55

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

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

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) femtosecond carrier processes in 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 tolerance 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

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

TABLE OF CONTENTS

A. DIRECTOR'S OVERVIEW

This document is the third year annual report of the Cornell Joint Services Electronics Program for the period from May 1, 1992 to April 30, 1993. The present Cornell program carries two themes: femtosecond carrier processes in compound semiconductors, and real time signal processing. The program has advanced according to plan. Seven task investigators, Profs. R. Shealy, C. Tang, C. Pollock, P. Krusius, A. Bojanczyk, F. Luk, and H. Torng, with their graduate students have contributed to JSEP research this year. A substitute task for G. Bilardi's effort with Prof. Adam Bojanczyk was started in September 30, 1991. F. Luk has been for most of the current period on a leave of absence from Cornell University. He resigned from Cornell during the fall of 1992 in order to take a position as the chairman of the Computer Science Department at Rensselaer Polytechnic Institute, but has continued to supervise his JSEP graduate students to the end of the current program period. Ten graduate students have been partially, or fully, supported by JSEP this year. A total of 29 publications and three theses were prepared in this period or are now in various stages of processing. Two PhD degrees and one M.S. degree have been awarded to JSEP supported students during this reporting period.

A proposal for the continuation of JSEP research for the next three year period, starting May 1, 1993, was submitted on August 1, 1992. In this proposal all research was focused into the fundamentals of high speed photonic devices including issues from materials to devices. An on site review of the proposal was held at Cornell on October 29 and 30, 1992.

B. DESCRIPTION OF SPECIAL ACCOMPLISHMENTS AND TECHNOLOGY TRANSITION

B.1. Femtosecond Carrier Processes in Compound Semiconductors

Several significant achievements have been reached in the research performed under the compound semiconductor theme. The new off-campus organometallic vapor phase epitaxial (OMVPE) compound semiconductor materials growth facility is in its second year of operation under the leadership of Prof. R. Shealy. The first reactor has been in routine operation, while two other reactors are being readied for materials growth. In the first JSEP task high purity GaAs films have been grown by flow modulation at low V/III ratios, which greatly reduce the environmental impact and life safety risk of the hydride based OMVPE growth processes. Good morphology was obtained for V/III ratios as low as unity above 610 C growth temperatures. 77K mobility values for GaAs reached 110,000 cm^2/Vs for a V/III ratio of 5. Gas phase reactions between TMAA, TEG, and AsH_3 have been examined in detail. This study provides the much needed understanding of AlGaAs growth chemistry with these gases. The maximum exciton linewidth of 2.2

meV measured from AlGaAs films grown with this understanding compares well with the best ever reported result for that composition. Finally, AlGaInAs layers lattice matched to InP substrates, and strain layer superlattices, have been grown with the flow modulation technique in Cornell's unique multi-chamber OMVPE reactor. This will facilitate the opportunity to study new ordered quaternary structures with improved electronic properties in this materials system.

The research into new tunable femtosecond sources has continued in the second task under Prof. C. Tang. This group has demonstrated the generation of high repetition rate femtosecond pulses in the blue for the first time by intracavity doubling of a mode locked Ti:sapphire laser using β-BaB_2O_4. Over 800 mW of 430 nm light was produced with intracavity pumping at a repetition rate of 72 MHz and pulse width of 54 fs. A first broadly tunable femtosecond optical parametric oscillator, working from the deep red to the mid ir, was demonstrated. Tunable femtosecond pulses from 720 nm to 3.5 µm at a repetition rate of 80 MHz and power level of a few mW have been generated. Via external cavity pumping with a Ti:sapphire laser, this group produced record output powers up to 3.8 W and tunable femtosecond pulses in the multi-hundred mW range at a total conversion efficiency of 55%. These new femtosecond sources will have a broad impact and provide a wealth of new probes to the scientific community.

C. Pollock's research group has refined their tunable femtosecond optical characterization system for equal pulse and pump-and-probe configurations. Quality data can now routinely be taken. New data on near band gap carrier relaxation in undoped InGaAs thin films on transparent InP substrates has been measured as a function of photon energy and light intensity. Measurements on the effect of background charge, introduced via doping or synchronized optical pulses, are in progress. The analysis of the complex carrier relaxation behavior has been done in collaboration with the Monte Carlo group of J.P. Krusius. This group has for the first time successfully simulated the relaxation of optically generated near band gap carriers (electrons and holes) in femtosecond pump-and-probe and equal pulse correlation measurements. Via correlations of measurements and simulations these two groups have been able to prove that band renormalization and dynamic screening are the two most important phenomena beyond the usual carrier scattering processes. The excellent agreement obtained between simulated and measured results shows that the self-consistent ensemble Monte Carlo method works well to time scales as short as 150 fs, if all important physical processes are included. The understanding of the role of band renormalization, dynamic screening, and the Coulomb enhancement of the optical transition elements is likely to have a profound effect on high speed optoelectronic devices, an area which will be examined by the Monte Carlo group under the new proposed JSEP work.

Further special accomplishments are listed in the description of research under each of the tasks.

B.2 Real Time Signal Processing

The investigators involved in the real time signal processing theme, Profs. A. Bojanczyk, F. Luk, and H. Torng, have continued the synergistic work. A. Bojanczyk's research was focused on recursive windowed least squares problems arising in real-time DSP systems derived from the covariance differencing principle. New highly concurrent algorithms amenable to efficient parallel implementations were developed. H. Torng's research group has continued work on interrupt handling, branch processing, and implementation of the dispatch stack in architectures with multiple functional units. These advances significantly enhance the performance of superscalar processors without raising the clock rate. H. Torng organized the fourth "Project 2000" meeting on June 2 and 3 1992 at Cornell to report on computer engineering advances in the past year. About 25 industrial representatives attended this two day meeting. F. Luk has again organized the Advanced Signal Processing Algorithms, Architectures and Implementations Symposium for the SPIE and edited the proceedings (July 1992).

Further special accomplishments are listed in the description of research under each of the tasks.

C. DESCRIPTION OF INDIVIDUAL WORK UNITS

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

Task #1

Task Principal Investigator: **James R. Shealy**
(607) 255-4657

OBJECTIVE

The program objective for the JSEP materials task is to explore the use of advanced OMVPE processes including Flow Modulation Epitaxy for improved III-V structures. Studies have been concluded on AlGaAs structures using a new Al source which results in the highest purity AlGaAs with ultra low oxygen and carbon contamination. We are currently emphasizing new structures lattice matched to InP for electronic and optoelectronic devices. These include ordered strained layer superlattices of the quaternary AlGaInAs. The transport properties of such structures will be studied in other JSEP tasks. Several device structures on InP will be prepared, such as quantum well lasers and advanced transistors structures, to confirm the improvements in materials synthesis sought after, as well as, to support the new investigators in the new JSEP program (R. Compton and Y-H Lo).

DISCUSSION OF STATE-OF-THE-ART

The following discussion of the state-of-the-art is organized into separate sections on growth of high quality GaAs with low hydride consumption, gas phase reactions during the OMVPE growth of AlGaAs using trimethylamine alane, triethylgallium, and arsine, and the growth of AlGaInAs on InP.

Low Hydride Consumption using Flow Modulation Epitaxy

One of the major disadvantages of hydride-based low pressure OMVPE has been the inefficient use of hydrides. For example, in conventional reduced pressure OMVPE growth of GaAs using trimethylygallium (TMG) and arsine (AsH3), high molar V/III ratios are needed to obtain high purity material [1]. Efforts have been made to reduce AsH_3 consumption, including precracking of the AsH_3[2] and substituting triethylgallium (TEG) for TMG [3]. In some cases, less toxic group V sources have been substituted for AsH_3, but this also requires V/III ratios of 10 or greater to yield high purity material [4]. Low V/III ratios are used in metalorganic molecular beam epitaxy [5], but best GaAs results are p-type and have carbon concentrations exceeding $5(10^{14})$.

Despite the attempts to reduce the AsH_3 requirement to produce high quality GaAs, most low pressure OMVPE growth is still done at relatively high V/III ratios. This poses a potential safety problem due to the expulsion of the excess hydride that does not participate in the growth process and to the increased handling of the source containers. Efforts to minimize high pressure cylinder storage include an on-demand arsine gas generator, but a low 77K mobility was observed [6]. Since conventional OMVPE is not able to produce high quality GaAs at V/III ratios near unity, a new growth process was sought.

Gas Phase Reactions of Trimethylamine Alane in the Growth of AlGaAs

The OMVPE growth of aluminum containing III-V compound semiconductors has traditionally been plagued with high oxygen and carbon incorporation. These problems are due, in large part, to the widely used aluminum source, trimethylaluminum (TMA). Triethylaluminum (TEA) is also used as an aluminum source, and it has demonstrated lower carbon incorporation in AlGaAs than TMA. However, some residual oxygen still remains using TEA, and TEA has a low vapor pressure (0.5 torr at 55°C) which is inconvenient for OMVPE.

Recently, trimethlyamine alane (TMAA) has received much attention as a viable alternative aluminum source in both OMVPE [7] and Chemical Beam Epitaxy (CBE) [8]. Previous reports indicate that using TMAA along with TEG and AsH_3 under the appropriate growth conditions (very high V/III ratios and gas velocities) can result in the high purity OMVPE grown AlGaAs [7]. This is believed to be due to the lack of a direct Al-C bond in TMAA and to its ability to form involatile Al-O compounds when reacted with oxygen and H_2O, resulting in reduced oxygen contamination.

However, TMAA suffers two major drawbacks. Its low thermal decomposition temperature leads to predeposition on the reaction cell sidewalls, and its strong room temperature reactions with conventional Ga sources (TMG and TEG) result in unwanted parasitic reactions during growth. The growth chemistry using these precursors in OMVPE must be understood in order to optimize film quality. Although studies investigating the effects of gas-phase reaction between TMAA and TEG in CBE have been reported [8], no previous study exists for OMVPE.

Short period Superlattices

Short period superlattices have been considered to be a viable alternative to the conventional use of bulk ternary and quaternary layers. The ordered arrangement of the constituent atoms in the superlattice can improve the optical and electrical properties of the superlattice over those of the bulk material. Potential fluctuations associated with the random arrangement of constituent atoms in the mixed crystal are absent from the superlattice structure, resulting in the reduction of alloy scattering and consequently to increased mobility [9]. Furthermore, exciton broadening due to clustering in the mixed crystal can be avoided by using short period superlattices.

The use of short period, strained layer superlattices is especially important in the growth of $Al_xIn_{1-x}As$ and $Ga_xIn_{1-x}As$ on InP. Because these material systems are lattice matched to InP when x is approximately 0.5, the alloy scattering and exciton broadening effects due to the mixed crystal are close to their maximum. Thus, short period, strained layer superlattices should improve optical and electrical quality in these alloys dramatically.

PROGRESS

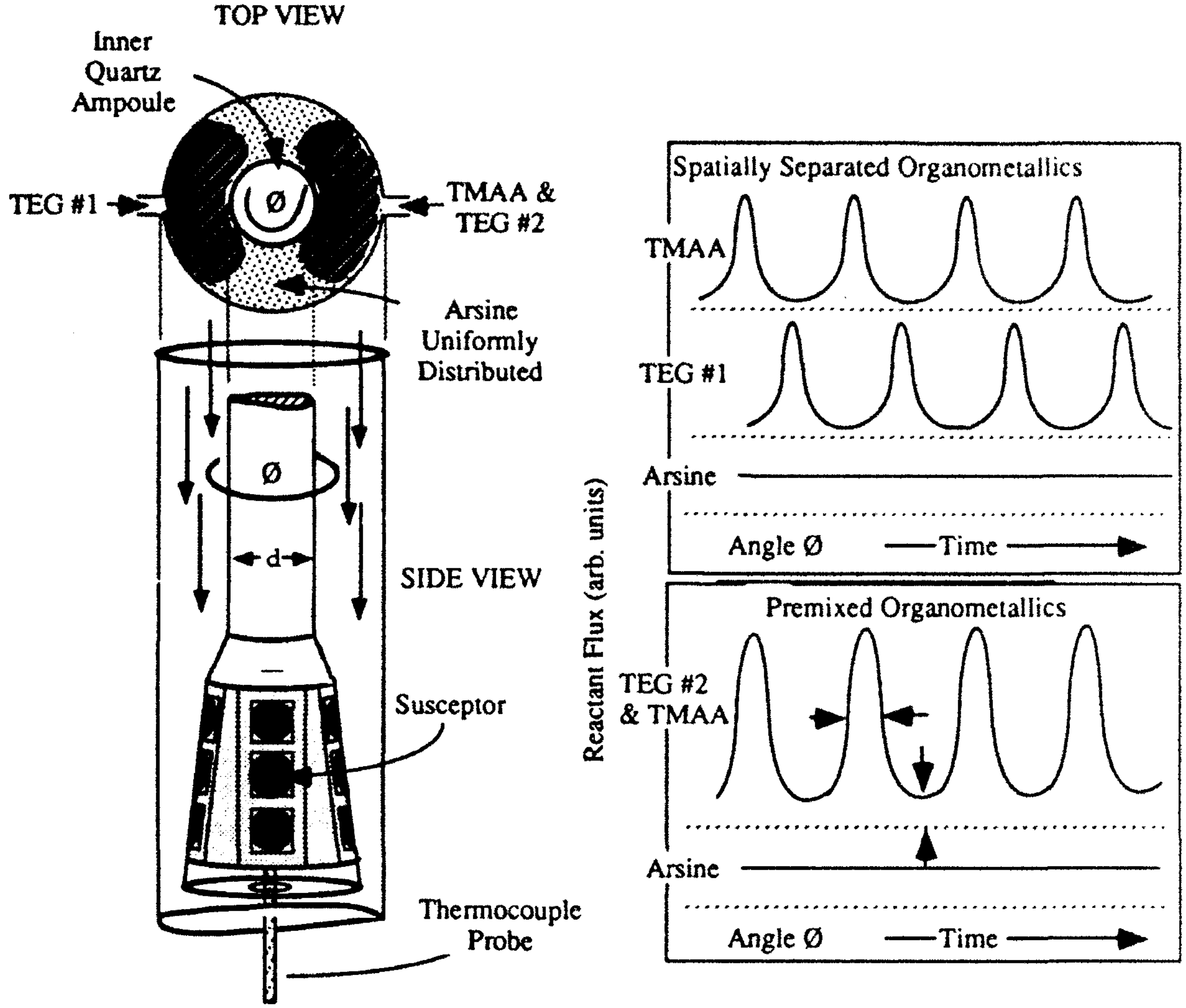
In this section, progress on the flow modulation growth of GaAs is presented, a basic study of the gas phase reactions of TMAA with TEG and AsH_3 is summarized, and recent progress of growth of AlGaInAs lattice matched to InP is outlined.

Flow Modulation growth of high purity GaAs

Using arsine and triethylgallium with flow modulation, a process has been developed that produces high purity GaAs with V/III ratios near unity and AsH_3 incorporation efficiency exceeding 30% [10]. This process greatly reduces the environmental impact and life safety risk of the hydride-based OMVPE method.

All films were grown using flow modulation epitaxy (FME) at 76 torr in a unique multichamber cell [11]. The substrates are rotated through group III and group V spatially rich zones without valve switching. During the group III exposure cycle, the local V/III ratio is estimated to be 25% of the average value. The growth cell and exposure cycle for the growth scheme are depicted schematically in Figure 1. The aluminum source is included for completeness and will be referred to in the next progress section.

The arsine flow requirements for obtaining good morphology have been determined over a wide range of growth temperatures. Good morphology was realized for V/III ratios as low as unity for growth temperatures exceeding 610°C. Below 610°C, the V/III ratio must be increased to maintain good surface morphology due to inefficient pyrolysis of the arsine.



(a) Reaction Cell

(a) Reactant Flow Schematic

Figure 1. (a) Schematic illustration of implementation of FME in the multichamber cell. Two TEG sources, one on each growth zone, allow for conventional premixed injection or spatially separated group III sources. The arsine is uniformly injected around the cell. The inner quartz ampoule (diameter d) serves to separate the reactant fluxes of each deposition zone. (b) Exposure cycle for premixed and spatially separated TMAA and TEG. The AsH_3 flow is uniformly distributed around the cell. Dotted lines represent the reactant zero flux reference. The degree of deposition zone separation is indicated schematically by the set of arrows in the lower diagram.

The electrical quality of the films grown at 635°C and with V/III ratios ranging from 1.8 to 22 was assessed using room temperature and 77 K Hall measurements. All films were n-type; no p/n crossover with decreasing V/III ratio was observed. The net impurity concentration varied from $3.7(10^{14})$ to $6.9(10^{14})$ cm^{-3} while the total impurity concentration varied from $7.7(10^{14})$ to $2.0(10^{15})$ cm^{-3}. The 77 K mobility ranged from 55,000 to 110,000 cm^2/V s, with the maximum value obtained at V/III ratio of 5. A V/III ratio of 1.8 (corresponding to an arsine efficiency of 17.2%) resulted in a 77 K mobility of

93,000 cm^2/V s. The highest room-temperature mobility was greater than 8000 cm^2/V s.

Optical quality of the samples grown at 635°C was evaluated using low temperature (1 K) photoluminescence (PL). For V/III ratios greater than 1.8, the dominant feature in the excitonic region of the spectra was the neutral donor exciton (D°, X). This along with a negligible acceptor exciton feature indicates strongly n-type material which is consistent with the Hall measurements. Films grown with V/III ratios greater than 1.8 demonstrated a single acceptor-related transition attributed to Mg, while those grown at V/III ratios between 1.8 and 1 showed acceptor transitions due to Mg and C. Significantly, normal excitonic features persisted until the V/III ratio was reduced to subunity. In the subunity film, new features near 1.503 eV, possibly due to defect related exciton emission, appeared. Films grown at subunity V/III ratios also exhibited poor surface morphology and reduced growth rates, indicative of arsenic limited growth.

Gas Phase Reactions of Trimethylamine Alane

Gas phase reactions between TMAA, TEG, and AsH_3 have been investigated [12]. Two dominant effects are observed: a TMAA-AsH_3 reaction results in varying film composition with V/III ratio, and a TMAA-TEG reaction which severely degrades thickness uniformity. The effects of each of these gas phase reactions in the upstream portion of the reaction cell were identified by spatially separating TMAA and TEG in the gas phase using the multichamber growth cell (see Figure 1). Material grown in the traditional premixed growth mode was compared with that grown in the new spatial separation injection scheme.

The TMAA-AsH_3 gas phase reaction was investigated in the conventional premixed growth mode by holding the growth temperature (670°C) and group III flux constant and varying the AsH_3 flow (V/III ratio). As shown in Figure 2, the Al content (determined by Raman scattering and confirmed by double crystal x-ray diffraction) was independent of V/III ratio at large V/III ratios (20-80) but increased as the V/III ratio was decreased below 20. This effect is attributed to the TMAA-AsH_3 reaction. At low V/III ratios, less AsH_3 is present to form involatile compounds with the TMAA and prevent its incorporation in the growing film. Although growth at low V/III ratios is more efficient, the optical quality degrades with decreasing V/III ratio. The full width at half maximum exciton linewidth of 2.2 meV compares favorably with the narrowest linewidth ever reported for AlGaAs of that composition [13], and the reduced but still clearly identifiable exciton for the sample grown with a V/III ratio of 7.5 is significant. The fact that large arsine flows are needed to produce AlGaAs with good optical quality may imply that the TMAA-AsH_3 reaction inhibits the TMAA-TEG reaction which is demonstrated below to severely degrade the quality of the AlGaAs films.

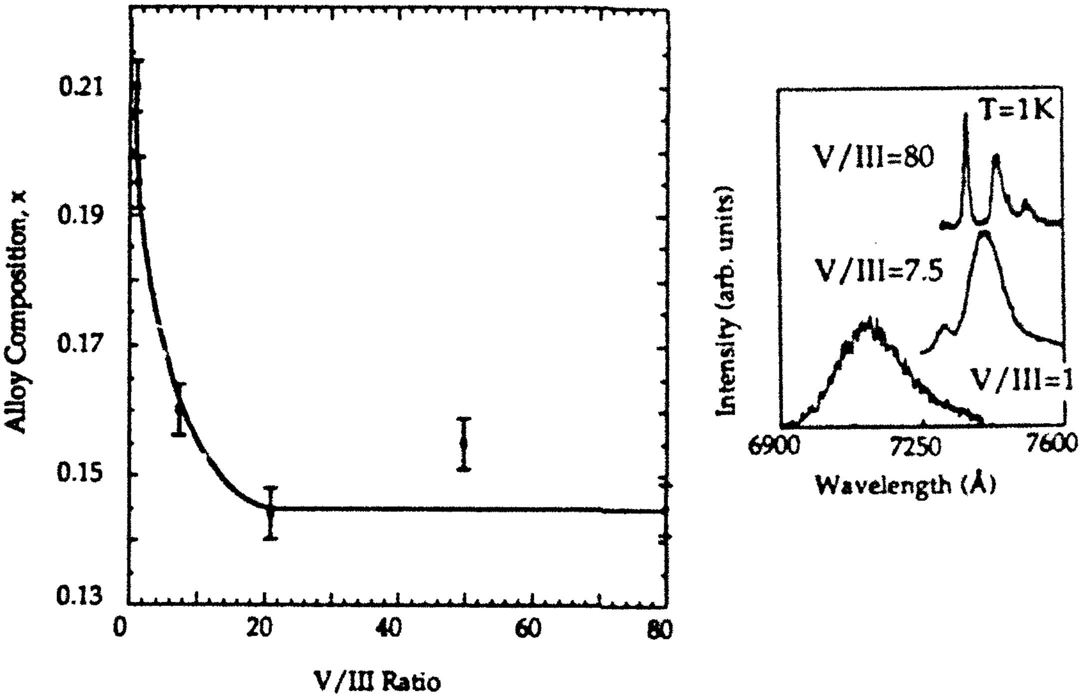


Figure 2. Dependence of Al composition on V/III ratio for constant TEG and TMAA fluxes at 670°C. The inset depicts the corresponding 1K photoluminescence spectra for several V/III ratios.

The reactant fluxes were held constant, and a relatively high V/III ratio (V/III=80) was used to eliminate the effects of arsine flows described above to investigate the effect of the TMAA-TEG gas phase reaction on the growth. Premixed growth of AlGaAs resulted in a growth rate that was approximately half that of GaAs grown with same TEG reactant flux. This was attributed to depletion of Ga in the gas phase by the TMAA-TEG reaction and confirmed by the drop in Al composition from 79% during premixed growth to 40% during spatially separated growth. Although both premixed and spatially separated grown material exhibited excellent compositional uniformity, the broad x-ray epitaxial peak characteristic of premixed grown material is believed to be due to compositional grading. Furthermore, severe thickness nonuniformity (16% over a 20 mm diameter) was characteristic of premixed growth while excellent thickness uniformity (1% over a 20 mm diameter) was obtained during spatially separated growth. Finally, the optical quality of the premixed growth as determined by low temperature photoluminescence was inferior to that of the spatially separated grown AlGaAs. A possible explanation for this effect is that in addition to the TMAA and TEG forming nonvolatile compounds which reduce the growth rate, volatile compounds are also present which participate in the growth process and incorporate non-radiative center in the epitaxial layers.

AlGaInAs Lattice Matched to InP

GaInAs and AlInAs lattice matched to InP are currently being grown by flow modulation. The structural quality and lattice parameter of the films was determined by Raman scattering and double crystal x-ray diffraction. Typical x-ray diffraction rocking curves are shown in Figure 3. The broad shoulder is indicative of graded indium content along the growth direction. Growth on InP has been suspended to modify the reactor and correct this grading problem. A Thomas Swan Epison gas concentration analyzer has been ordered and will be used to precisely monitor and control (w/ feedback) the TMIn flux out of the bubbler. Once the Epison is in place, growth of AlGaInAs will continue. After growth conditions of the mixed crystals are optimized, the study will be broadened to identify the effects of premixing and spatially separating the organometallic precursors and to compare the properties of the mixed crystal and the corresponding short-period superlattices. Finally, once the stimulated growth reactor is on-line, these basic studies will help to accelerate the selected growth of AlGaInAs on InP.

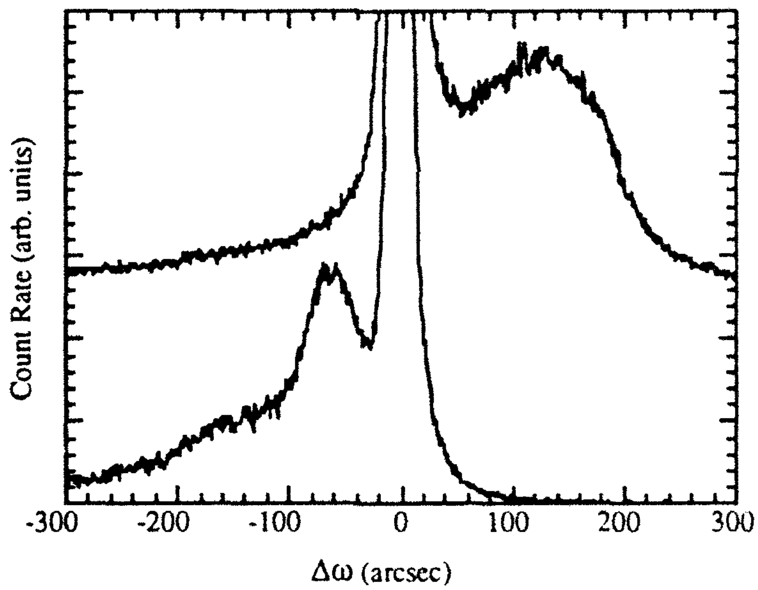


Figure 3. GaInAs on InP. Both In (bottom curve) and Ga (top curve) rich material exhibit broad epitaxial x-ray rocking curve peaks. The shoulder on the negative angle side of the each curve is indicative of graded material.

SCIENTIFIC IMPACT OF RESEARCH

The scientific impact of this research is threefold. First, the growth of high quality AlGaAs by flow modulation at low V/III ratios greatly reduces the environmental impact and life safety risk of the hydride based OMVPE growth process. Second, the investigation of the gas phase reactions of TMAA with

TEG and AsH_3 provides a much needed explanation of the growth chemistry of AlGaAs with these reactants. It should be noted that this study would have been impossible without Cornell's unique multichamber reactor. Finally, the growth of AlGaInAs alloys and stained layer superlattices on InP in the multichamber cell will allow the investigation new ordered quaternary structures for improved electronic properties.

DEGREES AWARDED

None.

REFERENCES

[1] J.R. Shealy, V.G. Kreismanis, D.K. Wagner, G.W. Wicks, W.J. Schaff, Z.Y. Xu, J.M. Ballantyne, L.F. Eastman, and R. Griffiths, "Characterization of High-purity GaAs Grown by Low-pressure OMVPE," *Inst. Phys. Conf. Ser.*, 65, 109-116 (1983).

[2] M. Ogura, Y. Ban, M. Morisaki, and N. Hase, "A New AsH3 Cracking Method for the MOCVD Growth of InGaAs," *Jpn. J. Appl. Phys..*, 22, L630-L632 (1983)

[3] Y. Seki, K. Tanno, K. Iida, and E. Ichiki., "Properties of Epitaxial GaAs Layers from a Triethyl Gallium and Arsine System," *J. Electrochm. Soc.*, 122, 1108-1112 (1975).

[4] G. Haacke. S.P. Watkins, and H. Burkhard, "Epitaxial Growth of High-Mobility GaAs Using Tertiarybutylarsine and Triethylgallium," *Appl. Phys. Lett.*, 56, 478-480 (1990).

[5] W.T. Tsar.g, "Progress in Chemical Beam Epitaxy," *J. Cryst. Growth*, 105, 1-29 (1990).

[6] S.G. Hummel, Y. Zou, C.A. Beyler, P. Grodzinski, P.D. Dapkus, J.V. McManus, Y. Zhang, B.J. Skromme, and W.I. Lee, "Characteristics of GaAs, AlGaAs, and InGaAs Materials Grown by Metalorganic Chemical Vapor Deposition Using and On-Demand Hydride Gas Generator," *Appl. Phys. Lett.*, 60, 1483-1485 (1992).

[7] W.S. Hobson, T.D. Harris, C.R. Abernathy, and S.J. Pearton, "High Quality $Al_xGa_{1-x}As$ Grown by Organometallic Vapor Phase Epitaxy Using Trimethylamine Alane as the Aluminum Precursor," *Appl. Phys. Lett.*, 58 (1991), 77-79.

[8] F. Kobayashi, S. Iio, T. Kuwahara, and Y. Sekiguchi, "Investigation of Gas-Phase Reaction between Trimethylamine Alane and Triethylgallium in Chemical Beam Epitaxy," *Jpn. J. Appl. Phys.*, 30 (1991), L1971-L1973.

[9] T. Yao, "A New High-Electron Mobility Monolayer Superlattice," *Jpn. J. Appl. Phys.*, 22 (1983), L680-L682.

[10] B.L. Pitts, D.T. Emerson, and J.R. Shealy, "Arsine Flow Requirement for the Flow Modulation Growth of High Purity GaAs Using Adduct-Grade Triethylgallium," *Appl. Phys. Lett.*, 61 (1992), 2054-56.

[11] J.R. Shealy, "Characterization of AlGaAs/GaAs Heterostructures Grown in a Multichamber OMVPE Apparatus," *J. Crystal Growth*, 87 (1988), 350-356.

[12] B.L. Pitts, D.T. Emerson, and J.R. Shealy, "Gas Phase Reactions of Trimethylamine Alane in low pressure Organometallic Vapor Phase Epitaxy of AlGaAs," *Appl. Phys. Lett.*, 62 (1993), 2054-56.

[13] D.C. Reynolds, K.K. Bajaj, C.W. Litton, P.W. Yu, J. Klem, C.K. Peng, H. Morkoc, and J. Singh, "Excitonic Photoluminescence Linewidths in AlGaAs Grown by Molecular Beam Epitaxy," *Appl. Phys. Lett.*, 48 (1986) 727-729.

JSEP PUBLICATIONS

1. B.L. Pitts, D.T. Emerson, and J.R. Shealy, "Arsine Flow Requirement for the Flow Modulation Growth of High Purity GaAs Using Adduct-Grade Triethylgallium," *Appl. Phys. Lett.*, 61, 2054-56 (1992).

2. J. Singletery and J. R. Shealy, "The Use of Ultraviolet Radiation at the Congruent Sublimation Temperature of Indium Phosphide to Produce Enhanced InP Schottky Barriers," *Journal of the Electrochemical Society*, 2961-2968 (October 1992).

3. B.L. Pitts, D.T. Emerson, and J.R. Shealy, "Gas Phase Reactions of Trimethylamine Alane in low pressure Organometallic Vapor Phase Epitaxy of AlGaAs," Accepted for publication, *Appl. Phys. Lett.* (Feb. 1993).

4. B.L. Pitts, D.T. Emerson, M.J. Matragrano, and J.R. Shealy, "The Effects of Gas Phase Reactions of Trimethylamine Alane on AlGaAs Films Grown by Organometallic Vapor Phase Epitaxy," to be submitted to *J. Crystal Growth* (March 1993).

FEMTOSECOND LASER STUDIES OF ULTRAFAST PROCESSES IN COMPOUND SEMICONDUCTORS

Task #2

Task Principal Investigator: **C. L. Tang**
(607) 255-5120

OBJECTIVE

The objective of this task is to develop new femtosecond sources and measurement techniques and to use such sources and techniques to study ultrafast processes in semiconductors and related quantum structures. On source development, current emphasis is on high repetition rate all-solid-state femtosecond sources and in extending the tuning range of such sources, in particular to the mid infrared and to the visible. On optical measurement techniques, current emphasis is on developing optical sampling techniques with femtosecond time resolution based on the up-conversion process. These sources and techniques are being applied to the study of the relaxation dynamics of non-equilibrium carriers in elemental and compound semiconductors and quantum well structures. The capture problem and the problem of tunneling of coherent wave packets in quantum wells are of particular interest at the present time.

DISCUSSION OF STATE-OF-THE-ART

Almost all the work on femtosecond optics and ultrafast processes in the past has been based on the use of the mode-locked Rh6G femtosecond dye laser as the primary source of short pulses of light. The trend recently has been to move away from the dye lasers to all-solid-state short pulse sources. CW mode-locked Ti-doped sapphire laser has been the most widely used new primary femtosecond laser source. The Ti:sapphire laser is tunable over the range of 720 nm to about 1 mm. The emphasis of our work has been to extend the useful spectral range of all-solid-state femtosecond lasers to beyond this range through nonlinear optical techniques. Very significant progress has been made in this effort during the past year and the results are discussed below in the Progress section.

In the case of femtosecond optical measurement techniques, most of the past studies of ultrafast phenomena have been based upon some sort of pump-probe measurement, including the related optical correlation spectroscopic techniques. All these techniques suffer from the fact that during the probing process, the system being measured is also disturbed to some extent. To avoid perturbing the system being measured, the time-resolved hot luminescence up-conversion technique has been developed to study the relaxation dynamics of non-equilibrium carriers in semiconductors at a number of laboratories recently, including Cornell. This technique allows optical sampling with a time resolution on the order of 50 fs of the very weak hot luminescence emitted by the carriers during the relaxation process. This technique has now been well developed where the dark noise count is down to half a photon per second and has been used successfully to yield unambiguous data on the cooling rates of hot carriers in bulk GaAs and GaAs/AlGaAs quantum wells at high carrier densities. The need is now to go beyond the GaAlAs systems and apply these techniques to other materials such as Si, porous Si, or the II-VI compounds and quantum structures.

In terms of the source wavelength, this source is ideal because photoexcitation at 2 eV can prepare carrier states up to 500 meV in GaAs and related materials and structures. With the new tunable sources now ranging from 0.5 eV to almost 4 eV, and with further extension to the 10 mm a distinct possibility [see, proposed work below], we propose to go beyond the GaAs-related materials. More specifically, we propose to study important electronic and optical materials such as Si, Ge, ZnSe, InGaAs, InGaAsP, etc., and in structures such quantum wells, quantum wires, super-lattices, and in particular porous Si. The current state-of-the-art is such that virtually no work on these materials has been done on the very short femtosecond time scale.

Si is obviously the most important electronic material. There is a clear need to explore the possibility of using some form of Si for optoelectronic applications. There have been recent reports that porous Si shows surprisingly strong photoluminescence and electro-luminescence in the visible. It has been suggested that the luminescence has something to do with reduced dimensionality and quantum confinement, but there is no consensus on either the experimental results or their interpretation. Femtosecond laser study is clearly another avenue for identifying the origin of the luminescence and clarifying the physics. We have recently developed a capability [in collaboration with IBM Yorktown Heights through an IBM Fellowship student working both at Yorktown Heights and at Cornell] to prepare reproducible porous Si samples. We hope to have preliminary results on the dynamics of photoexcited carriers in porous Si in the near future. At this point, we do not know what to expect, although we have some conjectures.

There has been a great deal of interest in ZnSe semiconductor lasers in the blue-green region recently. However, very little is known about the relaxation dynamics of hot carriers in ZnSe. A number of the important relaxation processes and the corresponding rates are not known. To develop room-temperature cw blue semiconductor lasers, which is a major direction of semiconductor laser research today, it is essential that these parameters be known. Our recently developed femtosecond laser in the blue is uniquely suited for such studies. We are preparing for experimental studies on this material in the femtosecond time domain.

Ternary and quarternary compound semiconductors are of importance to long wavelength (1.3 μm and 1.5μm) lasers. Ultrafast studies in these materials have been hampered by the lack of suitable ir femtosecond sources. The ir femtosecond OPOs developed in our laboratory are uniquely suited for studying these materials. The relaxation dynamics in these materials are more complicated than those in GaAs or AlGaAs. Understanding of the relaxation dynamics related to the Auger processes in InGaAs and InGaAs/InP quantum wells are of particular importance for long wavelength lasers; further work is needed.

In terms of different structures with reduced dimensionalities, planar 2-D GaAs/AlGaAs quantum wells have been studied extensively and are relatively well known. Because of the difficulties in fabricating quantum wires and boxes with dimensions on the order of 10 nm or less, much less is known about these structures. There is, however, a great deal of interest in such structures for a variety of electronic and optic device applications. J. R. Shealy has been able to fabricate quantum wires on the order of of 30x50 nm using MOCVD overgrowth on groves etched in GaAs by focused ion beam. With better control in the growth and ion beam etching processes, fabrication of wires of dimensions small enough to exhibit quantum effects should be possible. Ultrafast laser studies will elucidate the relaxation dynamics of hot carriers in such quasi-one-dimensional quantum structures. Because of the shifts in the excitation energies of various relevant states due to quantum confinement, wavelength tunability is, again, an important consideration in photoexcitation studies in such structures.

In all these studies, the significantly increased power of the new femtosecond OPO sources will greatly increase the sensitivity of time-resolved hot luminescence spectroscopy based upon the nonlinear optical upconversion process coupled with photon-counting. Since the signals in most such pump-and-probe measurement techniques are proportional to at least the square of the intensity of the fs pulses, an increase of 2 orders of magnitude in the intensity of the pulses as has been achieved with our new fs OPOs means an increase of at least 4 orders of magnitude in the sensitivity as compared to the Rh6G-based measurement systems currently in use. This improvement in sensitivity will vastly increase our capabilities to study various ultrafast processes.

PROGRESS

New Developments in Femtosecond Sources

A principal direction of research in the field of femtosecond optics and ultrafast processes is to extend the wavelength range of the femtosecond sources and, as a consequence, expand the types of ultrafast processes and materials that can be studied. In our last three-year proposal, one of the main objectives is to extend the femtosecond laser source into the uv through various nonlinear optical sum-frequency processes and into the ir through the optical parametric processes. We have been very successful in both areas.

In the uv, we have previously demonstrated [1] the shortest pulse (43 fs) at 315 nm, but the power was relatively low by today's standards, in the 10 mW range, but quite respectable by the standards prior to our most recent work in the last few months. Most recently we demonstrated [2] the generation of high-repetition-rate femtosecond pulses in the blue for the first time by intracavity doubling of a mode-locked Ti:sapphire laser using β-BaB_2O_4. To reduce the pulse broadening effect of group-velocity mismatch, an extremely thin β-BaB_2O_4 crystal was used. Over 800 mW of 430 nm light is produced at 72 MHz repetition rate and pulse widths as short as 54 fs are achieved.

From the deep red to the mid ir, we have demonstrated the first broadly tunable femtosecond optical parametric oscillator [3]. With intracavity pumping by a Rh6G dye laser, tunable femtosecond pulses from 720 nm to nearly 3.5 μm have been achieved at a pulse repetition rate of 80 Mhz and power level of a few mW [3-5]. Dramatically improved power levels have been achieved most recently using external-cavity pumping by Ti:sapphire lasers. Record femtosecond Ti:sapphire laser output powers up to 3.8 W have been achieved recently in our laboratory. Tunable femtosecond pulses in the multi-hundred mW range with a total conversion efficiency of 55% from femtosecond optical parametric oscillators (OPO) have been demonstrated most recently [6-10]. While the intracavity pumped fs OPO is extremely difficult to operate, the external-cavity Ti:sapphire laser pumped OPOs are relatively easy to construct and reproduce. Since our initial work, there has been extensive interest on the part of various laser companies to commercialize the technology. Technology-transfer agreements have been signed with one of the largest laser companies, Spectra-Physics. With the general availability of such powerful tunable fs sources, we expect that there will be rapid advances in the field of ultrafast studies over spectral regions that had been out of reach before. Figure 1 shows a summary of the current status of tunable high repetition rate femtosecond sources and the pump sources available. Apart from the two pump sources, Rh6G dye laser and the Ti:sapphire laser, all the other sources were first demonstrated in our laboratory under primarily JSEP sponsorship.

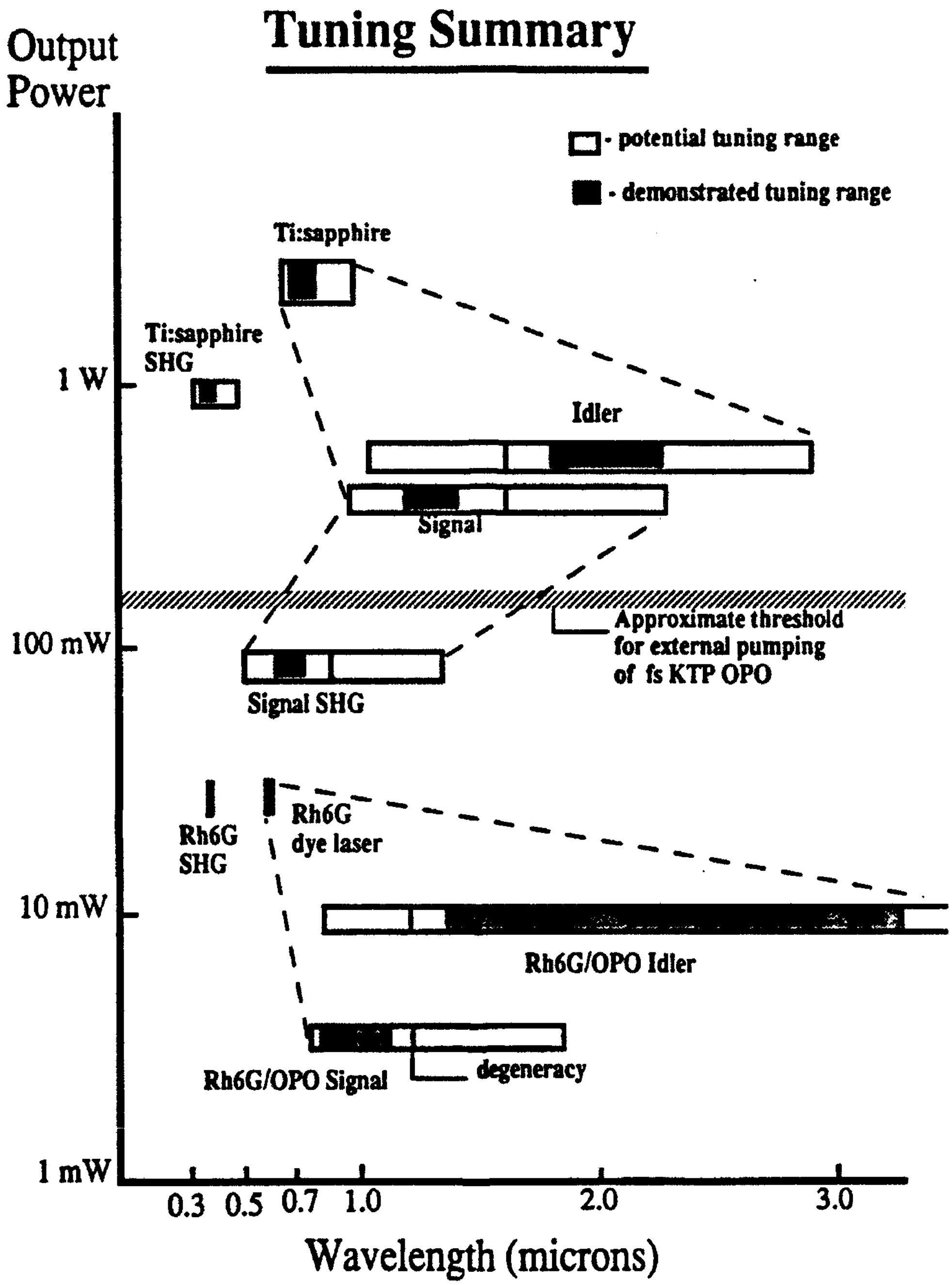


Figure 1

Ultrafast Relaxation Dynamics in Semiconductors

The investigation of hot-carrier relaxation in both quantum well (QW) and bulk semiconductor structures has been a topic of great importance and considerable interest. An investigation of the hot-carrier relaxation in GaAs/(Al,Ga)As quantum wells and bulk GaAs in the high-carrier-density limit has been carried out [11-13]. Using a time-resolved luminescence nonlinear optical up-conversion technique with ≤ 80-fs temporal resolution, carrier temperatures are measured in the 100-fs-to-2-ns range. The results show unambiguously that the hot-carrier cooling rates in the quantum wells are significantly slower than in the bulk for carrier densities greater than 2×10^{18} cm^{-3}. A comparison is made with previous publications to resolve the controversy concerning the difference in cooling rates in quasi-two- and three-dimensional systems. These results are important for applications related to injection of hot carriers in high speed electronic devices and high efficiency solar cells.

SCIENTIFIC IMPACT OF RESEARCH

The femtosecond sources and measurement techniques developed should be of great use to others in the scientific community. The results obtained on the dynamics of nonequilibrium carriers in III-V compounds and structures are of fundamental importance to the understanding of the physics and the design of ultra-high speed semiconductor electronic and optical devices.

DEGREES AWARDED

1. Wayne Pelouch
   "Multi-wavelength Ultrafast Source Development and Spectroscopy"
   Ph.D. in Applied Physics, January, 1993.

REFERENCES

[1] D. C. Edelstein, E. S. Wachman, L. K. Cheng, W. R. Bosenberg, and C. L. Tang, "Femtosecond Ultraviolet Pulse Generation in β-BaB_2O_4," *Appl. Phys. Lett.*, 52, 2211-2213 (June 1988).

[2] R. J. Ellingson and C. L. Tang, "High-repetition rate femtosecond pulse generation in the blue," *Optics Letters*, 17, 343-346 (March 1, 1992).

[3] D. C. Edelstein, E. S. Wachman, and C. L. Tang, "Broadly tunable high repetition rate femtosecond optical parametric oscillator," *Appl. Phys. Lett.*, 54, 1728-1730 (1 May, 1989).

[4] E. S. Wachman, D. C. Edelstein, and C. L. Tang, "Continuous-wave mode-locked and dispersion-compensated femtosecond optical parametric oscillator," *Optics Letters*, 15, 136-138 (January 15, 1990).

[5] E. S. Wachman, W. S. Pelouch, and C. L. Tang, "CW femtosecond pulses tunable in the near- and mid-infrared," *J. Appl. Phys.*, 70, 1893 (1 August, 1991).

[6] W. Pelouch, P. Powers, and C. L. Tang, "Ti:sapphire laser pumped high-repetition rate femtosecond optical parametric oscillator," Post Deadline paper, CLEO '92, May 14, Anaheim, CA.

[7] W. S. Pelouch, P. E. Powers, C. L. Tang, "Ti:sapphire laser pumped, high repetition rate femtosecond optical parametric oscillator," Post Deadline talk, IQEC Vienna (June 18, 1992).

[8] W. S. Pelouch, P. E. Powers, and C. L. Tang, "Ti:sapphire -pumped, high-repetition-rate femtosecond optical parametric oscillator," *Optics Letters* 17 (15), 1070-1072 (August 1, 1992).

[9] C. L. Tang, "Optical Parametric oscillators," Invited talk, International School on Nonlinear Optics and Optical Physics, Capri, Italy, (June 1-5, 1992).

[10] C. L. Tang, "Renaissance of optical parametric oscillators and amplifiers," invited talk at Nonlinear Optics: Materials, Fundamentals, and Applications, Maui, Hawaii (August 17-21, 1992).

[11] W. S. Pelouch, R. J. Ellingson, P. E. Powers, and C. L. Tang, "Comparison of hot-carrier relaxation in quantum wells and bulk GaAs at high carrier densities," *Phys. Rev. B.*, 45., 1450-1453 (15 January, 1992).

[12] W. S. Pelouch, R. J. Ellingson, P. E. Powers, C. L. Tang, D. M. Szmyd, and A. J. Nozik, "Investigation of hot carrier relaxation in quantum well and bulk GaAs at high carrier densities," *Proceedings of 7th International Conference on Hot-Carriers in Semiconductors*, Nara, Japan (July 1991).

[13] "Dynamics of hot carriers in quantum wells and bulk GaAs at high carrier densities: femtoseconds to nanoseconds," invited talk, SPIE Conference on Ultrafast Laser Probe Phenomena in Bulk and Microstructures in Semiconductors and Superconductors, Somerset, NJ (22-26 March, 1992).

JSEP PUBLICATIONS

1. R. J. Ellingson and C. L. Tang, "High-repetition rate femtosecond pulse generation in the blue," *Optics Letters*, 17, 343-346 (March 1, 1992).

2. W. S. Pelouch, P. E. Powers, and C. L. Tang, "Ti:sapphire -pumped, high-repetition-rate femtosecond optical parametric oscillator," *Optics Letters* 17 (15), 1070-1072 (August 1, 1992).

ULTRAFAST INTERACTIONS OF CARRIERS AND PHONONS IN NARROW BANDGAP SEMICONDUCTOR STRUCTURES

Task #3

Task Principal Investigator: **Clifford Pollock**
(607) 255-5032

OBJECTIVE

This task represents the device measurement portion of the JSEP program. We have been studying ultrafast relaxation processes in narrow bandgap semiconductors using tunable ultrashort optical pulsed lasers to determine carrier dynamics. These measurements have been made in bulk materials using thin films. Our goal is to generate solid experimental evidence of the carrier scattering processes that occur for energies near the bandedge of the semiconductor. Data has been used to test modelling calculations performed by Prof. Krusius' group. In this interaction, we hope to both improve the quality of the data we acquire, and to test the models developed by the Krusius group through rigorous comparison of data with modelling results. We also plan on improving our equipment by extending the tuning range and pulse performance of our probe lasers.

DISCUSSION OF STATE-OF-THE-ART

Studies of hot carrier relaxation have been carried out in GaAs and AlGaAs for several years. Only recently has there been the ability to study narrow bandgap semiconductors with femtosecond probes. Roskos et. al [1] have reported on the cooling of photoexcited carriers in undoped InGaAs samples similar to the films we are studying, only their work was done at 30 °K, and with significantly higher probe energy. Using a dye laser operating at 1265 nm, they created hot carrier populations approximately 200 meV above the conduction band minimum. They observed recovery of the transient absorption to occur with time constants 1 and 10 psec representing depletion of the electron and heavy hole states, respectively. Being at a higher excess energy than our measurements, the observed relaxation lifetimes are not directly comparable between their experiments and our experiments, as their carrier distribution is influenced more strongly by Fröhlich scattering. Our experiments provide a nice complement to this work, giving a picture of carrier thermalization mechanism's weights as one approaches the bandedge.

Nunenkamp et. al [2] describe an experiment similar to ours in a different material, AlGaAs. They probed carrier relaxation near the band-edge of the material using a tunable femtosecond laser. They developed a model including excitonic effects, however their data was taken at 30 °K compared to our 300 °K. Significantly, they found band-gap renormalization essential to fit the data to their simulations. This is consistent with the results we have obtained in our collaboration with the modelling efforts of Krusius' group.

PROGRESS

Progress in the last year is summarized below. We have worked closely with Krusius in measuring the carrier relaxation time in bulk GaInAs/InP layers. The data from our measurements have been used to test the models developed by Krusius and Bair, and results from their fitting has been used to modify our experiments. Emphasis in our group has been on the experimental side of the problem, especially with obtaining exceptionally high signal-to-noise data which can be used to test the models.

Several layers of InGaAs were obtained from the Cornell School of Electrical Engineering MBE facility, and were probed with our tunable color center laser. We studied three different thickness layers: 0.5 μm, 1 μm, and 3 μm. The color center laser we used was able to generate 100 fsec pulses over the 0.75 → 0.85 eV range, which is near the bandgap of $In_{0.53}Ga_{0.47}As$ lattice matched to InP. A 100 fsec pulse from the laser was used to create carrier densities ranging in the 10^{17} range. Using the equal pulse autocorrelation technique developed by Tang [3], a second pulse probed the recovery of the optical absorption as a function of time by entering the sample at the same point, but at a delayed time. As the carriers recover back to their equilibrium distribution, the optical absorption returns to its original value.

We observe changes in transmission on the order of approximately 1% for a carrier density change of $10^{17}/cm^3$. A typical data trace is shown in Fig. 1. The figure shows the measured autocorrelation transmission trace from a 0.5 μm thick InGaAs layer probed at 1.596 μm. The inner trace is the autocorrelation of the probe laser pulse. The quality of the data can be rapidly judged by the experimentalist through the symmetry of the plot. This data was fed to Krusius for analysis, and is discussed in his task. A joint publication is under preparation.

Fitting this data to simple single and double exponential decay models showed there were some trends with increased probe energy. Such data is shown in Fig. 2. However based on simulations and discussions with Krusius, it was decided that the decay process was too complicated to be described in terms of one or two lifetimes.

Initial comparison of our temporal data to the simulations of Krusius showed that our raw data was not sufficiently noise-free to make substantive claims as to the accuracy of the simulation. It was determined that the absolute reproducibility (not simply Signal-to-Noise ratio) had to be better than 1 part in 100. This posed quite a problem, as there are dozens of experimental effects

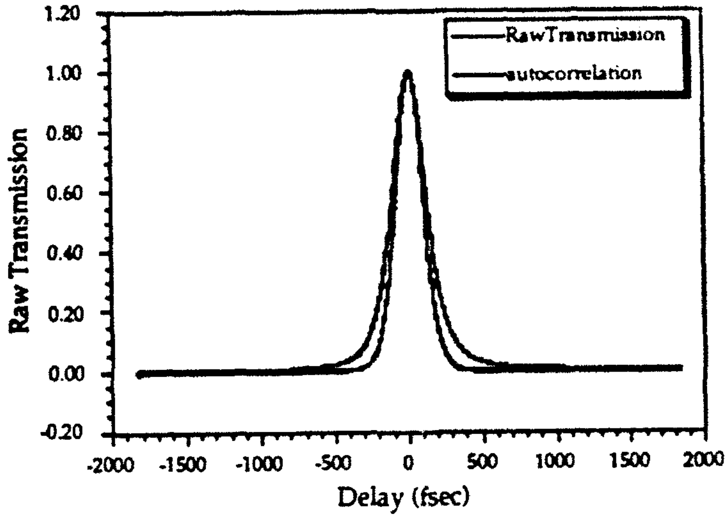


Figure 1. An equal pulse autocorrelation trace taken with InGaAs at a wavelength of 1.596 μm. The inner trace is the autocorrelation of the laser pulse; the outer trace is the transient transmission of the semiconductor sample.

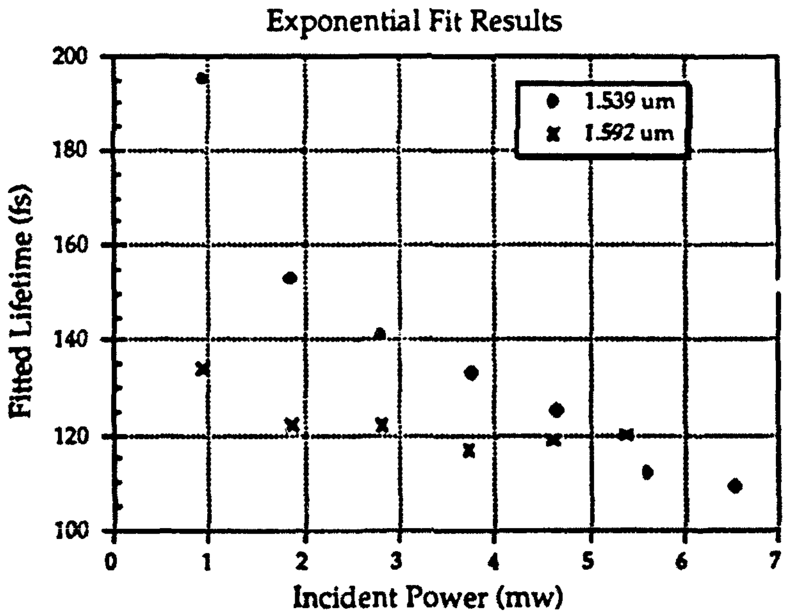


Figure 2. Exponential fits to the observed transient decay of the transmission, as a function of excess energy and carrier density.

which can systematically alter the overall shape by such magnitudes. We invested a significant effort into reducing systematic errors, by linearizing detectors, optimizing the PZT transducers and galvodrivers, and by carefully coating and selecting all optics to avoid unexpected reflections. The absolute reproducibility (measured from week-to-week, not run-to-run), has been improved dramatically to about 1%. An example of two data runs taken about four months apart, but under experimentally identical conditions, is shown in Fig. 3.

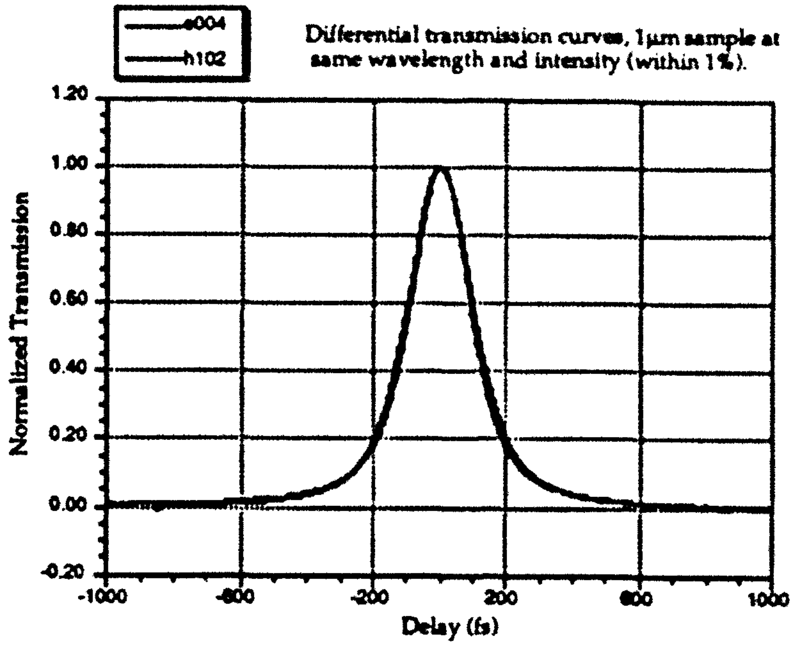


Figure 3. Two sets of data taken under nearly identical conditions, but at two different times. The absolute agreement of obtained data is within 1% from one run to the next.

This summer (1992), Krusius and his students modified their model to include bandgap renormalization effects. This modification has had a dramatic impact on the quality of the fit between data and model. We now find excellent agreement in the 100-400 fsec region, while there is still some disagreement for data beyond 500 fsec. We are repeating some key measurements at this time to determine whether the model or the data is at fault. This collaboration has finally paid off in terms of identifying the major physical processes that occur in hot carrier relaxation in InGaAs.

The tuning range of the color center laser used in these measurements coincides well with the bandgap of InGaAs, but it does not have sufficient tuning range to excite electrons more than about 2 LO phonons above the bandgap. Based on work reported by Petricevic [4] we built a forsterite laser that operates from 1.2 $\rightarrow$ 1.32 μm, or 0.93 $\rightarrow$ 1.03 eV. This energy range allows the excitation of carriers to about 250 meV above the conduction band edge in the InGaAs samples we are using, but does not lift them high enough to allow inter-valley scattering. This probe wavelength is similar to that used by Roskos et. al [1], except that we have much more power and a shorter pulse width, with a broader tuning range. We can stably generate 48 fsec pulses at an 80 MHz rate with an average power of 380 mW. We recently frequency doubled this laser to generate 25 mW of femtosecond pulses at 630 nm. Work is underway to apply this new source to some unique two wavelength pulse probe measurements. A paper has been submitted on the laser result [5], and another is in preparation on the frequency doubling.

SCIENTIFIC IMPACT OF WORK

The major impact of this work lies in the identification of carrier-carrier scattering as the dominant scattering process near the bandedge of InGaAs. The carrier densities we are exploring are similar to those used in semiconductor quantum well lasers, where high speed carrier capture is critical. We can now identify characteristic lifetimes for various physical scattering processes in this material system.

Secondly, this data has been used to critically test the Monte-Carlo simulations of the Krusius group, and have provided a motivation to modify and improve the model. Similarly, the modelling efforts have focussed our attention on eliminating systematic errors, and we are confident that the measurements we report represent the best data to date on relaxation phenomena near the bandedge of InGaAs.

DEGREES AWARDED

1. Timothy Carrig
   "Characterization of New Color Center and Transition Metal Ion Lasers"
   Ph.D., Applied and Engineering Physics, August 1992

REFERENCES

[1] H. Roskos, B. Rieck, A. Seilmeier, W. Kaiser, and G. Baumann, "Cooling of Photoexcited Carriers in Undoped and n-doped GaInAs Studied Within the First Few Picoseconds," *Phys. Rev. B*, 40, 1396-1399 (1989).

[2] J. Nunnenkamp, J. Collet, J. Klebniczki, J. Kuhl, and K. Ploog, "Subpicosecond Kinetics of Band-edge Absorption in AlGaAs," *Phys. Rev. B*, 43, 14047-14054 (1991).

[3] A. J. Taylor, D. J. Erskine, and C. L. Tang, "Equal Pulse Correlation Technique for Measuring Femtosecond Excited State Relaxation Times," *Appl. Phys. Lett.*, 43, 989 (1983).

[4] V. Petricevic, S. K. Gayen, R. R. Alfano, K. Yamagishi, H. Anzai, and Y. Yamaguchi, "Laser Action in Chromium-doped Forsterite," *Appl. Phys. Lett.*, 52, 1040-1042 (1988),

[5] A. Sennaroglu, C. R. Pollock, and H. Nathel, "Regeneratively Initiated, Self-modelocked, Cr:forsterite Laser," submitted to *Optics Letters* (January 1993).

JSEP PUBLICATIONS

1. A. Sennaroglu, T. J. Carrig, and C. R. Pollock, "Femtosecond Pulse Generation by Using an Additive-pulse Mode-locked Chromium-doped Forsterite Laser Operated at 77 K," *Optics Letters*, 17, 1216-18 (1992).

2. D. Cohen and C. Pollock, "Femtosecond Electron Relaxation in InGaAs Lattice-matched to InP," to appear in SPIE.

3. J. E. Bair, D. Cohen, J. P. Krusius, and C. R. Pollock, "Band Renormalization and Dynamic Screening in Near Bad Gap Femtosecond Optical Probing of InGaAs," manuscript submitted for publication to *Physical Review Letters*, April, 1992.

4. A. Sennaroglu and C. R. Pollock, "Generation of 48 fsec Pulses and Measurement of Crystal Dispersion by using a Regeneratively-initiated Self-mode-locked Chromium-doped Forsterite Laser," accepted for publication in *Optics Letters*.

5. A. Sennaroglu, C. R. Pollock, and H. Nathel, "Generation of Tunable Femtosecond Pulses in the Red by Frequency Doubling a Mode-locked Cr:forsterite Laser," submitted for publication to *Optics Letters*.

6. D. Cohen and C. Pollock, "Experimental Studies of Fast Carrier Relaxation Near the Bandedge of InGaAs," paper in preparation.

FEMTOSECOND CARRIER TRANSPORT AND OPTICAL INTERACTIONS IN COMPOUND SEMICONDUCTOR HETEROSTRUCTURES

Task #4

Task Principal Investigator: **J. Peter Krusius**
(607) 255-3401

OBJECTIVE

The objective in this work unit is to explore non-equilibrium carrier processes governing electron and hole transport and optical interactions in inhomogeneous compound semiconductor heterostructures theoretically. Electron and hole interactions among themselves, the semiconductor lattice, optical fields, and external electric fields are described using self-consistent ensemble Monte Carlo formulations. This work is done collaboratively with femtosecond optical measurements and materials growth efforts. In the area of femtosecond optical probing with tunable lasers, joint work is performed together with C. Pollock's research group in order to design samples, optical experiments, analyze measured data, and extract microscopic information of femtosecond carrier processes.

DISCUSSION OF STATE-OF-THE-ART

The relaxation of carriers excited by ultra-short optical pulses has been intensely studied both experimentally and theoretically for several years. Despite this a full understanding of the complex carrier dynamics in these highly non-equilibrium situations is still incomplete. On the theoretical side, current Monte Carlo models of carrier relaxation have achieved considerable success in explaining many qualitative features of experimental observations [1, 2]. However, a great deal of uncertainty remains. At the heart of these uncertainties lies the role of the Coulomb potential in the carrier-carrier interaction. This interaction expresses itself in a number of important phenomena. Most familiar is the free carrier dielectric function, which affects all carrier scattering processes through screening, but the Coulomb potential also enters the problem through carrier-carrier scattering and two carrier correlation effects, band renormalization, and the Coulomb enhancement of optical absorption.

Free carrier screening and carrier-carrier scattering have received a large amount of attention. Until recently models universally assumed that free carrier screening can be adequately described by using long wavelength static approximations. However, recent results have indicated that carrier-carrier scattering may be seriously underestimated in a static screening limit [3-6]. Since the two most important scattering processes in compound semiconductors, polar optic phonon scattering and carrier-carrier scattering, are both heavily dependent on free carrier screening, this issue is critical in the understanding of the role of these scattering mechanisms. The current consensus is that static screening is generally an inadequate approximation. A number of approximations, such as plasmon-pole are available, to approximate the full dielectric function, but these assume quasi-equilibrium conditions, which is not generally present on the femtosecond timescale in these experiments. Progress in developing more general methods has been made through a joint ensemble Monte Carlo/molecular dynamics approach [7], which has some success in correlations with measured data. However, this method appears to be limited to homogenous systems due to limitations arising from the size of the area that can be simulated, and the method does not naturally provide the free carrier dielectric function required for calculating the effect of screening on other scattering processes and Coulomb correlation effects. Thus it is unlikely that this method can be applied more widely to the modeling of other highly non-equilibrium phenomena, which are important in describing high speed compound semiconductor devices. Even more recently an extension of the standard Monte Carlo method that includes the free carrier polarization has been developed. However, this approach has to be classified as a radical departure from standard Monte Carlo, and hence it is not clear whether it retains the many features that make Monte Carlo widely used [8].

Band gap renormalization and Coulomb enhancement are processes that have received relatively little attention in theoretical studies. Most detailed simulations have completely neglected these effects [1,2,7] The few efforts that have included them typically use simple qualitative models. Nonetheless, a number of experiments have shown effects attributed to these processes [9-12]. Since these processes are known to decrease in importance away the band edge, experiments far from the band edge can perhaps be understood without them. However, their effect in the near bandgap regime, where the present femtosecond probing experiments are performed, is expected to be significant, and thus careful investigations must be launched.

The role of band renormalization and Coulomb enhancement effects in the relaxation of near band gap carriers on the femtosecond timescale femtosecond has been the main focus in this task during the past year. No major new issues have hence been explored in the non-equilibrium transport area.

PROGRESS

A dual carrier ensemble Monte Carlo program (OPTMC) capable of simulating near band gap femtosecond optical pump-probe experiments in $In_{0.53}Ga_{0.47}As$ thin films on transparent InP substrates has been developed. This formulation describes all scattering rates, including carrier-carrier, and an accurate **k.p** band structure for both electrons and holes. Carrier transport along the normal to the thin film is included through a self-consistently calculated electric field. Screening was initially handled self-consistently through the long wavelength limit of the Lindhard dielectric function. Optical absorbtion was calculated using the instantaneous carrier distribution function. This simultation code has been used in collaboration with femtosecond laser probing experiments performed by C. Pollock's group in task #3 of this JSEP program to explore the femtosecond dynamics of carries excited within 100 meV of the band gap.

A great a deal of effort has been expended, both experimentally and theoretically, in this area in recent years. However, our effort is in many ways unique. Most other groups have utilized the capabilities of femtosecond CPM lasers in the AlGaAs/GaAs system. The fact that this is not a tunable laser and has a fixed photon energy of about 2 eV restricts excitation of carriers to states very high in the band. Carrier relaxation dynamics in this energy range are dominated by intervalley transfer and hence most of the work has focused on determining these rates. In the present work the availability of the unique tunable femtosecond color-center laser developed by C. Pollock has been exploited. This laser enables $In_{0.53}Ga_{0.47}As$ thin films to be excited over a range from the band edge to 100 meV above it. Thus the complications of intervalley transfer are avoided and a very different regime of carrier dynamics can be explored.

Due to the critical role of screening in the formulation, it was concluded that the consequences of the approximations made in the static screening model needed to be examined in greater detail in this femtosecond domain. To this end a new more accurate dynamic screening model has been developed and implemented in the Monte Carlo code. The new model is derived from the Lindhard dielectric function and fully incorporates the energy and wavelength dependence of free carrier screening. The use of an approximate parabolic band structure and the neglect of the anisotropy in the bands, electron and hole carrier distributions, and the dielectric function were the only simplifications made. Comparison of the dielectric functions derived from the two models resulted in a number of critical differences. With dynamic screening the dielectric function falls of rapidly with energy resulting in significantly reduced screening for inelastic scattering processes. The dynamic dielectric function also falls off faster at small wavelengths resulting from a more accurate inclusion of the finite size of the carrier wave packets compared to the static long wavelength approach. The plasma resonance at high energy and long wavelength is also properly included in the dynamic model resulting in the enhancement of some high energy carrier-carrier interactions. Further, at early times in the simulations (within 100 fs of the center of the exciting pulse) the extremely non-equilibrium distribution function results in an unexpected plasma resonance with the heavy holes taking the form of a narrow ridge running from small energies and large wavelengths to large energies and small wavelengths diagonally across the dielectric function. We have found that the dielectric response of the holes is not fully integrated into the total dielectric function and as long as the holes are constrained to this narrow energy region the plasma will exhibit a second plasmon spectrum due to the heavy-holes. This results from the fact that the heavy holes are excited into an extremely small energy range initially (2-3 meV). As the heavy hole distribution relaxes toward equilibrium the resonance shrinks in size and eventually disappears as the holes are dispersed throughout the band. The significance of this effect for the simulations discussed here is that the resonance crosses a region of momentum-energy space important two a number of carrier-carrier scattering processes and results in a significant enhancement of these processes at ultrashort times.

The new dynamic screening model resulted in dramatically faster relaxation times when compared to the static screening model. This is primarily the result of two effects. Within this formulation scattering processes involving optical phonons are largely unscreened because of the large energy and momentum transfers involved. Thus the resulting electron-optical phonon scattering rates are nearly doubled, while heavy hole-optical phonon rates show only a small increase because of the extremely large transfers of momentum involved. As anticipated the carrier-carrier scattering rates were greatly enhanced within the new screening model. At longer times (beyond 100 fs) when the carrier distribution functions are approaching an equilibrium distribution, the electron-electron scattering rates are found to be approximately doubled while the heavy hole-heavy hole rate shows a 10-15% increase. The effect on heavy hole-heavy hole scattering is weaker, because it results in smaller energy transfers but large momentum transfers. This heavy hole scattering rate benefits less from the energy dependence of the dynamic dielectric function and at the same time is more weakly screened initially. Electron-hole scattering is negligibly affected because it is approximately elastic. At early times, interaction with the heavy-hole resonance results in a spectacular enhancement in all carrier-carrier scattering rates except the heavy hole-heavy hole interaction. Thus we find that the use dynamic screening results in a large increase in the magnitude of the dominant scattering rates, a finding important for many non-equilibrium transport conditions. Here the effect of dynamic screening is reflected in the decrease of the relaxation time extracted from the simulated probe transmission, from 200 to 145 fs for a pulse consisting of 810 meV photons with an intensity of 5.0×10^{13} eV/cm^2. This exhibits itself in the distribution functions as an extremely rapid appearance of carriers at the band edge and an extremely rapid washing out of the excitation peak structure as well as the elimination of distinct phonon replicas. Optical phonon scattering is still an important scattering process, but carrier-carrier scattering has increased greatly in its relative importance and is now found to contribute significantly to the total scattering times. Further details of this effect are discussed in JSEP publication [3].

Investigations have been completed examining the effect of the energy of the exciting photons and the intensity of the exciting pulse and comparing these to the results from static screening. In both cases the results are qualitatively similar to those for static screening although with significantly reduced relaxation times. The extracted relaxation times still decrease with increased photon energy, however, the distinct shoulder at the first phonon threshold has disappeared. This is indicates the increased importance of carrier-carrier scattering, which serves to wash out sharp features. The relaxation times are still found to increase with pulse intensity although to a lesser degree. In the dynamic model enhanced screening with increased carrier density plays only a small role with the dominant effect being the degeneracy in the conduction band, which reduces electron scattering rates and consequently increases relaxation times.

New attempts to correlate with recent measured data from the group of C. Pollock have been undertaken with the new dynamic screening model. Significant gains in describing carrier behavior have been made with relaxation times derived from the dynamic model. The resulting simulated results are much closer to those determined from measured data than the those extracted from the static screening model. Still, the differences are larger than could be removed by tuning the model parameters and research has been continued to determine what significant physical process is responsible.

Consequently two additional physical processes, band renormalization and the Coulomb enhancement of the optical matrix elements, were investigated. Band renormalization has been included within the quasi-static approximation of Haug and Schmitt-Rink [13]. This formulation allows the renormalization for both the valence and conduction bands to be calculated from the carrier distribution functions and free carrier dielectric function. As implemented in OPTMC the renormalizations are calculated at each ti· ꞉ step self-consistently using the instantaneous values of these functions available within the simulation. The result is then included in the simulation through a rigid band shift for each band. The conduction and valence bands are each shifted in energy by an amount calculated at their respective band edges (G point). The effect of the band gap shrinkage is fully included in the model of optical interactions. Both the shift in energy of carriers excited by photons of a given energy and the decrease in the photon absorption length are included. The variation in the position of the bands across the thickness of the sample arising from variations in the carrier density are included within the same interpolation scheme used for the carrier distribution functions and dielectric function. Transport effects arising from this spatial dependence are dealt with by including the renormalization energy for the band as an effective potential in the semi-classical equations of motions for the carriers in that band.

The effect of band renormalization on the simulated results was found to be quite dramatic, and for the range of photon energies of interest here (20-60 meV above the band) of even greater importance than dynamic screening. For example, simulated pulse-probe experiments performed assuming an excitation pulse energy 1.9 x 10^{13} eV/cm^2 and a photon energy of 0.87 eV (37 meV above the band show a 199 fs effective relaxation time over the range 200-800 fs of the delay between the pulse and probe pulse for a simple model with only static screening. With dynamic screening the result is 169 fs, while for static screening with band renormalization, and dynamic screening with band renormalization the results are 139 and 100 fs respectively. Thus the band renormalization is the larger of the two effects. These results can be compared to the experimental result for a dual-pulse-correlation experiment of 116 fs. In this case there is a reduction in the band gap of approximately 13 meV, which is a significant portion of the initial excess energy of 37 meV, and results in a significant change in the relationship between the pump pulse and the excited carrier distribution. Not surprisingly band renormalization takes on increasing importance when the band edge is approached. Details of these results are available in JSEP publications [3, 4].

A detailed comparison of the results of this model with the experiments of C. Pollock's group indicates that the simulation now reproduces measured optical transmission data as a function of probe delay fairly accurately. The initial transients are well reproduced with effective relaxation times agreeing with experiment within 10-15% without any adjustable parameters. It seems clear that this model contains the physics necessary to provide a fundamental understanding of these experiments. However some details still need to be addressed for a full quantitative understanding. The simulation does not generally reproduce the long time measured transients (larger than 600 fs) accurately and as a result does not describe the small variations seen as a function of photon energy and excitation pulse intensity, which are largely as result of changes at these late times. It should be emphasized, however, that the remaining problems are small details and that the basic response is well reproduced. We believe that this model provides an adequate basis for the fundamental understanding and analysis of these femtosecond experiments.

Some of the remaining discrepancy is likely the result of significant uncertainty in some of the materials parameters and can likely be removed by fine tuning the model. However, it has become clear that full agreement with experiment cannot be obtained by tuning within the range of material parameters, which is physically reasonable. Preliminary results do indicate that much of the remaining discrepancy can be addressed through the inclusion of the Coulomb enhancement of the optical transition elements. In this process the Coulomb interaction between the excited electron-hole pair enhances the effective transition rate. This process becomes increasingly important as the band edge is approached and can be included as a multiplicative factor in the optical transition matrix element. It has been implemented within OPTMC through a phenomenological formulation developed by Banyai and Koch [14]. This formulation allows the inclusion of the effects of free carrier screening on the enhancement and is implemented self-consistently within the Monte Carlo method. Early results indicate that this improvement will significantly reduce remaining discrepancies with experiment. Specifically the accuracy of the long time transients has been improved and correct trends as a function of pulse intensity have been restored. These results will be discussed in detail in JSEP publication [5].

SCIENTIFIC IMPACT OF RESEARCH

Free carrier screening and band renormalization have been shown to be critical processes (beyond the usual scattering mechanisms) for understanding carrier relaxation on the femtosecond timescale in the near bandgap regime. We think that the Coulomb enhancement may also be an important process in this regime. In order to quantitatively investigate the effect of these processes new and unique techniques have been developed to allow there inclusion in the standard Monte Carlo method with a minimum of assumptions. Static screening appears inadequate in the modeling femtosecond carrier relaxation. The inclusion of dynamic screening provides a much more accurate understanding of the microscopic processes involved and the increased importance of carrier-carrier scattering, when combined with dynamic screening.

Based on our findings it appears that the effect of carrier-carrier scattering should be reassessed in other situations where highly non-equilibrium distribution functions are involved. Such conditions are found in state of the art high speed heterojunction devices. Our results further suggest that band renormalization may play a non-negligible role in many semiconductor devices, in which it has not yet been considered. Finally in our work on the simulation of femtosecond optical experiments we have developed the formulations and numerical methods, which are required to quantitatively investigate optoelectronic devices, such as light detectors and sources, on the first principles level. This is the direction we have proposed to take in the continuation task within the new JSEP program.

DEGREES AWARDED

None

REFERENCES

[1] C. J. Stanton, D. W. Bailey, and K. Hess, "Femtosecond-Pump, Continuum-Probe Nonlinear Absorption in GaAs," *Physical Review Letters*, 65, (2), 231 (1990).

[2] R. P. Josh, R. O. Grenadine, and D. K. Ferry, "Monte Carlo Simulation of Electron-hole Thermalization in Photo-exited Bulk Semiconductors," *Physical Review B*, 42, (9), 5685 (1990)

[3] J. A. Kash, "Carrier-carrier Scattering in GaAs: Quantitative Measurements from Hot (EPA) Luminescence," *Physical Review B*, 40, (5), 3455 (1989).

[4] Thomas Elsaesser, Jagdeep Shah, Lucio Rota, and Paolo Lugli, "Initial Thermalization of Photo-Excited Carriers in GaAs Studied by Femtosecond Luminescence Spectroscopy," *Physical Review Letters*, 66, (13), 1757 (1991).

[5] Jeff F. Young, Norm L. Henry, and Paul J. Kelly, "Full Dynamic Screening Calculatic.' of Hot Electron Scattering Rates in Multicomponent Semic..ductor Plasmas," *Solid State Electronics*, 32, 12 (1989).

[6] R. Binder, D. Scott, A. E. Paul, M. Lindberg, K. Hennebergerger, and S. W. Koch, "Carrier-carrier Scattering and Optical Dephasing in Highly Excited Semiconductors," *Physical Review B*, 45, (3), 1107 (1992).

[7] M. J. Kann, A. M. Kriman, and D. K. Ferry, "Role of Electron-electron Scattering on Ultrafast Probe Phenomena of Photoexcited Carriers in GaAs," Ultrafast Laser Probe Phenomena in Bulk and Microstructure Semiconductors III, Robert R. Alfano, *Editor, Proc. SPIE*, 1282, 98 (1990).

[8] Tilmann Kuhn, and Fausto Rossi, "Monte Carlo Simulation of Ultra-fast Processes in Photoexcited Semiconductors: Coherent and Incoherent Dynamics," *Phys. Rev.* B 46, 7496 (1992).

[9] J. H. Collet, J. Klebnicxki, J. Kuhl and K. Ploog, "Subpicosecond Kinetics of Band-edge Absorbtion in $Al_{0.25}Ga_{0.75}As$," J. Nunnenkamp, *Physical Review B*, 43, 14047 (1991).

[10] T. Gong, P.M. Fauchet, Jeff. F. Young, and P.J. Kelley, "Femtosecond Gain Dynamics Due to Initial Thermalization of Hot Carriers Injected at 2 eV in GaAs," *Phys. Rev.* B 44, 6542 (1991).

[11] T. Gong, P. Mertz, W.L. Nigham Jr., and P.M. Fauchet, "Femtosecond Refractive Index Spectral Hole Burning in Intrinsic and Doped GaAs," *Appl. Phys. Lett.* 59, 721 (1991).

[12] H. Roskos, B. Reik, A. Seilmeier, W. Kaiser, and G.G. Baumann, "Cooling of Photoexcited Carriers in Undoped and n-doped

$Ga_{0.47}In_{0.53}As$ Studied Within the First Few Picoseconds," *Phys. Rev.* B 40, 1396 (1989).

[13] H. Haug and S. Schmitt-Rink, "Electron Theory of the Optical Properties of Laser-excited Semiconductors," *Progress in Quantum Electronics*, 9, 3 (1984).

[14] L. Banyai, and S. W. Koch, Z. "A Simple Theory for the effects of plasma screening on the optical spectra of highly excited semiconductors," *Phys. B*, 63, 283 (1986).

JSEP PUBLICATIONS

1. S.R. Weinzierl and J.P. Krusius, "Heterojunction Vertical FET's Revisited: Potential for 225 GHz Large Current Operation, *IEEE Transactions on Electron Devices* 39, (5), 1050-1055 (1992).

2. S.R. Weinzierl and J.P. Krusius, "Space Charge Effects on Ballistic Injection Across Heterojunctions," *IEEE Transactions on Electron Devices*, 39, (7), 1780-1782 (1992).

3. J. E. Bair and J. P. Krusius, "Investigation Of The Role Of Free Carrier Screening During The Relaxation Of Carriers Excited By Femtosecond Optical Pulses," Ultrafast Laser Probe Phenomena in Semiconductors and Superconductors, R.R. Alfano (Editor), *Proceedings of the SPIE*, Vol. 1677, 157 (1992).

4. J. E. Bair, D. Cohen, J. P. Krusius, and C. R. Pollock, "Band Renormalization and Dynamic Screening in Near Bad Gap Femtosecond Optical Probing of InGaAs," manuscript submitted for publication to *Physical Review Letters*, April, 1992.

5. J. E. Bair and J. P. Krusius, "Monte Carlo Simulation of Near Band Gap Femtosecond Optical Experiments in Compound Semiconductors," manuscript in preparation, to be submitted for publication in July 1993.

6. S.R. Weinzierl and J.P. Krusius, "Design of the Modified Planar Doped Barrier for Maximizing the Effects of Ballistic Acceleration," manuscript in progress, to be submitted for publication in Summer 1993.

7. S.R. Weinzierl and J.P. Krusius, "Ballistic Injection Cathodes in High Speed Three Terminal Devices: Do They Really Make a Difference," manuscript in progress, for publication in Summer 1993.

PARALLEL STRUCTURES FOR REAL-TIME ADAPTIVE SIGNAL PROCESSING

Substitute Task #5

Task Principal Investigator: **Adam W. Bojanczyk**
(607) 255-4296

OBJECTIVE

Adaptive computational schemes are typical in real-time signal processing where the data collected from the environment rapidly changes in time. A basic step in adaptive processing is to discard a portion of the "old" data which no longer represents the environment, add new data, and then "adapt" the current knowledge about the environment according to the change in the data. Such processing arises for example in sensor array processing [1]. Sensor array processing is characterized by the need to implement multiple algorithms (i.e., weight vector computation, and application, eigenanalysis, residual evaluation, etc.). Serial implementation of such multiple tasks most often can be readily done via standard numerical schemes: the challenge in multiple tasking lies in devising efficient parallel implementations which are needed in real-time analysis of data derived from an array of sensors. Our three major objectives are: (i) development of strategies for adding and deleting information from the covariance matrix in multi-direction beamforming (least squares), (ii) development of strategies for tracking the eigenstructure of the array data after addition and deletion of data (covariance differencing), (iii) evaluation of procedures in (i) and (ii) on emerging parallel processor architectures.

DISCUSSION OF STATE-OF-THE-ART

Least squares problems are ubiquitous in engineering, science, operations research, etc. The linear least squares problem can be posed as follows:

Given an $n \times p$ matrix A, $p < n$, and a vector **b**, find the vector **x** which solves

$$\min_{\{\mathbf{x}\}} (A\mathbf{x} - \mathbf{b})^{\dagger} (A\mathbf{x} - \mathbf{b}) \tag{1}$$

Here † denotes conjugate transposition. As long as the data matrix A has full rank the solution is given by the solution to the corresponding normal equations,

$$\mathbf{x} = (A^{\dagger}A)^{-1} A^{T}\mathbf{b} \tag{2}$$

The quantity $K = A^{\dagger}A$ is referred to in the signal processing literature as the *covariance* matrix, and $\mathbf{x}$ is referred to as the *weight* vector.

In applications various constraints are imposed on the weight vector $\mathbf{x}$. Typical constraints are linear equality constraints, linear inequality constraints, or quadratic constraints.

Because K has twice the dynamic range as the data matrix A, it requires about twice the wordlength to work with on a computer, for a given problem. Therefore it is often advisable to *not* form K explicitly. The method of choice for solving full rank least squares equations is to proceed by a unitary transformation Q that "compresses" the data matrix A to the "information equivalent" triangular matrix U. This triangular matrix is known as a Cholesky factor of K. The desired least squares solution is next determined by solving the corresponding triangular system of linear equations.

In recursive least squares equation (1) needs to be solved repeatedly after some rows of A are removed and additional rows are added. The addition and the deletion are known as updating and downdating the Cholesky factor, respectively, or simply as a modification of the Cholesky factor. The reasons for updating and downdating vary. Perhaps the data to be deleted is unrepresentative of the data at large and so its effects on the weight vector (or parameter estimate) $\mathbf{x}$ must be excised (robust statistics). Or perhaps the data is changing with time and old data must be deleted (adaptive beamforming).

The combined process of updating and downdating the Cholesky factor is called a *sliding rectangular window* process, and is the focus of this work. An alternative to a sliding window is an exponential weighting, where all past data is employed in the least squares estimate, albeit with an increasingly smaller weighting, controllable by the decay rate of the window [2]. Single vector updating for exponential weighting, and an efficient parallel implementation of the process is described in [3]. In this work we are concerned with the (slowly varying) nonstationary case which necessitates multiple vector addition/deletion from the data matrix.

Processing of recursive least squares problems on a sequential machines is now well understood. It is known that while the updating process is numerically sound, the downdating can be very sensitive to rounding errors [4]. The most numerically accurate downdating methods requires formation of Q and downdating Q itself ([5]). This however results in quite high computational cost, and additional memory requirements for storing Q. One scheme that avoids computing Q is a single vector downdating scheme based on plane rotations [4]. Another way of downdating the Cholesky factor is the so-called *Stabilized Hyperbolic Householder* scheme [6] (see also [7]) which is half as expensive (for multiple vector updating/downdating problems), in terms of number of operations, as the method described in [4], and hence the preferable method for sequential rectangular sliding window process.

In parallel computing the additional cost of interprocessor communication has to be taken into account in assessing the cost of executing algorithms. Most discussions surrounding multiprocessor computers for signal processing have centered on planar (triangular) arrays [8], [9], [3], [10], [11], [12]. Perhaps the sole exception has been the important contribution by Rader in [13]. Both triangular and linear arrays considered in [10] or [13] are designed to implement efficiently the exponential weighting method. The exponential weighting method is very attractive for parallel implementation as it can be realized by a single updating process. On the other hand, the sliding window process is a composite tasks in the sense that each recursive step involves updating and downdating of the triangular factor followed by solving the resulting triangular systems of linear equations. None of the architectures proposed in [3] or [13] can efficiently deal with the sliding window process described above.

This work is interested in updating and downdating methods that can add or remove an arbitrary number of rows at a time. The advantage of such methods over repetitive applications of single row updating and downdating is that they make use of matrix-vector and matrix-matrix operations, as opposed to mainly vector-vector operations for single row methods. This may make such block methods faster on processors with caches and parallel computers.

PROGRESS

We have produced algorithms that can update and downdate an arbitrary number of rows from the triangular factor corresponding to the least squares problem. Below we describe the method of removing an arbitrary number of rows from the data (downdating). Updating (adding rows to the data) can be derived in a similar way.

Assume we wish to remove the first k rows, $(Z\ d)$, of the data $(A\ b)$. This would be a *rank-k downdating* of the linear least squares problem. To accomplish this, we must first examine the QR factorization of an augmented problem,

$$\begin{pmatrix} (Z\ \ d) & I_k \\ (\check{A}\ \ \check{b}) & 0 \end{pmatrix} = \bar{Q}\,\bar{R} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} R & Q_{11}^T \\ 0 & Q_{12}^T \end{pmatrix}, \tag{3}$$

where R is the upper triangular factor of the data matrix $(A\ b)$. Note that from this equation

$$Q_{11}Q_{11}^T + Q_{12}Q_{12}^T = I_k \tag{4}$$

implying that the first k *rows* of the orthogonal columns are orthogonal as well.

Now let H be an orthogonal matrix such that

$$\begin{pmatrix} (Z \;\; d) & I_k \\ (\check{A} \;\; \check{b}) & 0 \end{pmatrix} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} H H^T \begin{pmatrix} R & Q_{11}^T \\ 0 & Q_{12}^T \end{pmatrix} = \begin{pmatrix} 0 & P \\ \check{Q} & 0 \end{pmatrix} \begin{pmatrix} \check{U} & \check{u} & 0 \\ 0 & \check{\rho} & 0 \\ \check{Z} & \check{d} & P^T \end{pmatrix} \tag{5}$$

where $\check{U}$ is upper triangular. $\check{U}$ is the triangular factor corresponding to the data $(\check{A} \;\; \check{b})$. Then we solve $\check{U}\,\check{x} = \check{u}$, and we are done. The two-norm of the residual corresponding to the downdated problem is $\check{\rho}$.

The relation (3) is fundamental to all rank-k downdating methods we have developed. The major quantity of interest is $(Q_{12}^T Q_{22}^T)^T$ and the downdating methods will differ in how $(Q_{12}^T Q_{22}^T)^T$ is computed.

Methods that maintain Q and R or just R throughout a recursive least squares problem are described below:

Classical Gram-Schmidt on an augmented problem. The first method is to use classical Gram-Schmidt with reorthogonalization (CGS) to build on the orthonormal columns already given. Equation (3) represents a QR factorization, which could have been accomplished by classical Gram-Schmidt. Since we have Q_{11}, Q_{21}, and R already, we have completed n iterations (Note that the ith iteration of classical Gram-Schmidt produces the ith column of Q and the ith column of R). We may then proceed with the remaining k iterations of the orthogonalization process to get the new orthogonal columns.

Modified Gram-Schmidt on an augmented problem. Modified Gram-Schmidt (MGS) can also be used to get the new orthonormal columns. The ith iteration of MGS produces the ith column of $\bar{Q}$ and the ith row of $\bar{R}$ in (3) and updates the columns of $\bar{Q}$ to be formed in later iterations. Again we use the fact that (3) represents a QR factorization that is partially completed. After $n + 1$ iterations of MGS, we have the following:

$$\begin{pmatrix} (Z \;\; d) & I_k \\ (\check{A} \;\; \check{b}) & 0 \end{pmatrix} = \begin{pmatrix} Q_{11} & T_1 \\ Q_{21} & T_2 \end{pmatrix} \begin{pmatrix} R & Q_{11}^T \\ 0 & I_k \end{pmatrix} \tag{6}$$

The only term in the above that we do not have is $T^T = (T_1^T \; T_2^T)$. We can find T by performing only the update portion of MGS (i.e., $T = T - q_i q_i^T T$) for each of the $n + 1$ orthogonal columns from $(Q_{11}^T \; Q_{21}^T)^T$ one at a time in ascending order on the new k columns of T.

Using Corrected Semi-Normal Equations. This method is one of the numerically best that only maintains R and not Q (but uses the original data (A b)). The basis of this method is in corrected semi-normal equations [14].

The method computes T in (6) by first solving the normal equations corresponding to the linear least squares problem. The accuracy of T is next improved by the technique of iterative refinement. Finally, the QR decomposition of T produces the desired submatrix Q_{12}.

Numerical Results

The following two figures display the relative error of solutions for each of the three methods described. Both tests represent a window of size 8 sliding down a 50 x 5 matrix three rows at a time (i.e., three rows are updated and three rows are downdated each time the window "slides" down). The first matrix tested in Figure 1 is well-conditioned except for a large element at entry (18,3). The second matrix tested in Figure 2 is an overall ill-conditioned matrix, especially in the central rows. Each method is compared against a full QR decomposition method, and a method extended from LINPACK. The full QR method should result in the best possible accuracy since it has complete information, but has the resulting disadvantage of being slow from maintaining all of its information. The LINPACK method is an example of a "maintain R only" method that is well known and commonly used.

Figures 1 and 2 illustrate that the methods developed in this work achieve the numerical accuracy of a full QR decomposition for numerically "bad" sliding window scenarios that commonly occur in practice. However, the methods presented involve less work than a full QR decomposition method. Note also that our newly developed methods perform significantly better than the established method of LINPACK.

Parallel results

Distributed memory parallel architectures are well suited for matrix processing as application and machine topology are often well matched. The critical aspect in the performance of matrix computations on such architectures is the distribution of the matrix data among the processors. Standard matrix data distributions are by rows, columns, or subblocks, either block- or wrap-distributed to the processors. The choice of data distribution affects communication requirements, and the load balance among the processors.

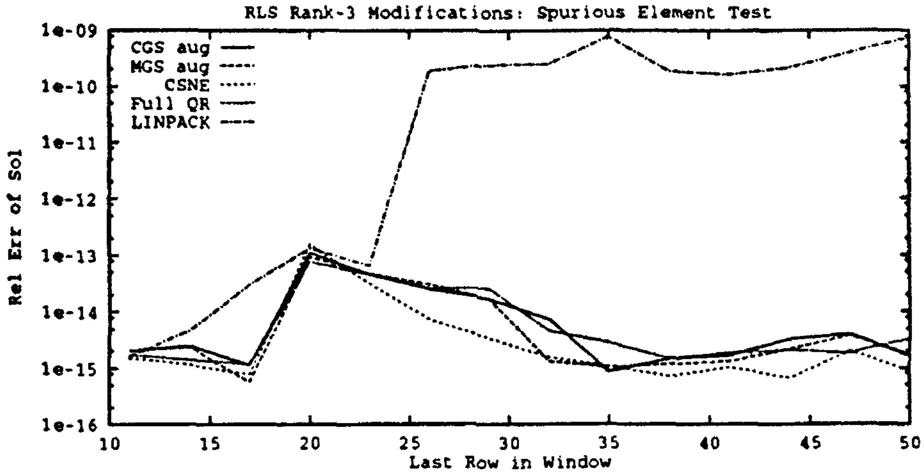


Figure 1. Sliding Window Test - Spurious Element.

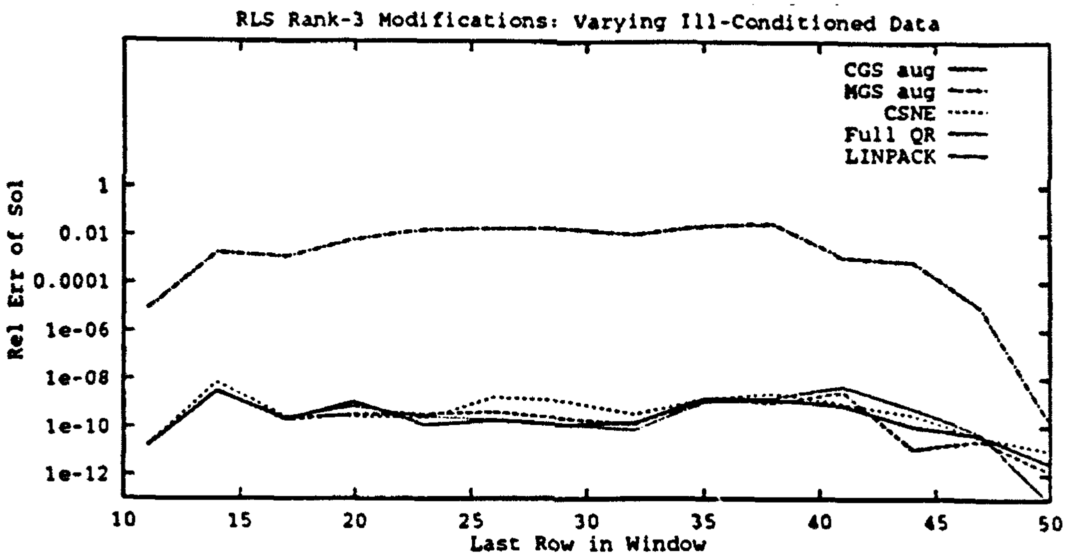


Figure 2. Sliding Window Test - Varying Ill-Conditioned Data.

In light of this, we have examined experimentally the parallel properties of two of the fundamental algorithms used in recursive least squares: back substitution and Givens rotation sequences.

Figure 3 shows the back substitution algorithm and Figure 4 shows a typical Givens rotation sequence needed in recursive least squares (for forming H in (5)) running on an Intel iPSC/860 Hypercube for various matrix data distributions. Notice how the distribution choice greatly affects execution time. Also note that our newly developed block distribution implementations perform better than published wrap implementations [15] (denoted by "LC89"), experimentally supporting the theory that block-oriented algorithms are preferable in parallel environments.

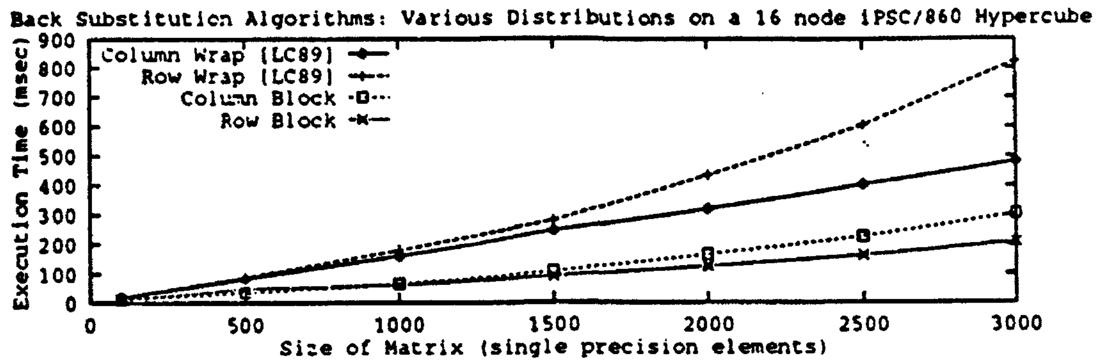


Figure 3. Comparison of Back Substitution Parallel Implementations.

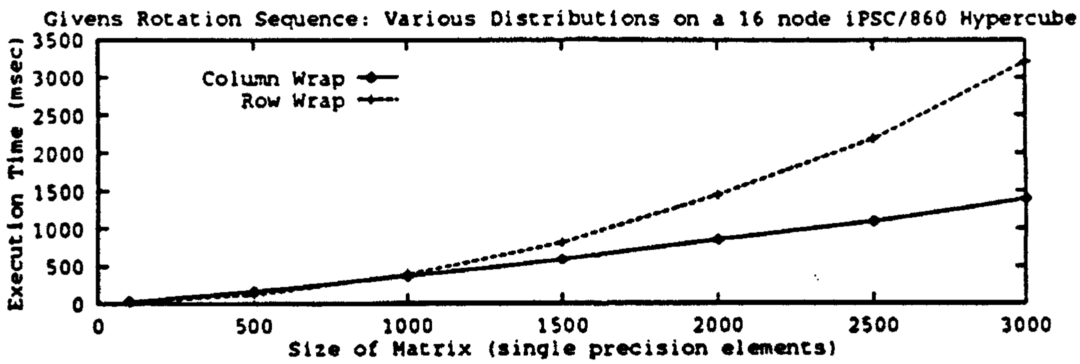


Figure 4. Comparison of Givens rotation sequence parallel implementations.

SCIENTIFIC IMPACT OF RESEARCH

The research concluded so far concentrated mainly on recursive windowed linear least squares problems arriving in real-time DSP systems derived from the covariance differencing principle. The contribution of this research is twofold. First, new highly concurrent algorithms amenable to efficient parallel implementation were developed and their numerical properties analyzed. Second, tests have been conducted indicating the feasibility of a block approach to recursive linear least squares problems. What remains to be done is the formulation of parametric models that can predict the best numerical and parallel implementation for recursive least squares problems based on the specified requirements of the user.

DEGREES AWARDED

None

REFERENCES

[1] R. Monzingo and T.W. Miller, Introduction to Adaptive Arrays, Wiley (1980).

[2] S. Haykin, Adaptive Filter Theory, Prentice-Hall (1986).

[3] J.V. McCanny and J.G. McWhirter, "Some Systolic Array Developments in the United Kingdom," *IEEE Computer*, 20, 51-63 (July 1987).

[4] A. Bojanczyk, R. Brent, P. Van Dooren, and F. De Hoog, "Downdating the Cholesky Factorization," *Algorithms and Appplications on Vector and Parallel Computers*, H.J.J. de Riele, Th.J. Dekker and H.A. van der Vorst, eds., Elsevier Science Publishers (1987).

[5] G. Golub and C. Van Loan, Matrix Computations, John Hopkins Press, 2nd edition (1989).

[6] A.W. Bojanczyk and A.O. Steinhardt, "Stabilized Hyperbolic Householder Transformations," *IEEE ASSP*, 37, (8), 1286-8 (1989).

[7] C. Rader and A. Steinhardt, "Hyperbolic Householder Transformations," *IEEE ASSP*, 1589-1602 (Dec. 1986).

[8] A.W. Bojanczyk and F.T. Luk, "A Unified Systolic Array for Adaptive Beamforming," *J. Parallel and Distributed Computing*, 8, (4), 388-392 (April 1990).

[9] W.M. Gentleman and H.T. Kung, "Matrix Triangularization by Systolic Arrays," *SPIE Proceedings*, 298, 12-26 (1982).

[10] J.G. McWhirter, "Algorithmic Engineering - an Emerging Discipline," *Proceedings of the SPIE*, 1152, 2-15 (1989).

[11] R. Schreiber and W.P. Tang, "On Systolic Arrays for Updating Cholesky Factorization," *BIT*, 26, 451-66 (1986).

[12] R. Schreiber, "Implementation of Adaptive Array Algorithms," *IEEE ASSP*, 34, (5) (October 1986).

[13] C.M. Rader, "Wafer-Scale Integration of a Large Systolic Array for Adaptive Nulling," *The Lincoln Laboratory Journal*, 4, (1), 3-29 (1991).

[14] Å . Björck, "Stability Analysis of the Method of Seminormal Equations for Linear Least Squares Problems," *Linear Algebra and Its Applications,* 88-89:31-48 (1987).

[15] G. Li and T. Coleman, "A New Method for Solving Triangular Systems on Distributed-Memory Message-Passing Parallel Processors," SIAM Journal of Scientific and Statistical Computing, 10 (2), 382-396 (March 1989).

JSEP PUBLICATIONS

1. A.W. Bojanczyk, J. Nagy and R. Plemmons, "Row Householder Transformations for rank-k Inverse Modifications," *Linear Algebra and its Applications,* to appear.

2. A.W. Bojanczyk, J. Lebak and S. Olszanskyj, "Rank-*k* Modification Methods for Recursive Least Squares Problems," Technical Report, School of Electrical Engineering, Cornell University (January 1993).

3. A.W. Bojanczyk and P. Van Dooren, "On Propagating Orthogonal Transformations in a Product of 2×2 Triangular Matrices," *Proceedings of the Kent Conference on Scientific Computing,* Kent State University, Kent, Ohio (April 1992).

4. A.W. Bojanczyk and P. Van Dooren, "Reordering Diagonal Blocks in Real Schur Form," *Proceedings of the NATO ASI Workshop on Real-time and Large Scale Computing,* Leuven, Belgium (August 1-15, 1992).

5. A.W. Bojanczyk, G.H. Golub and P. Van Dooren, "The Periodic Schur Decomposition. Algorithms and Applications," *Proceedings of the SPIE's Annual Meeting,* San Diego (July 19-24, 1992).

FAULT TOLERANT BEAMFORMING ALGORITHMS

Task #6

Task Principal Investigator: **Franklin T. Luk**
(518) 276- 8326

OBJECTIVE

Multidimensional signal processing in the context of processing signals received by an array of sensors has many important applications. The type of filtering that can be conveniently applied to signals carried by propagating waves is beamforming, which seeks to isolate signal components that are propagating in a particular direction. Although computationally expensive, the beamforming procedure has been rapidly rising in popularity due to advances in both matrix algorithms and systolic arrays. Most systolic arrays will be deployed in harsh environments and are thus susceptible to frequent transient errors. The principal objective of this task is to develop systolic fault tolerant beamforming techniques. Special attention will be paid to computing complex matrix decompositions, avoiding numerical overflows, differentiating between errors arising from numerical roundoff buildups and those from hardware failures, and interrupting the operation of systolic arrays for error correction.

DISCUSSION OF STATE OF THE ART

The compatibility of systolic arrays and algorithms with both matrix computations and today's VLSI technology guarantees their future use as key components in any signal processing system. An especially important systolic algorithm is the orthogonal triangulation algorithm (QR decomposition) for least squares minimization, a crucial step in most adaptive antenna processing algorithms. The importance of these problems is evidenced by two major systolic array projects; one at MIT's Lincoln Laboratory [1] and the other at the United Kingdom's Royal Signals and Radar Establishment (RSRE) [2]. However, traditional fault tolerance techniques such as modular redundancy have been regarded as too costly and unwieldy to implement on these systolic arrays. In [3], a JSEP supported work, we presented a simple fault tolerance scheme for the QR decomposition and showed how it can be easily incorporated into the RSRE systolic arrays for recursive least squares minimization. Our work has already won recognition at the RSRE as a possible fault tolerance technique for their systolic arrays [4].

Data matrices that are ill-conditioned call for a more robust and more expensive numerical technique known as the singular value decomposition (SVD). An SVD systolic array designed by us has been adopted for hardware implementation at both the RSRE [5] and Computational Engineering, Inc. [6]. The implementation of the latter will be used for real time system control; its application in the wing flutter analysis of supersonic planes has been proven in a wind tunnel test at an Air Force Laboratory in Ohio. The problem of fault tolerant computation of the singular value decomposition awaits a nice solution. Schemes were reported in [7], but they are so complicated that triple modular redundancy may well be a better choice.

Existing fault tolerance schemes have often been ignored by systolic array designers because they are too costly and unwieldy to implement. An attractive new idea came in the form of algorithm-based fault tolerance. This approach employs three steps; encode the input data, execute the algorithm on the encoded input to produce encoded output, and decode the output to detect and perhaps correct errors. Both checksum and weighted checksum encoding schemes have been developed by Abraham et al. [8, 9], who showed that a variety of matrix operations preserves the checksum property.

In [9, 10] a linear algebraic interpretation of the weighted checksum scheme was proposed. Such a model allows parallels to be drawn between algorithm-based fault tolerance and coding theory, and makes it possible to examine in detail the difficulties in choosing weight vectors such that the correction vector can be explicitly resolved. The hard problem of how to determine the exact number of errors that have occurred has been solved in [11]. For error correction, prior to [11], it was known only how to correct a weighted checksum scheme for the cases of one error [9] and two errors [10]. In [11] a theoretical framework was given which would enable one to solve the correction problem for the general case.

The weighted checksum technique has been demonstrated to be effective in multiple error detection. It has been shown that, in order to guarantee error detection, the chosen weight vectors must satisfy some very specific properties about linear independence. Previously, appropriate sets of weight vectors have been proposed which are powers of integers [9, 12]; these suffer from the fact that the weights can become very large. In [13, 14] a new scheme was presented that generates weight vectors to meet the requirements about independence and to avoid the difficulties with overflow.

PROGRESS

The problem of linearly constrained least squares has many applications in signal processing. In a paper that will appear in the journal Integration, we present a perturbation analysis of a linearly constrained least squares algorithm for adaptive beamforming. The perturbation bounds for the solution as well as for the latest residual element are derived. We also propose an error estimation scheme for the residual element, which can be incorporated into a systolic array implementation of the algorithm.

A new approach to the parallel solution of the Singular Value Decomposition (SVD) has been devised by us. The recursive and hierarchical structure of the new ordering coincides with the hierarchical connection network, in which neighboring processors are more tightly coupled than processors far apart, minimizing the communication overhead. The fat tree was used as a first example of a hierarchical network. The resulting algorithm scales up nicely when the processors are configured as a two-dimensional fat-tree. The best approach to parallel SVD computation is one of the Jacobi type; such algorithms have been developed for hypercube, linear array and ring architectures. Motivated by the fat tree architecture, in this research we have developed an implementation of a Jacobi method on the massivley parallel CM-5. We propose new Jacobi orderings for efficient utilization of the communication network and analyze its behavior. Our paper has been presented at an Army Conference on Applied Mathematics and Computing, and will appear in its Conference Proceedings.

SCIENTIFIC IMPACT OF RESEARCH

Our work is making a significant impact in that it is getting lots of attention so that several researchers are attempting to improve on our work [15-17]. We are most proud of our result in discovering the relationship between the famous Berlekamp-Massey algorithm for decoding the Reed-Solomon code, and the well known Lanczos algorithm in numerical computing.

DEGREES AWARDED

None

REFERENCES

[1] C. M. Rader, "Wafer-Scale Systolic Array for Adaptive Antenna Processing," *Proceedings IEEE ICASSP*, New York, NY, 2069-2071 (1988).

[2] J. V. McCanny, and J. G. McWhirter, "Some Systolic Array Developments in the United Kingdom," *IEEE Computer*, 20, 51-63 (July 1987).

[3] C. J. Anfinson, F. T. Luk, and E. K. Torng, "A Novel Fault Tolerance Technique for Recursive Least Squares Minimization," *Proceedings of SPIE Vol. 975, Advanced Algorithms and Architectures for Signal Processing III*, 268-276 (1988).

[4] I. K. Proudler, Private communication at Conference on Mathematics in Signal Processing, Warwick, England, December 1988.

[5] G. de Villers, "Simulation of Luk's SVD Array Using a Transputer," Paper presented at Conference on Mathematics in Signal Processing, Warwick, England, December 1988.

[6] W. E. Larimore, J. Galdos, D. Martin, and F. T. Luk, "A Transputer Implementation of System Identification and Control Using SVD's," *Proceedings SPIE, Vol. 1069, Advanced Algorithms and Architectures for Signal Processing IV*, Paper 11 (August 1989).

[7] C. Y. Chen and J. A. Abraham, "Fault-Tolerant Systems for the Computation of Eigenvalues and Singular Values," *Proceedings SPIE, Vol. 696, Advanced Algorithms and Architectures for Signal Processing I*, 228-237 (1986).

[8] K. H. Huang and J. A. Abraham, "Algorithm-Based Fault Tolerance for Matrix Operations," *IEEE Transactions on Computers*, C-33, 518-528 (1984).

[9] J. Y. Jou and J. A. Abraham, "Fault-Tolerant Matrix Arithmetic and Signal Processing on Highly Concurrent Computing Structures," *Proceedings IEEE*, 74, 732-741 (1986).

[10] C. J. Anfinson and F. T. Luk, "A Linear Algebraic Model of Algorithm-Based Fault Tolerance," *IEEE Transactions on Computers*, C-37, 1599-1604 (1988).

[11] R. P. Brent, and F. T. Luk, "A Theoretical Foundation for Weighted Checksum Methods," C. J. Anfinson, *Proceedings of SPIE Vol. 975, Advanced Algorithms and Architectures for Signal Processing III*, 10-18 (1988).

[12] F. T. Luk, "Algorithm-Based Fault Tolerance for Parallel Matrix Equation Solvers," *Proceedings SPIE*, 564, Real Time Signal Processing VIII, 49-53 (1985).

[13] R. P. Brent, F. T. Luk, and C. J. Anfinson, "Choosing Small Weights for Multiple Error Detection," *Proceedings of SPIE*, 1058, High Speed Computing II, 130-136 (1989).

[14] R. P. Brent, F. T. Luk, and C. J. Anfinson, "Checksum Schemes for Fault Tolerant Systolic Computing," Paper presented at Conference on Mathematics in Signal Processing, Warwick, England (December 1988).

[15] J. Chun and F. T. Luk, "Algorithmic Sensor Failure Detection on Passive Antenna Arrays," *Proceedings of SPIE Vol. 1566, Advanced Signal Processing Algorithms, Architectures and Implementations II, SPIE -*

The International Society for Optical Engineering, Bellingham, Washington, 483-492 (1991).

[16] D. L. Boley, R. P. Brent, G. H. Golub and F. T. Luk, "Algorithmic fault tolerance using the Lanczos method," *SIAM Journal on Matrix Analysis and Applications,* 13, 312-332 (1992).

[17] D. L. Boley and F. T. Luk, "A Well Conditioned Checksum Scheme for Algorithmic Fault Tolerance," *Integration, the VLSI Journal,* 12, 21-32 (1991).

ISEP PUBLICATIONS

1. F. T. Luk, Editor, *Proceedings of the SPIE Vol. 1770 Advanced Signal Processing Algorithms, Architectures, and Implementations III* - The International Society for Optical Engineering, July 19-21, 1992, San Diego, CA, 468 pages.

2. G. E. Adams, A. W. Bojanczyk, and F. T. Luk, "Computing the PSVD of Two 2 x 2 Triangular Matrices," *Proceedings of the SPIE Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations III,* 1770, 43-59 (1992).

3. F. T. Luk and S. Qiao, "Analysis of a Linearly Constrained Least Squares Algorithm for Adaptive Beamforming," *Proceedings of the SPIE Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations III,* 1770, 210-218 (1992).

4. T. J. Lee, F. T. Luk, and D. L. Boley, "Computing the Singular Value Decomposition on a Fat-Tree Architecture," Department of Computer Science, Rensselaer Polytechnic Institute, Technical Report No. 92-33 (November 1992).

INTERRUPT AND BRANCH HANDLING FOR REAL-TIME SIGNAL PROCESSING SYSTEMS

Task #7

Task Principal Investigator: **H. C. Torng**
(607) 255-5191

OBJECTIVE

All the emerging processors, such as the Pentium from Intel- to be announced in the spring of 1993, and the IBM RS6000 "Power" architecture, employ multiple functional units, issue and execute multiple, and possibly out-of-order, instructions from an instruction stream. The objective of our task is to address two important issues that confront the development of these processors: interrupt handling and branch processing.

An indispensable requirement for modern processors is that they handle interrupt promptly and precisely. 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, a processor first stores its processor state; this has to be done precisely so that the interrupted process can be resumed at the very point of interruption.

In a modern processor with multiple functional units, which may include arithmetic and logic units (ALU's), multipliers, branch handlers, and memory access units, multiple instructions are at various stage of execution at a given time instant; it is extremely difficult to identify and store a precise interrupt point with an acceptable latency.

Realistic programs always contain conditional branch instructions. It takes time to resolve each and every one of these conditional branches; and there is an inevitable delay to access the target instruction once the branch uncertainty is resolved. These undesirable effects are even more prominent in computer systems which issue multiple and out-of-order instructions.

We seek solutions to these two problems, which are important for real-time signal processing systems as well as for general computer applications.

DISCUSSION OF STATE-OF-THE-ART

Interrupts and exceptions can be classified into three types: external interrupts, exception traps, and software traps. External interrupts are initiated by sensors and other outside elements; the computer system being interrupted is called upon to perform certain tasks of higher priority. When a processor encounters an abnormality in its processing, such as division by zero, overflow, or illegal operations, it generates exception traps. Software traps are installed by programmers in an instruction stream to initiate interrupt requests; with these pre-planned interrupts, programmers gain additional controls of the processing of their programs.

In response to an interrupt request, a processor saves its processor state, then load and execute an appropriate interrupt handler. Upon completion of the interrupt handling routine, the processor restores the saved processor state and then restarts the interrupted process.

A processor state should contain only enough information so that the interrupted process can be restarted at the **precise** point where it was interrupted. To be able to resume an interrupted process, the processor state should consist of the contents of the general purpose registers, the program counter, the condition registers, all index registers and the relevant portion of the main memory. A processor state consumes a large amount of precious memory.

For processors which issue at most one instruction per machine cycle, it is relatively easier to identify precisely the point where a process is interrupted. Among other vital items, the address of a specific instruction, say instruction a, is saved. All instructions that precede instruction a have been executed. And instruction a and those that follow it have not. Instruction a thus provides a **precise** interrupt point.

Unfortunately, modern processors, such as the Intel Pentium and IBM RS 6000, may issue more than one instruction per machine cycle and some of these instructions are issued out-of-order. The identification of a precise interrupt point becomes very difficult. We have addressed this task and have found a promising solution. A brief summary of the state of the art is presented below:

The machines designed by Cray [1, 2] have multiple functional units and do allow instructions executed out-of-order. In response to interrupt requests, these machines allow instructions under execution to complete before the processor state is stored; this means that they do not respond to interrupt requests promptly. For machines designed for off-line scientific computing, this is acceptable. It is however not desirable or even permissible for real-time applications. In the IBM 360/91 [3], a similar "long-latency" approach was adopted by allowing all issued instructions to complete their execution. The IBM 360/91 does allow an "imprecise" interrupt which ignores the interrupt point; in other words, the interrupted process can not be restarted later.

Checkpointing has been proposed to facilitate precise interrupt handling [4]. In such a scheme, a minimum of two checkpoints and hence two additional processor states have to be maintained. Clearly, the checkpointing approach proposed degrades system performance, both in the time required to process a given task, and in the time required to restore to a consistent processor state upon receiving an interrupt request. The processing time for a task is increased by the the need to update state information as the states change, and by the execution of additional read instructions which must precede all instructions which alter the memory. It should be pointed out also that each checkpoint corresponds to a processor state; the memory and other hardware required to implement each and every one of these checkpoints are extremely costly.

Smith and Pleszkun [5] presented several schemes to realize the classical precise interrupts. By introducing "re-order registers", "history buffers" and "future files", they make sure that the instructions are completed "in-order", even though they may be executed out-of-order. In so doing, a precise interrupt point can be identified and saved. This is done with unavoidable performance degradations. We should and can do better.

The effectiveness of processors with multiple functional units in realizing performance gains is hindered to a great extent by the presence of branches, especially conditional branches, in system and application codes.

One is tempted to learn to live with a conditional branch by waiting patiently for its resolution. However, the number of instructions that can be examined at run time by the processor for possible concurrent issuance will be limited by the size of a basic block, which is usually not that large. We have to deal with conditional branches more aggressively.

Schemes have been proposed and implemented to predict the outcome of a conditional branch [6, 7], and to fetch and execute instructions on the predicted path. Of course, it should be expected that we may encounter more conditional branches as we process along the selected path. What has been gained is that with branch prediction we have essentially enlarged a basic block and made many more instructions available for concurrent processing. There is one caveat: some of the branch predictions are going to be wrong. When that happens, the processor has to be able to restore the processor state to the point where the correct path is to be taken. Not only do we lose the work that has been done on the wrongly predicted path, but also we have to take time to restore the processor state.

In order to make a branch prediction scheme worthwhile, we have to have the following:

1. a good branch prediction mechanism, which makes most of the predictions correctly; and

2. an efficient recovery mechanism, which does not degrade the processor performance in normal operation and it restores the processor to an appropriate state when a prediction made turned out to be incorrect.

PROGRESS

We have made considerable progress in the following investigations:

Interrupt Handling [8, 9, 10, 11]

We have identified factors that must be considered in evaluating the effectiveness of interrupt and exception handling schemes: latency, cost, and performance degradation.

We have formulated and evaluated an efficient hardware mechanism, the Instruction Window (IW), and a new approach, which allows for precise, responsive and flexible interrupt and exception handling.

We have examined in detail the implementation of the IW. The design of an 8-cell IW has been carried out. We have established that such an instruction window can work with a very short machine cycle time; this is important because we do not want the introduction of the IW to reduce the processor clock rate.

We have undertaken a comparison of all interrupt and exception handling schemes for out-of-order execution processors.

A paper reporting this topic appears in the January 1993 issue of the *IEEE Trans. on Computers*. And the patent application pending since January 1990 should be resolved soon.

We have completed the study of the Fast Dispatch Stack (FDS) system, which provides another approach to fast, precise interrupt handling. In addition to instruction issuance, the FDS serves as a "re-order" buffer for the processor; this extends considerably the state of the art in this important area.

Branch Processing [9, 10, 11]

We have established that the Fast Dispatch Stack (FDS) can be used to reduce the penalty when a branch is predicted incorrectly. Conditional branch instructions along with those instructions on the predicted branch path are brought into the FDS to be executed speculatively. The execution results are written into an additional set of registers, called the "working registers" for temporary storage and accesses by subsequent instructions. These results are copied into the "architected registers" once the instructions are retired from the FDS.

We now require that a new instruction removal mode be instituted: only those instructions at the top of the FDS may be retired; multiple instructions are retired at once if they form a contiguous sequence of instructions at the top.

When the prediction made for a conditional branch is found to be correct, the branch instruction can be removed from the FDS when it is a member of a contiguous segment of completed instructions, including the top one. Again, please note that when an instruction is retired from the FDS, its result is made permanent by being copied into an architected register.

When the prediction made for a conditional branch is found to be incorrect, the branch instruction will not be retired from the FDS. All instructions which follow the branch are removed immediately. When the branch instruction reaches to the top of the FDS, the contents of the architected registers are copied into the working registers. The FDS is then filled with instructions from the correct path. In other words, the correction required can be accomplished very quickly, often in one machine cycle; this reduces the penalty for wrong guesses and thus facilitates the use of branch prediction as a means for performance enhancement.

New Implementation Schemes [10, 11, 12]

We have addressed the concerns that the Dispatch Stack that we have developed may have adverse impact on processor cycle time, and developed several new schemes: the use of bit vectors; the use of pointers; and finally a block based window. Details can be found in the cited publications.

SCIENTIFIC IMPACT OF RESEARCH

We have developed solutions to several outstanding problems in the development of modern processors, which can make the best use of the advancing device technologies. We have made considerable progress in interrupt handling, branch processing and dispatch stack implementation. These advances enhance significantly the performances of superscalars processors without raising the clock rate, and provides timely and much needed solutions to outstanding problems. We continue to have ongoing discussions with IBM and Intel.

DEGREES AWARDED

1. Deborah T. Marr
   "A Block-Based Dispatch Window for Multiple Out-of-Order Instruction Issue," M.S., Electrical Engineering, June 1992.

REFERENCES

[1] R. M. Russell, "The CRAY-1 Computer System," *ACM Communications*, 21, (1), 63-72 (January 1978).

[2] J. E. Thornton, DESIGN OF A COMPUTER: The Control Data 6600, Scott Foresman (1970).

[3] R. M. Tomasulo, "An Efficient Algorithms for Exploiting Multiple Arithmetic Units," *IBM Journal*, 11, 25-33 (January 1967), and related articles in the same issue.

[4] Hwu, Y. N. Patt, "Checkpoint Repair for High Performance Out-of-Order Execution Machines," *IEEE Trans. Computer.*, C-36, (12), 1496-1514 (Dec. 1987).

[5] J. E. Smith and A. R. Pleszkun, "Implementing Precise Interrupts in Pipelined Processors," *IEEE Trans. Computer.*, C-37, (5), 562-573 (May 1988).

[6] J. K. F. Lee and A. J. Smith, "Branch Prediction Strategies and Branch Target Buffer Design," *IEEE Computer*, 17, (1), 6-22 (January 1984).

[7] S. McFarling and J. Hennessy, "Reducing the Cost of Branches," *Proc. 13th Sym. Computer Architecture*, 396-404 (1986).

[8] H. C. Torng and Martin Day, "Interrupt Handling for Out-of-Order Execution Processors," *IEEE Trans. on Computers*, 42 (January 1, 1993).

[9] H. Dwyer and H. C. Torng, "An Out-of-Order Superscalar Processor with Speculative Execution and Fast, Precise Interrupts," *Proceedings of 25th Symposium on Microarchitectures*, 272- 281 (December 1992).

[10] H. C. Torng, H. Dwyer, and D. Marr, "On Instruction Windowing for Fine Grain Parallelism in High-Performance Processors," invited paper and in *Proceedings of 1993 International Phoenix Conf. on Comp. and Communications* (March 23-26, 1993).

[11] H. Dwyer, "A multiple, Out-of-Order, Instruction Issuing System for Superscalar Processors," Ph.D. Thesis, School of Electrical Engineering, Cornell University (1991).

[12] D. Marr, "A Block-Based Dispatch Window for Multiple Out-of-Order Instruction Issue," M. S. Thesis, School of Electrical Engineering, Cornell University (1992).

ISEP PUBLICATIONS

1. H. C. Torng and Martin Day, "Interrupt Handling for Out-of-Order Execution Processors," *IEEE Trans. on Computers,* 42 (January 1, 1993)

2. H. Dwyer and H. C. Torng, "An Out-of-Order Superscalar Processor with Speculative Execution and Fast, Precise Interrupts," *Proceedings of 25th Symposium on Microarchitectures,* 272- 281 (December 1992).

3. H. C. Torng, H. Dwyer, and D. Marr, "On Instruction Windowing for Fine Grain Parallelism in High-Performance Processors," invited paper and in *Proceedings of 1993 International Phoenix Conf. on Comp. and Communications* (March 23-26, 1993).