

THIS FILE COPY

1

AD-A211 394

ANNUAL REPORT

ULTRA HIGH SPEED COMPOUND SEMICONDUCTORS
AND
REAL TIME SIGNAL PROCESSING

MAY 1, 1988 - APRIL 30, 1989

CONTRACT # F49620-87-C-0044

DTIC
ELECTE
AUG 15 1989
S D D

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

89 8 14 124

REPORT DOCUMENTATION PAGE

For: Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT Approval for public release. Distribution unlimited		
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Cornell University		6b. OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research		
6c. ADDRESS (City, State, and ZIP Code) 119 Phillips Hall Ithaca, NY 14853-5401			7b. ADDRESS (City, State, and ZIP Code) Building 410, Bolling Air Force Base Washington, D.C. 20332-6448		
8a. NAME OF FUNDING / SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (if applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-87-C-0044		
8c. ADDRESS (City, State, and ZIP Code)			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.
					WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) Ultra High Speed Compound Semiconductors and Real Time Signal Processing					
12. PERSONAL AUTHOR(S) J. Peter Krusius					
13a. TYPE OF REPORT Annual Report		13b. TIME COVERED FROM 5/1/88 TO 4/30/89		14. DATE OF REPORT (Year, Month, Day) 5/1/89	15. PAGE COUNT
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Compound semiconductor, organometallic vapor phase epitaxy, femtosecond laser probing, Monte Carlo simulation, VLSI, fault tolerance, computer architecture		
FIELD	GROUP	SUB-GROUP			
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This report is the annual report on research conducted under the auspices of the Joint Services Electronics Program at Cornell University. The research is grouped under two themes: (a) ultra high speed compound semiconductors, and (b) real time signal processing. Results on OMVPE materials growth, femtosecond laser probing of hot carriers, and ensemble Monte Carlo simulations are reported on under the first theme. Accomplishments on VLSI algorithms, fault tolerant architectures, and architectures with multiple functional units for signal processing are given under the second theme.					
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> 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

CONTENTS

	Page
A. Director's overview	1
B. Description of Special Accomplishments and Technology Transition	2
B.1 Femtosecond Carrier Processes in Compound Semiconductors	2
B.2 Real Time Signal Processing	3
C. Description of Individual Work Units	4
C.1 Femtosecond Carrier Processes in Compound Semiconductors	
Task #1 OMVPE Growth of III-V Alloys and Structures for New High Speed Electron Devices	5
Task #2 Femtosecond Laser Studies of Ultrafast Processes in Compound Semiconductors	10
Task #3 Ultrafast Studies of Device Structures and of Hot Carriers in Narrow Bandgap Semiconductors	15
Task #4 Physics of High Speed Meso Scale Compound Semiconductor Devices	20
C.2 Real Time Signal Processing	
Task #5 Novel VLSI Algorithms and Architectures for High Data Rate Digital Filtering	27
Task #6 Fault Tolerant Signal Processing Architectures	31
Task #7 Real-Time Signal Processing Systems with Multiple Functional Units	36

A. DIRECTOR'S OVERVIEW

This document is an annual report of the current two year Cornell Joint Services Electronics Program for the period from May 1, 1988 to April 30, 1989. This is the first year after the program was broadened from an exclusive focus on compound semiconductor materials and devices into a two theme approach. One of these themes continued the compound semiconductor research concentrating on more fundamental phenomena, femtosecond transport and optical phenomena in heterostructures, while the other theme added research in a new area, real time digital signal processing, to the program. The new objectives brought four new faculty (C. Pollock, G. Bilardi, F. Luk and H. Torng) to a total of seven principal investigators now participating in the program.

As a consequence of these major changes in the program new research objectives have been pursued, common ground between the new set of investigators has been charted, and collaborative research initiated. As a means to define the identity of the new Cornell JSEP program the director initiated the publication of an entire issue of the Engineering Quarterly, a periodic journal published by the College of Engineering, on JSEP research [A New Thrust in Electronics Research, Engineering: Cornell Quarterly, Vol. 23, No. 1, Autumn 1988]. This Quarterly issue was distributed to the full JSEP distribution in addition to the wide Cornell distribution to industries, businesses, firms, government organizations, foundations, alumni, faculty, and students.

Efforts to establish a new compound semiconductor growth facility at Cornell have been the main collaborative issue for the JSEP faculty during the current program period. The new facility is expected to require continued attention well into future years of the program. The old organometallic vapor phase epitaxy (OMVPE) operation on the fourth floor of Phillips Hall had become substandard because of ever tightening hazard gas safety regulations. The new growth laboratory, a shared facility operated under the technical direction of Prof. R. Shealy and an oversight committee consisting of users and administration, is being established in an existing building off-campus. Plans for the new facility call for the installation of three independent OMVPE systems to be used for specialized growth tasks. One of these reactors will be the rebuilt reactor moved from Phillips Hall, the second a fully operational reactor donated by General Electric Company, and the third a reactor currently under construction. Much of the effort in R. Shealy's task has been devoted to the establishment of this new OMPVPE facility, which will provide truly unique capabilities for compound semiconductor heterostructure growth. Compound semiconductor JSEP faculty in addition to R. Shealy have contributed to the planning, fund raising, and extensive discussions with university administration because of the importance of this facility to JSEP and related research.

In addition to joint efforts centered around the new OMPVPE facility, research interactions within the compound semiconductor theme have intensified. As an example, the ongoing tunable femtosecond optical relaxation experiments have been designed jointly by experimental and theoretical investigators (C. Pollock and P. Krusius) in order to maximize the results from the complex measurements. Synergism between the four research groups of R. Shealy, C. Tang, C. Pollock, and P. Krusius on other efforts in materials growth, device processing, femtosecond measurements, and theoretical analysis has been growing as well. This evolution will be even more strongly reflected in the results anticipated for the second year of the current program period.

The three task investigators in the present program contributing to the real time signal processing theme, C. Bilardi, F. Luk and H. Torng, have held regular meetings, identified overlapping research areas, and started to define unifying research issues. Significant results have already been accomplished by this group despite the fact that they started in the Cornell JSEP program in May 1988.

B. DESCRIPTION OF SPECIAL ACCOMPLISHMENTS AND TECHNOLOGY TRANSITION

B.1 Femtosecond Carrier Processes in Compound Semiconductors

Significant accomplishments both in facilities and research have been achieved. While the new off-campus multi-reactor OMVPE facility is not expected to become operational until sometime during the second year, considerable progress has been made. The planning, fund acquisition, reactor construction, and the operational management structure have for the most part been completed, and interior construction is under way. Once completed this facility is likely to provide one of the most modern and versatile OMVPE growth capabilities for a variety of compound semiconductor materials. It can also serve as an example OMVPE facility designed for ultra safe handling of toxic hydride source gases.

Three new unique femtosecond laser sources have been developed. The first one, a high repetition rate UV femtosecond source is based on intracavity frequency doubling in a BaB_2O_4 crystal. The second broadly tunable red to mid IR femtosecond laser source employs resonant parametric oscillation and a KTiOPO_4 . The third source is a high power color center laser tunable in the 0.7 to 0.85 eV photon energy range. In the coming years these femtosecond laser sources will be used to study the dynamics of carrier processes in a variety of compound semiconductor materials, heterostructures, and devices. In parallel the ensemble Monte Carlo transport simulation approach has been extended to describe femtosecond optical interactions and dual carriers. It is expected that significant advances in the insight and understanding of femtosecond carrier processes will accrue from the joint design of femtosecond laser experiments and their subsequent detailed microscopic analysis using this dual carrier simulation capability.

Four Ph.D. degrees and one Master of Science degree have been awarded to graduate students working on JSEP research under this theme. Two of the JSEP investigators have spent their sabbatical leaves at research laboratories on problems related to JSEP research. C. Pollock worked for six months at the NRL in Washington, D.C. on new color center lasers. P. Krusius was for the entire academic year 1988/89 at IBM Research at Yorktown Heights working on hot carrier transport in the strain layer $\text{Ge}_x\text{Si}_{1-x}/\text{Si}$ materials system.

Largely building on years of leading research in optoelectronics under JSEP support, C. Tang served as the coordinator and principal investigator to a research proposal to DARPA as a response to the broad agency announcement BAA#89-09. In this document the establishment of the National Optoelectronic Materials Center with East and West Coast Divisions was proposed. Research groups primarily from Cornell University and the University of California Santa Barbara were involved.

B.2 Real Time Signal Processing

All task investigators in this theme started JSEP research just a year ago. Despite this significant accomplishments have already been achieved. The Naval Ocean Systems Center (San Diego, CA) is building a linear algebra parallel processor based on the work of F. Luk. The Boston based company, Computational Engineering, is considering building a systolic array processor, also based on F. Luk's work, for real time analysis of airplane wing flutter for the Air Force. A patent for the dispatch stack, a new method for speeding up RISC processors, was issued. H. Torng is continuing his work on the dispatch stack for real time computing systems with multiple functional units.

One Ph.D. degree and one Master of Engineering degree were awarded to graduate students working on JSEP research under this theme.

One of the JSEP investigators, H. Torng, is organizing the first meeting of "Project 2000", an interactive partnership between academia and industrial researchers working on high speed computers for the future. The first technical meeting is to be held on the Cornell Campus June 22-23, 1989.

C. DESCRIPTION OF INDIVIDUAL WORK UNITS

	Page
C.1 Femtosecond Carrier Processes in Compound Semiconductors	
Task #1 OMVPE Growth of III-V Alloys and Structures for New High Speed Electron Devices	5
Task #2 Femtosecond Laser Studies of Ultrafast Processes in Compound Semiconductors	10
Task #3 Ultrafast Studies of Device Structures and of Hot Carriers in Narrow Bandgap Semiconductors	15
Task #4 Physics of High Speed Meso Scale Compound Semi- conductor Devices	20
C.2 Real Time Signal Processing	
Task #5 Novel VLSI Algorithms and Architectures for High Data Rate Digital Filtering	27
Task #6 Fault Tolerant Signal Processing Architectures	31
Task #7 Real-Time Signal Processing Systems with Multiple Functional Units	36

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

TASK #: 1

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

OBJECTIVE

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

DISCUSSION OF STATE-OF-THE-ART

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

The conventional OMVPE process has produced many of the III-V materials and structures which find applications in high speed electron devices. The vast majority of published literature involves the AlGaAs materials system.

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

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

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

PROGRESS

In the past 1 year period progress has been made on several aspects of Cornell's JSEP materials program. The program has involved growth and characterization of materials, device feasibility studies, upgrading of the "JSEP OMVPE reactor" with industrial support, and finally, the design and construction of an expanded facility to support the needs of the JSEP program, as well as other programs requiring compound semiconductor materials.

Novel Heterostructures and Insulators on InP

One objective of this research is to exploit a recently discovered technique that increases the Schottky barrier of InP from 0.5 V to over 0.8 V [9]. If the Schottky barrier is stable, high performance MESFET's can be realized which take advantage of the high velocity and breakdown field strength of InP. It was found that an ozone exposed InP surface will produce approximately 10 Å of native oxide which enhances the Schottky barrier. Almost ideal IV curves exhibiting very low leakage current have been observed [10]. An ultraviolet light source was used in dry air at room temperature followed by a post anneal at 300°C in this process. The construction of a oxidation chamber with UV illumination has been completed to extend this process for use device applications.

While at an elevated temperature and in a nitrogen/oxygen ambient, the UV light interacting with the oxygen produces ozone that interacts with the InP surface. Preliminary experiments with InP samples were performed in N₂, O₂, and H₂ ambients at 500°C. When InP was annealed in the O₂ atmosphere, thin films were detected. The oxide thickness for InP annealed in the dark and under UV illumination was 200 and 350 Å, respectively. This data clearly shows that the modest UV exposure (~ 5 mW/cm²) will enhance the thermal oxidation process. At lower temperatures the ozone produced oxidation dominates the thermally activated process. Recently we have obtained rectifying Schottky barriers on ozone treated InP surfaces.

OMVPE Growth Wide Bandgap III-V Alloys

Several wide bandgap materials systems have been investigated in this program for their potential suitability as electron confinement layers in high speed electron device structures. These materials are also useful for high temperature transistor operation. These systems studied include AlGaInP/GaAs and AlGaP/GaP, both of which have been grown by OMVPE. The latter of these represents the widest bandgap alloy, excluding the nitrides, which is offered by the group of III-V materials. It is worth noting that high temperature operation (550°C) of heterojunction bipolar transistors have been reported with AlGaP/GaP structures [11].

The progress to date on these materials has included optimizing the growth conditions for high quality alloys and heterostructures. The optical properties have been determined with Raman, Electroreflectance, and Photoluminescence Spectroscopies. This study has produced the first reliable energy bandgap data on epitaxial AlGaP and AlGaInP. Finally, the microscopic features of defects present in AlGaInP films due to lattice mismatch have been studied with Transmission Electron Microscopy.

Raman Spectroscopy of 2 Dimensional Electron Gas (2DEG) Systems

An attempt was made to characterize the energy subbands in 2 dimensional electron systems as a means to characterize the space charge transfer at modulation doped heterojunctions. First, a series of experiments were successfully completed on samples containing multiple modulation doped interfaces. These measurements yielded the subband energy separation in the quantum size potential troughs, but attempts to measure Raman Scattering on single interfaces at room temperature failed. Unfortunately, the structure that is best suited to high speed device applications is the single modulation doped heterojunction. In order to enhance the optical interaction of the Raman probe and the 2DEG, the feasibility of performing Raman (defining the selection rules) in a waveguide geometry was established.

With the eventual goal of examining single heterojunctions containing a 2DEG, inelastic light scattering in optical waveguides was studied. It is anticipated that enhanced interactions with 2 dimensional electron gases will occur in this scattering geometry. The usual perturbation theory (2nd order) for light scattering in bulk crystals [12] has been extended to account for the added complexity of guided modes. This approach accounts for the observed differences between the spectra of bulk crystals and waveguides. Currently, these measurements have confirmed the theoretically predicted selection rules and work is underway to produce epitaxial structures by MBE which contain a single 2DEG in a surface waveguide and measure the electronic Raman scattering from such structures at room temperature.

POTENTIAL SCIENTIFIC IMPACT OF RESEARCH

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

DEGREES

None.

REFERENCES

1. "UV Laser Stimulated MOVPE of GaAs", P. Balk, Third Biennial OMVPE Workshop Abstracts, Brewster, MA (1987).
2. "Low Temperature Selective Epitaxy of III-V Compounds by Laser Assisted Chemical Vapor Deposition", N. H. Karam, H. Liu, I. Yoshida, B.-L. Jiang, and S. M. Bedair, Journal of Crystal Growth, 93, 254-258 (1988).
3. "A New Versatile, Large Size MOVPE Reactor", P. M. Frijlink, Journal of Crystal Growth, 93, 207-215, (1988).
4. "Characterization of AlGaAs/GaAs Heterostructures Grown in a Multi-chamber OMVPE Apparatus", J. R. Shealy, Journal of Crystal Growth, 87, 350-356 (1988).
5. "Raman Scattering Characterization of Interface Broadening in GaAs/AlAs Short Period Superlattices Grown by Molecular Beam Epitaxy", B. Jusserand, F. Alexandre, D. Paquet, and G. Le Roux, Applied Physics Letters, 47 (3) (August 1985).
6. "MOCVD Growth of AlGaInP at Atmospheric Pressure Using Triethylmetals and Phosphine", M. Ikeda, K. Nakano, Y. Mori, K. Kaneko, and N. Watanabe, Journal of Crystal Growth, 77, 380-385 (1986).
7. "Growth Temperature Dependent Atomic Arrangements and their Role on Band-Gap of InGaAlP Alloys Grown by MOCVD", C. Nozaki, Y. Ohba, H. Sugawara, S. Yasuami, and T. Nakanisi, Journal of Crystal Growth, 93, 406-411 (1988).

8. "Uniform and Abrupt InGaP/GaAs Selectively Doped Heterostructures Grown by MOVPE for HEMT ICs", T. Otori, M. Takechi, M. Suzuki, M. Takikawa, and J. Komeno, *Journal of Crystal Growth*, 93, 905-910 (1988).
9. "Schottky Barrier Height Enhancement in n-InP for MESFET Applications", A. A. Iliadis, to appear in *Proc. of GaAs Conference*, Atlanta, GA (1988).
10. "Schottky Diode and Field-Effect Transistor on InP", S. Loualiche, A. Ginoudi, H. L'Haridon, M. Salvi, A. Le Corre, D. Lecrosnier, and P. N. Favennec, *Applied Physics Letters*, 54 (13), 1238-1240, 27 March 1989.
11. "GaP/Al_xGa_{1-x}P Heterojunction Transistors for High-Temperature Electronic Applications", T. E. Zipperian and L. R. Dawson, *Journal of Applied Physics*, 54 (10) 6019-6025 (October 1983).
12. "Light Scattering Spectra of Solids", G. B. Wright, *Proceedings of the International Conference on Light Scattering Spectra of Solids*, New York University, NY, September 3-6, 1988.

ISEP PUBLICATIONS

1. "Organometallic Vapor Phase Epitaxial Growth of (Al_xGa_{1-x})_{0.5}In_{0.5}P and Its Heterostructures", D. P. Bour and J. R. Shealy, *IEEE Journal of Quantum Electronics*, 24 (9) 1856-1863 (September 1988).
2. "Direct Observation of Ordering in (GaIn)P", S. McKernan, B. C. DeCooman, C. B. Carter, D. P. Bour, and J. R. Shealy, *Journal of Materials Research*, 3 (3) 406-409 (May 1988).
3. "TEM Studies of Ordering in MOCVD-Grown (GaIn)P on GaAs", S. McKernan, B. C. DeCooman, C. B. Carter, D. P. Bour, and J. R. Shealy, *Proceedings of Materials Research Symposium*, 104, 637-640 (1988).
4. "Disordering, Intermixing, and Thermal Stability of GaInP/AlInP Superlattices and Alloys", S. O'Brien, D. P. Bour, and J. R. Shealy, *Applied Physics Letters*, 53 (19) 1859-1861 (November 7, 1988).
5. "Optical Investigation of Organometallic Vapor Phase Epitaxially Grown Al_xGa_{1-x}P", D. P. Bour, J. R. Shealy, A. Ksendzov, and F. Pollak, *Journal of Applied Physics*, 64 (11) 6456-6459 (1988).
6. "The Characterization of Misfit Dislocations at {100} Heterojunctions in III-V Compound Semiconductors", B. C. DeCooman, C. B. Carter, K. T. Chan, and J. R. Shealy, submitted to *Journal of Act. Mat.* (January 28, 1989).

FEMTOSECOND LASER STUDIES OF ULTRAFAST PROCESSES IN COMPOUND SEMICONDUCTORS

TASK #: 2

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

OBJECTIVE

The basic objective of this task is to study the dynamics of non-equilibrium electrons and holes in compound semiconductors and related structures using recently developed femtosecond laser sources and measurement techniques. There have been several major breakthroughs in the development of new femtosecond laser sources in the last year or so. Femtosecond sources in the uv to 315 nm and in the ir from 700 nm continuously to 4.5 μm have been developed for the first time. This wavelength flexibility coupled with the femtosecond optical correlation spectroscopic and hot-luminescence spectroscopic techniques previously developed in our laboratory will allow us to study a wide range of processes, materials, and structures.

DISCUSSION OF STATE-OF-THE-ART

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

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

Recent developments in Professor Pollock's laboratory and in our (Tang's) laboratory have led to the first truly broadly tunable femtosecond laser sources in the infrared and the first extension of the femtosecond sources into the uv. Pollock's source is based on color-center lasers and is tunable from 1.4 to 1.8 μm . Our source is based upon the optical parametric oscillator in the ir and is tunable from 700 nm to 4.5 μm at 10^8 Hz rate, but the

power level is in the mW range and lower than the color center lasers. The uv source is based upon the intra-cavity second harmonic generation technique using the new nonlinear optical crystal, β -barium borate, grown and fabricated in our laboratory. This led to nearly 100% conversion of the 630 nm fundamental light to 315 nm at the same pulse repetition rate. Thus, a uv femtosecond source with comparable characteristics as those of the Rh6G laser is now available for the first time. Combined with other dye lasers, this technique should extend the accessible wavelength range in the uv down to approximately 240 nm. Together with the femtosecond optical parametric oscillators, tunable femtosecond sources are now available for studying ultrafast processes from 240 nm to 4.5 μm . This vastly extends the accessible wavelength range for studying ultrafast dynamic processes.

The relaxation dynamics of nonequilibrium carriers excited by 2 eV photons in bulk GaAs have been studied extensively in the last few years. It is now generally agreed that the shortest relaxation component for electrons is 30 to 40 fs, as first reported by us over three years ago, not 13 fs as reported by the group at MIT. Our result was confirmed by Becker, et al. at Bell using the 6 fs laser. Because the 6 fs laser is at low repetition rates and very noisy, the data they obtained were very crude and not able to resolve the individual relaxation processes as well as could be done with our high repetition rate femtosecond laser. Nevertheless, they have been able to conclude that the shortest component was about 35 fs as we reported. On the whole, the relaxation dynamics of hot electrons in bulk GaAs up to approximately 0.5 eV in the conduction band are now fairly well understood, although there may still be some unresolved inconsistencies between the conclusions reached on the basis of the time-resolved experiments and a recent low temperature cw luminescence experiment. It is generally accepted that polar optical phonon scattering time is on the order of 150 to 200 fs and the Γ to L deformation potential is $0.8 - 0.9 \times 10^9$ eV/cm as we measured. The dynamics of electrons in the Γ valley of GaAs are, therefore, well understood. The dynamics of the holes are still far from understood, however. There have been some recent studies addressing this issue, although the picture is still far from clear. The main difficulty is that the hole relaxation process is expected to be even faster. With the femtosecond lasers restricted to near 2 eV, the holes generated are very near the top of the valence band and relax quickly. To study the hole dynamics optically, the holes must be created far from the zone center. This means femtosecond pulses at shorter wavelengths than 630 nm are needed. With the new uv source we have developed, this will be possible for the first time. This is one area that we plan to investigate in the future.

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

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

a subpicosecond time scale. Also, the effects of applied electric field on such structures can be studied with the help of such sources.

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

PROGRESS

(1) The completion of our study on the relaxation dynamics of nonequilibrium electrons excited by 2 eV photons in GaAs and AlGaAs: With the time resolution available from our visible femtosecond laser system and the extremely accurate and versatile optical correlation spectroscopy technique, a vast amount of data on the dynamics of nonequilibrium electrons photo-excited by 2 eV femtosecond photons have been collected and analyzed. A consistent and quantitative picture of the fate of such carriers is now more or less complete. Details are described in our earlier JSEP publications, Refs. [1-6] and current JSEP publications [4-6].

(2) The development of the first high repetition rate uv femtosecond laser source: Intracavity doubling in β -BaB₂O₄ (BBO) of femtosecond pulses into the ultraviolet with high efficiency is demonstrated (JSEP publication [3]). Pulse widths down to 43 fs at 10⁸ Hz repetition rate and outputs as high as 20 mW per arm of the femtosecond laser on a continuous-wave basis have been achieved. The ultra-violet pulse widths were determined through detailed cross-correlation measurements based on sum-frequency mixing to 210 nm in ultra-thin BBO crystals.

(3) The development of the first truly broadly tunable femtosecond laser source from the deep red to mid-ir (JSEP publication [7]): a singly resonant optical parametric oscillator based on a thin crystal of KTiOPO₄ is pumped by intracavity femtosecond pulses at 620 nm from a visible femtosecond laser. Oscillation results in stable continuous outputs of femtosecond pulses at 10⁸ Hz repetition rate and milliwatt average power levels in both signal and idler beams. Tuning from 820 - 920 nm and 1.90 - 2.54 μ m with a set of mirrors has been demonstrated. With multiple sets of mirrors, continuously tunable outputs from \sim 0.72 to \sim 4.5 μ m should be possible, making this a uniquely versatile femtosecond laser source.

POTENTIAL SCIENTIFIC IMPACT OF RESEARCH

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

femtosecond laser sources and techniques that might be used for a wide range of material and device studies.

DEGREES

None.

REFERENCES

1. "Femtosecond laser studies of ultrafast processes in semiconductors and large molecules", C. L. Tang, F. W. Wise, and I. A. Walmsley, invited, Special Issue on Femtosecond Lasers and Ultrashort Phenomena, Revue de Physique Applique, Vol.22 1695-1703(December, 1987).
2. "Picosecond relaxation of hot carrier distribution in GaAs/GaAsP strained-layer superlattices", D. C. Edelstein, A. J. Nozik, and C. L. Tang, App. Phys. Lett., Vol.51, 48-50(July 6, 1987).
3. "Application of linear prediction least-squares fitting to time-resolved optical spectroscopy", F. W. Wise, R. J. Rosker, G. L. Millhauser, and C. L. Tang, J. Quant. Elect.,QE-23, 1116(July, 1987).
4. "Experimental determination of hot-carrier scattering processes in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ", by F. W. Wise, I. A. Walmsley, and C. L. Tang, App. Phys. Lett., Vol.51, 605-607(August 24, 1987).
5. "Femtosecond relaxation dynamics of nonequilibrium carriers in GaAs and related compounds and structures", C. L. Tang, F. W. Wise, and I. A. Walmsley, pp.439-442, Proceedings of Fifth Int. Conf. on Hot Carriers in Semiconductors, Boston, MA (July 20-24, 1987); Solid-State Electronics, Vol.31,439-442(1988).
6. "Ensemble Monte-Carlo simulations of femtosecond energy relaxation in bulk GaAs and AlGaAs/GaAs quantum well structures", D. Bailey, C. Stanton, M. Artaki, K. Hess, F. Wise, and C. L. Tang, Proc. of Fifth Int. Conf. on Hot Carriers in Semiconductors, Boston, MA(July 20-24, 1987).

ISEP PUBLICATIONS

1. "Simultaneous formation of solitons and dispersive waves in a femtosecond ring dye laser", F. W. Wise, I. A. Walmsley, and C. L. Tang, Optics Lett. Vol. 13, 129-131(Feb., 1988).
2. "Measurement of ultrashort optical pulses with barium meta-borate", K. Cheng, W. Bosenberg, F. Wise, I. Walmsley, and C. L. Tang, App. Phys. Lett. (scheduled for Feb. 15, 1988).
3. "Femtosecond ultraviolet pulse generation in $\beta\text{-BaB}_2\text{O}_4$ ", D. C. Edelstein, E. S. Wachman, L. K. Cheng, W. R. Bosenberg, and C. L. Tang, App. Phys. Lett., Vol. 52, 2211-2213(June 27, 1988).

4. "Femtosecond spectroscopy of semiconductors and large molecules", C. L. Tang, F. Wise, and I. A. Walmsley, invited, J. of Modern Optics, Vol.35 (December 1988).
5. "Femtosecond laser studies of the relaxation dynamics of semiconductors and large molecules", C. L. Tang, F. W. Wise, M. J. Rosker, and I. A. Walmsley, invited review, IBM Journal of Research and Development, (scheduled for July, 1989).
6. "Subpicosecond luminescence study of hot-electron relaxation in GaAs quantum wells", F. W. Wise and C. L. Tang, Sol. State Comm. (accepted for publication).
7. "Broadly tunable high repetition rate femtosecond optical parametric oscillator", D. D. Edelstein, E.S. Wachman and C. L. Tang, App. Physics Lett. (scheduled for May 1, 1989); also, Post Deadline paper, CLEO '89, April 27, 1989, Baltimore, MD.

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

TASK #: 3

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

OBJECTIVE

The temporal relaxation of hot carriers in narrow bandgap semiconductors has been studied using infrared pulses of 60-200 fs duration. The femtosecond pulses are obtained from a color center laser, tunable from 0.7 to 0.85 eV. This wavelength range is directly useful for GaInAs based materials, and has been frequency doubled to provide a tunable probe in the 1.4 - 1.7 eV range. Tunability with femtosecond resolution provides the unique ability to measure relaxation rates of carriers lying anywhere between the bottom of the conduction band up to approximately the L and X valleys in certain semiconductors. Measurement of both energy and momentum relaxation are examined in bulk InGaAs as a function of probe energy.

The femtosecond pulses will also be used as an optical sampling probe for measurement of the switching dynamics of device structures. The electro-optic effect of III-V compounds will be used to sample the electrical waveform in a device, with subpicosecond resolution.

DISCUSSION OF STATE-OF-THE-ART

There has been widespread application of femtosecond pulses to the study of ultrafast phenomena in semiconductors. The techniques used for these studies (for example, pulse-probe or two pulse correlation) are well known, and are the techniques we will use in this proposed work. However most femtosecond spectroscopy has been done using fixed energy 2 eV photons. Our work is distinct in its use of pulses at a different energy (0.8 eV) and which are tunable. So in a sense, our work is establishing the state-of-the-art for femtosecond spectroscopy in narrow bandgap materials.

There has been extensive work in the measurement of the scattering rates of carriers in GaAs, AlGaAs, and quantum well structures using these materials. Tang (Cornell) [1], Ippen (MIT) [2], and Knox (Bell Labs) [3] have all pioneered techniques, instrumentation, and measurements on these samples. The current ability of optical probing is well established for determining both the rate at which carriers relax from the initial state, and the mechanisms (such as intervalley scattering) responsible for the short-lived carrier distributions in GaAs/AlGaAs. There has not been as much work on direct measurement of specific devices or structures with the aim of improving the mobility of the carriers in the material; most work to date has primarily focussed on determining the dynamics of the carrier relaxation within a given structure.

To date, because of the lack of femtosecond sources in the 0.8 eV photon energy range, little work has been done on the InGaAs system. Chemla (Bell Labs) [4] has recently concluded a study of exciton absorption in InGaAs quantum wells using 200 fs resolution pulses. His work concentrated only on

exciton interactions with the lattice and free carriers, and confirms earlier theory about the scaling of binding energy with bandgap and well dimension. There was no discussion of hot carrier dynamics.

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

PROGRESS

We have been working for one year under JSEP support. To date, we have developed our NaCl laser source to the point where it can deliver 75 fs pulses over the tuning range from 1.47 to 1.75 μm , which is an ideal range for studying InGaAs. In addition, we have frequency doubled this source to provide tunable 70 fs pulses in the 750-850 nm range for use in collaborative studies with other JSEP members. We stress that the femtosecond source is tunable, making it unique not only by the fact that the output is at approximately 0.8 eV, but by the fact that this output can be adjusted over approximately a 0.1 eV range. This is in contrast to most other prior femtosecond sources.

The direct output of the NaCl femtosecond laser has been used to study the temporal relaxation rates of hot carriers in InGaAs bulk material. Graduate students Chris Yakymyshyn and Brian Zook have been leading our effort to study GaInAs samples (provided by the Cornell crystal growth facilities). Using the two pulse correlation method first developed by Prof. Tang, we have probed the excited carrier lifetimes in GaInAs/InP wafers approximately 1 μm thick. The probe pulses have been tuned from 1.675 μm (which corresponds to photons with barely enough energy to promote an electron across the bandgap) to 1.53 μm , which corresponds to photon energy that is about 70 meV above the band gap (70 meV is roughly equal to the energy in two LO phonons). Results have not yet been fully analyzed or published, but the data shows clear trends. Fig. 1 shows the raw data from the two-pulse correlation experiments. The lower trace shows the nonlinear transmission of a sample excited with 1.53 μm photons (sufficient energy to place the carriers about 70 meV above the conduction band minimum). The sample transmission recovers within about 200 fs, indicating the time it takes for excited carriers to scatter out of their initial excited states. The upper trace shows the transmission of the sample when pumped by photons with energy near the bandgap energy. The relaxation time becomes noticeably longer, displaying long-lived tails. This response is consistent with a model of carrier scattering due to LO phonons: at low excitation energy there is not enough excess energy for the carriers to relax by LO phonon emission, so they cannot rapidly cool. Our measurements are preliminary at this point, and have not been fully analyzed. However, student Zook has developed a good deconvolution routine for separating the optical pulse shape from the observed relaxation, and Yakymyshyn has developed the laser and instrumentation to the point where our signal-to-noise ratio in the wings of the relaxation signal is on the order of 60 dB. The data we are obtaining is being shared with Prof. Krusius' group.

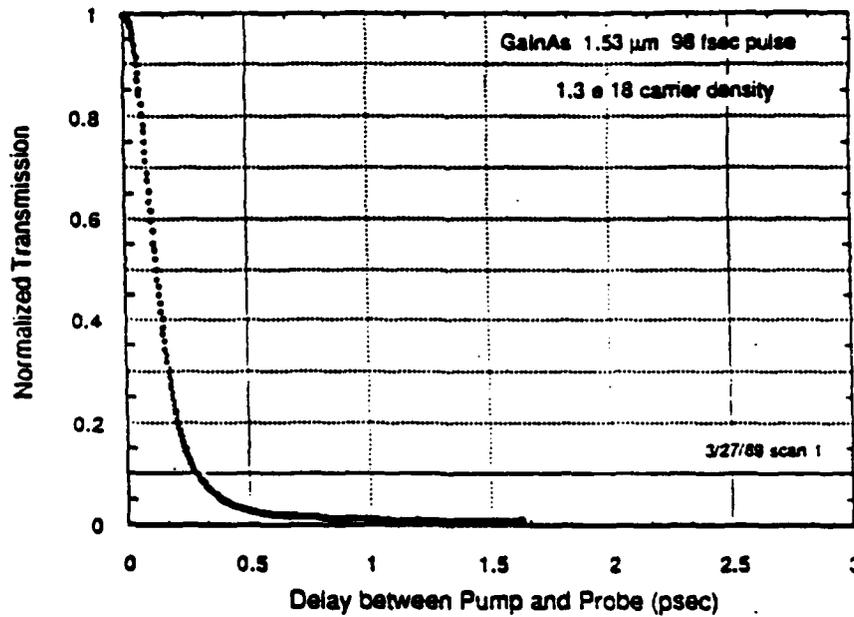
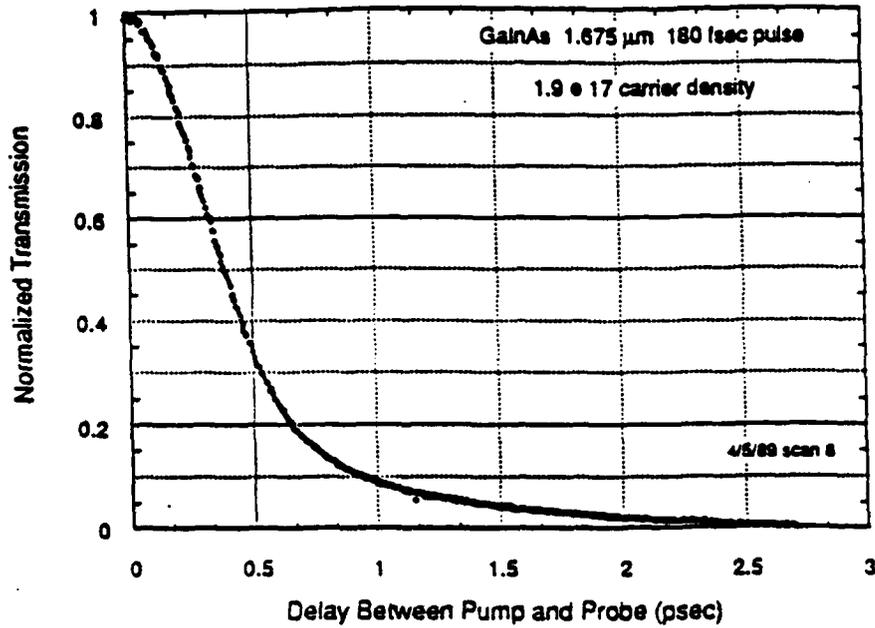


FIGURE 1. Normalized optical transmission as a function of delay between pump and probe pulses. Upper and lower figures for carrier densities of $1.9 \times 10^{17} \text{ cm}^{-3}$ and $1.3 \times 10^{18} \text{ cm}^{-3}$.

who are using it to compare with numeric simulations of the carrier dynamics for this material.

There is the possibility that some of the relaxation we are measuring is due to surface recombination, so Dr. Bill Schaff is growing some new MBE wafers which sandwich the GaInAs between appropriate transparent materials to eliminate such problems. We anticipate that by the end of the second year of this work, we will have essentially characterized the relaxation dynamics of bulk GaInAs as a function of carrier energy near the conduction band minimum, and as a function of carrier concentration.

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

There has also been a great deal of informal collaboration between my group and Prof. Tang's group. The two-pulse correlation technique we are using was pioneered by the Tang group, and we have benefitted from his experience. Also, when developing the second harmonic ability for our APM laser, we discussed many aspects of the problems of short pulse propagation in doubling crystals with his students, notably Walt Bosenberg.

SCIENTIFIC IMPACT

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

DEGREES

1. Joseph F. Pinto
Ph.D, Electrical Engineering, 1988
"Development of a color center laser in NaCl", Cornell University, 1988.

2. Christopher P. Yakymyshyn,
Ph.D, Electrical Engineering, 1989
"The additive pulse modelocked NaCl laser", Cornell University, 1988.
3. Efstratious T. Georgiou
Ph.D, Physics, 1989
"Formation and optical properties of new laser-active color centers in sodium chloride and potassium chloride", Cornell University, 1988.
4. Timothy J. Carrig
M.S., Applied Physics, 1988.

REFERENCES

1. C. L. Tang, F. W. Wise, and I. A. Walmsley, Solid State Electronics, 31, 439-442, (1988)
2. W-Z Lin, R. W. Schoenlein, J. G. Fujimoto, and E. P. Ippen, IEEE JQE 24, 267-275 (1988)
3. W. H. Knox, D. S. Chemla, G. Livescu, J. E. Cunningham, and J. E. Henry, Phys. Rev. Lett 61, 1290-1293, (1988)
4. M. Wegener, I. Bar-Joseph, G. Sucha, M. N. Islam, N. Sauer, T. Y. Chang, and D. S. Chemla, "Femtosecond dynamics of exciton absorption in the infrared InGaAs quantum wells", submitted for publication
5. M. J. Barlow, B. K. Ridley, M. J. Kane, and S. J. Bass, Solid State Electronics 31, 501-505 (1988)
6. K. Kash and J. Shah, Jour. Luminescence 30, 333-339 (1985)

ISEP PUBLICATIONS

1. C. P. Yakymyshyn, J. F. Pinto, and C. R. Pollock, "Additive-Pulse Mode-Locked NaCl:OH- Laser", Optics Letters 14, 621-623 (1989)
2. C. P. Yakymyshyn, and C. R. Pollock, "Frequency Doubled Additive Pulse Modelocked NaCl Laser", Optics Letters (to appear Aug. 1, 1989)

PHYSICS OF HIGH SPEED MESO SCALE COMPOUND SEMICONDUCTOR DEVICES

TASK #: 4

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

OBJECTIVE

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

DISCUSSION OF STATE-OF-THE-ART

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

A significant part of the hot carrier processes, namely those involving both electrons and holes, and their interactions under non-equilibrium conditions, has however received little attention. Recent experimental and theoretical indicators are pointing to the importance of electron-hole interactions and hole processes in the behavior of hot carriers. A few examples illustrate this statement. Optically generated hot electron and hole distributions have been shown to thermalize on drastically different time scales with the distribution function itself influencing thermalization dynamics [1]. Transport of minority carriers in dense semiconductor plasmas has been demonstrated to be so strongly affected by the electron-hole interaction that negative minority carrier mobilities have been measured (electron-hole drag)

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

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

PROGRESS

Research during the first year of this two year program has focused on two non-equilibrium carrier problems: (1) carrier transport across graded heterojunctions, and (2) femtosecond carrier thermalization processes in narrow band gap heterostructures. A third research issue for the current two year period, dual carrier transport (3), will replace problem (1) after its completion during the second year.

1. Non-equilibrium Carrier Transport Across Graded Heterojunction

Hot electron injection across graded and abrupt III-V compound semiconductor heterostructures has been explored using a self-consistent time-dependent ensemble Monte Carlo transport formulation. Electron bands are described within a position dependent $k.p$ framework in combination with the virtual crystal approximation to account for pseudobinary alloy effects. Scattering processes include intra and inter valley phonons (optical and acoustic), ionized impurities (Ridley screening), and the alloy effect (chemical disorder, Harrison and Hauser formulation). Scattering rates were calculated within the $k.p$ theory including all overlap integrals. This transport formulation, described in detail in an earlier and a current JSEP publication (see Ref.

[11] and JSEP publication [1]), has during the current period been implemented in a two-dimensional time-dependent computer code for non-equilibrium electron injection studies. This code, 2D-TCMC, allows one to simulate the behavior of non-equilibrium electrons in a rectangular domain comprised of several compound semiconductor regions including compositional grading and imbedded abrupt or graded heterojunctions. Ohmic and Schottky contacts can be placed anywhere on the periphery of the rectangular domain and with a small amount of rework also into the interior. Ohmic contacts are described via an interaction with external thermal reservoirs using correct injection statistics. Particle conservation is not explicitly enforced. Two-dimensional space charges are fully included. Charge assignment is performed using the cloud-in-cell method and Poisson's equation for the ensemble is solved using Hockney's Fast Fourier Transform based technique.

Hot electron transport processes across laterally uniform one-dimensional and laterally non-uniform two-dimensional graded heterojunctions have been explored. It was found that ensemble phenomena dominate the carrier injection process across the heterojunction influencing distribution functions, drift velocities, and ballistic carrier fractions. Space charges and current continuity play a crucial role. In a study of hot electron injection in the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ materials system across one-dimensional heterojunctions it was established that drift velocities downstream from the heterojunction can vary by a factor of four depending on the state of the space charge at the junction. The injection efficiency can be influenced via grading, doping density, temperature, and applied voltage (JSEP publication [2]). Flat band conditions desirable for high current drive device applications are not attainable for any applied voltages, if the injection junction has not been correctly designed, an important finding for high speed device design (JSEP publications [2,3]). Lateral electrodes can be used to shape the space charge at the heterojunction in two-dimensional device structures. Lateral control electrode placement can influence the device current almost by an order of magnitude for a similar current modulation capability. Two-dimensional space charge phenomena have been studied in vertical FET (VFET) devices (JSEP publication [4]). Both steady state and transient ballistic carrier transport across heterojunctions has been shown to be controlled by lateral space charges. Our findings also explain why measured cut-off frequencies have not reached expected values (JSEP publications [5,6]).

Based on these results it is now possible to understand hot electron injection in a variety of compound semiconductor devices. Optimization issues for a class of heterostructure unipolar devices, including the VFET, the permeable base transistor (PBT) and the planar doped barrier transistor (PDBT) are currently being simulated on the IBM 3090-600E supercomputer in order to predict their ultimate high speed potential (JSEP publications [7,8]).

2. Femtosecond Carrier Thermalization

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

The physics of hot carrier thermalization in these thin films is described with a dual carrier ensemble Monte Carlo formulation including both electrons and holes. Because of the homogeneity of the films and the slowness of ambipolar diffusion compared to the thermalization times, it is not necessary to include spatial inhomogeneities. Electron and hole bands are again calculated within the $k.p$ formulation with corrections for higher bands included through second order perturbation theory. Due to the multiplicity and warping these expressions are quite involved. Both acoustic and optical phonons are included inelastically through the long range (polar, piezoelectric) and short range (deformation potential) interactions. Electron-electron, hole-hole and electron-hole scattering are included with the simplification that carriers are not permitted to be scattered from the valence bands to the conduction bands, or vice versa. Scattering between valence bands is fully included. All interactions are statically screened. For this pseudobinary materials system it is necessary to include both chemical and structural disorder when treating alloy scattering because of local anion site related bond length and angle distortions. This can be accomplished using the molecular coherent potential approximation [12]. Optical interactions are included within first order time dependent perturbation theory with full inclusion of photon polarization effects. Derivations of all energy dependent scattering rates and polarization dependent optical interactions have been completed along with the $k.p$ band formulation. Computer code is being developed for the simulation of the tunable pump/probe experiment including photon polarization effects. Considerable leverage is derived from the 2D-TCMC code developed for non-equilibrium electron transport studies (see section 1.). It is expected that this code development and the analysis of the first measurement will be completed prior to the end of the current calendar year.

With the adopted synergistic experimental and theoretical design the energy dependence of the dual carrier thermalization dynamics can be studied without a condensation of the data into a set of effective exponential time constants, a considerable advantage compared to earlier pump/probe studies.

3. Dual Carrier Transport

As mentioned, above dual carrier transport issues, band structure, carrier dynamics and scattering processes, will be based on work performed in areas (1) and (2). In addition it will be necessary to develop a formulation for impact ionization. A simple threshold field dependent model derived by Keldysh will be used as the starting point. Space charge effects and carrier plasma interactions can be included in a natural fashion because of the self-consistency and adopted microscopic model for device boundaries. Work in this area will begin during the second year of the current program after the completion of the carrier injection studies.

POTENTIAL SCIENTIFIC IMPACT OF RESEARCH

In order to fully understand femtosecond electronic and optical processes in compound semiconductors it is necessary to examine non-equilibrium electrons and holes simultaneously. The dual carrier ensemble particle methods developed in this task will allow us to analyze these processes in full detail in realistic inhomogeneous heterostructures. Although these methods will be extremely complex, and take considerable amount of time to

develop, they are necessary for the unambiguous interpretation of the femto-second laser measurement of carrier processes in compound semiconductor heterostructures. Once these methods have been developed, correlations with optical and transport measurements performed, and their validity established, they can be applied to the analysis, design and optimization of a large number of ultrafast electronic and optoelectronic devices.

DEGREES

1. A. Al-Omar,
Ph.D., Electrical Engineering, 1988
"Hot Electron Transport in Nanometer Scale Graded Ternary III-V Semiconductor Devices", Cornell University, 1988.

REFERENCES

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

9. "A Self-Consistent Particle Simulation for (AlGa)As/GaAs HBTs with Improved Base-Collector Structures", R. Katoh, M. Kurata, and J. Yoshida, Proceedings of the International Electron Devices Meeting, p. 248 (1987).
10. "Monte Carlo Simulation of AlGaAs/GaAs Heterojunction Bipolar Transistors", K. Tomizawa, Y. Awano, H. Hashizume, IEEE Electron Devices Letters, 5, #9, 362 (1984).
11. "Self-Consistent Monte Carlo Study of High Field Transport in Graded Heterostructures", A. Al-Omar and J. P. Krusius, Journal of Applied Physics, 62, 9, 3825 (1987).
12. "Molecular Coherent Potential Approximation for Zinc-Blende Pseudobinary Alloys", R. J. Lempert, K. C. Hass, and H. Ehrenreich, Physical Review, B36, 2, 1111 (1987).

ISEP PUBLICATIONS

1. "Hot Electron Transport in Nanometer Scale Graded Ternary III-V Semiconductor Devices", A. Al-Omar, PhD Thesis, Cornell University, Ithaca, NY, 1988, 343 pages.
2. "Microscopic High Field Transport in Graded Heterostructures", A. Al-Omar and J. P. Krusius, Solid State Electronics, 31, #3/4, 329 (1988).
3. "Conditions for Space Charge Reversal at Thermionic Heterojunctions Designed for Ballistic Electron Injection", A. Al-Omar and J. P. Krusius, IEEE Electron Device Letters, 9, 2, 81 (1988).
4. "Space Charge Effects on Heterojunction Cathode (Al:Ga)As Gunn Oscillators", A. Al-Omar J. P. Krusius, Z. Greenwald, A. R. Calawa, and L. F. Eastman, Proceedings of the IEEE/Cornell Conference on Advanced Concepts in High Speed Semiconductor Devices and Circuits, p. 365 (1988).
5. "Investigation of Lateral Effects on Electron Launching in Graded AlGaAs-GaAs Heterostructure Devices Using Two-Dimensional Self-Consistent Monte Carlo Method", S. Weinzierl and J. P. Krusius, accepted for presentation at the NASECODE VI Conference, Dublin, Ireland, July 11-14, 1989, and accepted for publication in proceedings.
6. "Lateral Space-Charge Effects on Ballistic Electron Transport Across Graded Heterojunctions", S. Weinzierl and J. P. Krusius, accepted for presentation at the 6th International Conference on Hot Carriers, July 1989, accepted for publication in proceedings in Solid State Electronics.
7. "Optimization of the Design of the Heterojunction Vertical Field Effect Transistor Using a Two-Dimensional Self-Consistent Monte Carlo Method", S. Weinzierl and J. P. Krusius, accepted for presentation at the 1989 IEEE/Cornell Conference on High Speed Devices and Circuits, Ithaca, NY, August 1989, and to be submitted for proceedings.

8. "Non-Equilibrium Carrier Transport Phenomena in High Speed Compound Semiconductor Heterostructure Devices", S. Weinzierl and J. P. Krusius, manuscript in progress.

NOVEL VLSI ALGORITHMS AND ARCHITECTURES FOR HIGH DATA RATE DIGITAL FILTERING

TASK #: 5

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

OBJECTIVE

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

DISCUSSION OF STATE-OF-THE-ART

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

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

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

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

PROGRESS

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

Filtering and Prefix Computations.

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

The desire to obtain a deeper understanding of the relationship between filtering and prefix computation has motivated an extensive investigation of the latter. Various resource tradeoffs have been completely characterized in terms of algebraic properties of the semigroup underlying the prefix computation (Ref. [8], JSEP publication [5]). These results have independent interest since variants of the prefix operation (e.g., fetch-and-add on the Ultracomputer, scan on the Connection Machine, and multiprefix on the Fluent Machine) play an important role in parallel computing.

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

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

Multidimensional Signal Processing.

VLSI signal processing can be extended to multidimensional signals. In JSEP publication [3] we have taken a step in this direction for the *multidimensional discrete Fourier transform*. The cases of complex arithmetic and modular arithmetic have both been investigated. Area-time optimal designs have been developed for a wide range of computation times. Previously published lower bounds on the area-time performance were based on fallacious arguments, and completely new arguments have been developed to establish performance lower bounds.

The results published in JSEP publication [3], in common with almost all the studies on DFTs, make specific assumptions on the factorability of the size of the transform. In recent unpublished work we have succeeded in constructing optimal circuits in the general case.

Distributed Implementation of Shared Memory.

The design of a special-purpose VLSI system typically involves the choice of a suitable parallel algorithm and architecture for the desired task. The architecture should support well the execution of that algorithm, and have the smallest layout compatible with this requirement. It is often the case that parallel algorithms have already been proposed in the literature. However, most algorithms are developed for a shared-memory model of computation such as the *parallel random access machine* (PRAM). It would be of great interest if one could automatically transform a PRAM algorithm into a VLSI algorithm.

As a step in this direction, together with Kieran Herley (a graduate student supported by JSEP), we have studied the problem of simulating a PRAM

on a *bounded-degree network*, a model more suited to VLSI implementation. Optimal simulation schemes have been obtained in JSEP publication [2] for the case in which the memory size grows at least as a polynomial function in the number of processing elements. Further results have been obtained recently by Herley in JSEP publication [6] in the case of smaller memory. It should be mentioned that this work is technically difficult and that further progress depends on the solution of some deep questions in graph theory.

Lower Bounds.

Communication is often the critical factor limiting the speed of parallel algorithms. In JSEP publication [4], the constraints on the computation time posed by propagation of information are analyzed for a specific class of functions. The goal is to gain a better understanding on what properties of a function constrain its parallel computation time. A general technique is developed to obtain lower bound on the parallel computation time of *monotone* boolean functions in terms of the length of their largest prime implicant or prime clause.

SCIENTIFIC IMPACT OF RESEARCH

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

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

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

The problem of translating shared-memory algorithms into distributed-memory algorithms is of fundamental importance. Any implementation of a large memory is bound by technological constraints to be a distributed one. Thus, the work reported in JSEP publications [2] and [6] has consequences for VLSI implementation of shared-memory algorithms.

DEGREES

Kieran Herley, Ph.D. expected in 1989.

REFERENCES

- [1] C.D. Thompson, "A Complexity Theory for VLSI", Ph.D. Thesis, Department of Computer Science, Carnegie-Mellon University (Aug. 1980).
- [2] R.P. Brent and H.T. Kung, "The Chip Complexity of Binary Arithmetic", *J. ACM*, 28 (3) 521-534 (July 1981).

- [3] J.D. Ullman, *Computational Aspects of VLSI*, Computer Science Press, Rockville, MD. 1983.
- [4] G.M. Baudet, F.P. Preparata and J.E. Vuillemin, "Area-Time Optimal VLSI Circuits for Convolution", *IEEE Trans. on Computers*, C-32 (7) 684-688 (July 1983).
- [5] C.D. Thompson, "Fourier Transforms in VLSI", *IEEE Transactions on Computer*, 32 (1983), pp. 1047-1057.
- [6] G. Bilardi and M. Sarrafzadeh, "Optimal VLSI Circuits for Discrete Fourier Transform", in *Advances in Computing Research*, Vol. 4, (F.P. Preparata, Editor), JAI Press, Greenwich, CT, pp. 87-101.
- [7] G. Bilardi and F.P. Preparata, "Digital Filtering VLSI", (invited paper), *Proc. Aegean Workshop on Computing: VLSI Algorithms and Architectures*, Loutraki, Greece, 1-11, Springer Verlag (July 1986).
- [8] G. Bilardi and F.P. Preparata, "Size-Time Complexity of Boolean Networks for Prefix Computation", *Journal of ACM*, 36 (2) 362-282, April 1989.
- [9] R.E. Ladner and M.J. Fisher, "Parallel Prefix Computation", *J. ACM*, 27 (4) 831-838 (Oct. 1980).

ISEP PUBLICATIONS

- [1] G. Bilardi, Area vs. time: A tradeoff in a VLSI computer, *Engineering Cornell Quarterly*, 23(1) 26-30, Autumn 1988.
- [2] K. Herley and G. Bilardi, Deterministic simulations of PRAMs on bounded-degree networks, *Proceedings of the 26th Annual Allerton Conference on Communication, Control, and Computing*, 1084-1093, Monticello, Illinois, September 1988. An extended version appears in the Technical Report 88-951, Department of Computer Science, Cornell University, November 1988.
- [3] G. Bilardi, S. Hornick, and M. Sarrafzadeh, Optimal VLSI architectures for multidimensional DFT, *1989 ACM Symposium on Parallel Algorithms and Architectures*, Santa Fe, New Mexico, June 1989, 265-272.
- [4] G. Bilardi and A. Moitra, Time lower bounds for CREW-PRAM computation of monotone functions, *International Symposium on Automata, Languages, and Programming*, Stresa, Italy, July 1989. Also Technical Report 89-1012, Department of Computer Science, Cornell University, May 1989.
- [5] G. Bilardi and F.P. Preparata, Characterization of associative operations with prefix circuits of constant depth and linear size, *SIAM Journal on Computing*, accepted for publication.
- [6] K. Herley, Efficient simulations of small shared memories on bounded degree networks, Extended Abstract, accepted at *30th Annual Symposium on Foundations of Computer Science*, North Carolina, October 1989.

FAULT TOLERANT SIGNAL PROCESSING ARCHITECTURES

TASK #: 6

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

OBJECTIVE

Systolic arrays have emerged as a preferred means for performing the many matrix computations central to real-time signal processing [12]. Indeed, the Naval Ocean Systems Center in San Diego is building a machine, the Systolic Linear Algebra Parallel Processor (SLAPP) [3], based on our theoretical design. Undoubtedly, systolic arrays will also be used in a great many other environments. What we need to do, then, is provide a wide range of cost effective fault tolerance options to meet the varying needs of all users of systolic arrays. What are these needs? At the minimum, error detection is necessary because errors in matrix computations are essentially undetectable by a mere examination of the results. For the real-time environment, we need totally reliable arrays that do not reduce the throughput of the original array.

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

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

In this task, we are primarily concerned with hard and soft processor calculation errors. We ignore transmission and memory errors because effective error correcting codes such as Hamming codes already exist to deal with these errors. We propose to examine critically both existing and new techniques for achieving fault tolerance, first in systolic arrays and then in real-time signal processing systems. Further, we propose to formulate and evaluate schemes for fault-tolerance in such arrays and systems, including the digital filtering structures being investigated by Bilardi and the configurations with multiple functional units being studied by Torng.

DISCUSSION OF STATE-OF-THE-ART

Algorithm-based fault tolerance, proposed by Jacob Abraham and his students at the University of Illinois [1,2,4,5], is a technique specially tailored for systolic algorithms and architectures. By encoding the input data as checksums and by redesigning algorithms for the encoded data, one can detect and correct transient errors that have occurred during the computations. This approach requires a very low overhead and uses simple arithmetic; Abraham et al. apply it to basic operations such as matrix-matrix multiplication, LU decomposition and matrix inversion. For correcting a transient error in Gaussian elimination, they propose a computation rollback to the point just before the error occurred. However, it is hard to execute rollbacks on systolic arrays. In [8] we show how to avoid rollbacks by computing the correct decomposition from the erroneous one.

In [10] we extend Abraham's checksum scheme for the LU decomposition to a unified scheme for three different triangularization procedures: LU decomposition, Gaussian elimination with pairwise pivoting, and QR decomposition. We show how to represent the error as a rank-one perturbation to the data, and develop a new error model where the occurrence time of the error is not involved.

Although in exact arithmetic the checksum scheme works well (an inconsistent checksum indicates the presence of a transient error), in floating point arithmetic an undesirable growth of rounding errors may cause confusion. There is a need to establish a tolerance to decide if an inconsistent checksum were caused by a (large) transient error or by roundoffs. In [9] we analyze the effects of rounding errors on the checksum scheme and establish a tolerance for transient error detection. Furthermore, we show that the tolerance is necessarily large for the LU decomposition and for Gaussian elimination with pairwise pivoting, but is acceptably small for the QR decomposition.

The guiding principle of virtual redundancy is the same as that of any space redundancy technique: make the same calculation on different processors and compare results to detect and correct errors. However, instead of replicating the hardware, one replicates the data and takes advantage of idle processors to make the redundant calculations (cf. Kim and Reddy [6]).

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

PROGRESS

We introduced a new algorithm-based fault tolerance technique specifically designed for use in recursive least squares minimization. Through monitoring a single scalar $x(n)$, we get error detection. The same quantity can also be used as an indicator for correction. This technique applies in equally effective fashion to many other systolic arrays, as it provides the same properties for the QR decomposition phase of the algorithms. Our scheme does not

require the summation process during the decoding stage; it requires only the monitoring of $x(n)$. Furthermore, due to the flow of data into the array, which is in a wavefront pattern, $x(n)$ is the only reasonable quantity to examine continually. The chief attribute of our technique is its remarkable simplicity.

Error correction has proved to be a much more difficult problem to solve than error detection when using weighted checksums. We provided a theoretical basis for the correction problem. We showed that for a distance $d+1$ weighted checksum scheme, if a maximum of $\text{floor}(d/2)$ errors ensue then we can determine exactly how many errors have occurred. We further demonstrated that in this case we can correct the errors and gave a procedure for doing so.

To avoid numerical overflows, in [4] and [5], two methods were proposed that use modular arithmetic to compute weighted checksums. A new scheme was derived by us. For means of comparing these three methods, we consider fixed point arithmetic since it was used in [4] and [5]. Let r denote the word length. Huang and Abraham [4] suggested that we compute the actual checksum values modulo 2^r , while in Jou and Abraham [5], a similar technique was proposed where the checksum column is computed modulo 2^r and the weighted checksum column is computed modulo N , where N is the largest prime less than 2^{r+1} . They proved that for these choices of weights, the schemes detect errors as desired. One difference between the methods is that the new one is all pre-processing while the other two add extra time to the algorithm as the checksums must be computed using modular arithmetic. Furthermore, the prime that is selected by Jou and Abraham is much larger than the prime in our work. For example, in [5], if the word length $r = 32$ then the prime $N = 8,589,934,583$. Our scheme is relatively independent of the word length; it depends upon the values of n and d . For $n = 1000$, a large value in light of the dimensions required by current signal processing problems, and $d = 50$, the prime $p = 1051$. We should also compare the sizes of the weight elements for various weight generating schemes. Suppose that $n = 500$ and $d = 10$. Then the largest weight generated by the scheme in [5] equals 2^{4491} , and by the scheme in [8], equals 500^9 . For our new scheme, the smallest prime p satisfying $p > n+d = 510$ is $p = 521$. Therefore, every weight element will be bounded above by 521. Thus, we see that for this moderately sized problem, our new scheme generates a parity-check matrix whose elements are likewise of moderate size.

SCIENTIFIC IMPACT OF RESEARCH

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

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

We believe that the investigations being proposed will yield approaches, more cost effective and flexible than traditional techniques such as modular redundancy and quadded logic, and may initiate new areas in the study of fault tolerance.

DEGREES

1. Joseph Cavallaro
Ph.D., Electrical Engineering, 1988
"VLSI Cordic Processor Architectures for the Singular Value Decomposition", Cornell University, 1988.

REFERENCES

- [1] J.A. Abraham, "Fault Tolerance Techniques for Highly Parallel Signal Processing Architectures", *Highly Parallel Signal Processing Architectures*, K. Bromley, Ed., Proc. SPIE, 614, 49-65 (1986).
- [2] C.Y. Chen and J.A. Abraham, "Fault-Tolerant Systems for the Computation of Eigenvalues and Singular Values", *Advanced Algorithms and Architectures for Signal Processing I*, J.M. Speiser, Ed., Proc. SPIE, 696, 228-237 (1986).
- [3] B.L. Drake, F.T. Luk, J.M. Speiser and J.J. Symanski, "SLAPP: A Systolic Linear Algebra Parallel Processor", *IEEE Computer (Special Issue on Systolic Arrays)* 45-49 (July 1987).
- [4] K-H. Huang and J.A. Abraham, "Algorithm-Based Fault Tolerance for Matrix Operations", *IEEE Trans. Comput.*, C-33, 518-528 (1984).
- [5] J-Y. Jou and J.A. Abraham, "Fault-Tolerant Matrix Arithmetic and Signal Processing on Highly Concurrent Computing Structures", *Proc. IEEE*, 74, 732-741 (1986).
- [6] J.H. Kim and S.M. Reddy, "A Fault-Tolerant Systolic Array Using TMR Method", *Proc. ICCD*, 769-773 (1985).
- [7] H.T. Kung and M.S. Lam, "Fault-Tolerant VLSI Systolic Arrays and Two-Level Pipelining", *Real Time Signal Processing VI*, K. Bromley, Ed., Proc. SPIE, 431, 143-158 (1983).
- [8] F.T. Luk, "Algorithm-Based Fault Tolerance for Parallel Matrix Equation Solvers", *Real Time Signal Processing VIII*, W.J. Miceli and K. Bromley, Eds., Proc. SPIE, 564, 49-53 (1985).
- [9] F.T. Luk and H. Park, "Analysis of Algorithm-Based Fault Tolerance Techniques", *Advanced Algorithms and Architectures for Signal Processing I*, J.M. Speiser, Ed., Proc. SPIE, 696, 222-227 (1986).
- [10] F.T. Luk and H. Park, "Fault-Tolerant Matrix Triangularizations on Systolic Arrays", *IEEE Trans. on Computers*, to appear.
- [11] F.T. Luk and E.K. Torng, "Fault-Tolerance Techniques for Systolic Arrays", *Real Time Signal Processing X*, J.P. Letellier, Ed., Proc. SPIE, 827, Paper 5 (1987).

- [12] J.M. Speiser and H.J. Whitehouse, "A Review of Signal Processing with Systolic Arrays", Real Time Signal Processing VI, K. Bromley, Ed., Proc. SPIE 431, 2-6 (1983).

ISEP PUBLICATIONS

- [1] F.T. Luk, Editor, Proceedings of SPIE Vol. 975, Advanced Algorithms and Architectures for Signal Processing III, SPIE - The International Society for Optical Engineering, Bellingham, Washington, 1988.
- [2] "A Theoretical Foundation for the Weighted Checksum Scheme", C.J. Anfinson, R.P. Brent, and F.T. Luk, Proceedings of SPIE Vol. 975, Advanced Algorithms and Architectures for Signal Processing III, 10-18 (1988).
- [3] "A Novel Fault Tolerance Technique for Recursive Least Squares Minimization:", C.J. Anfinson, F.T. Luk, and E.K. Tornig, Proceedings of SPIE Vol. 975, Advanced Algorithms and Architectures for Signal Processing III, 268-276 (1988).
- [4] "VLSI Arrays for Signal Processing", F.T. Luk and D.E. Schimmel, Engineering: Cornell Quarterly, 23, 1, 15-17 (Autumn 1988).
- [5] "A Novel Fault Tolerance Technique for Recursive Least Squares Minimization", F.T. Luk, E.K. Tornig, and C.J. Anfinson, Journal of VLSI Signal Processing, to appear in Vol. 1, number 3 (1989).
- [6] "Checksum Schemes for Fault Tolerant Systolic Computing", R.P. Brent, F.T. Luk, and C.J. Anfinson, Paper presented at Conference on Mathematics in Signal Processing, Warwick, England (December 1988). Proceedings will appear in 1989.
- [7] "Computing Generalized Canonical Correlations", L.M. Ewerbring and F.T. Luk, Paper presented at NATO Advanced Study Institute on Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms, Leuven, Belgium (August 1988). Proceedings will appear in 1989.

REAL-TIME SIGNAL PROCESSING SYSTEMS WITH MULTIPLE FUNCTIONAL UNITS

TASK #: 7

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

OBJECTIVE

Our task, initiated in May 1988, addresses one of the pressing problems, if not THE problem, in the field of signal processing: how to perform computations fast enough to meet stringent real-time requirements.

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

Concurrent processing can be classified according to two levels: task and instruction. In task-level concurrent processing, a given job is partitioned into partially ordered subtasks, and these subtasks are assigned to cooperating processors. These processors communicate, as they must, through shared memory and/or other interconnection mechanisms.

In instruction-level concurrent processing, multiple functional units, installed in a single processor, are assigned instructions from a single instruction stream. In such a structure, several instructions are being processed at the same time, and this brings about performance enhancement. In our study, we have concentrated on systems with multiple functional units (SMFU).

CRAY and CDC machines achieve their high performances through exactly such configurations. Recently announced RISC processors: Intel 80960, Motorola 88000 and AMD 29000, also employ multiple functional units. For example, Motorola 88000 processor has one integer unit, one floating-point unit, and 6 optional special-function units.

These special functional units may very well be the SVD processors, being studied by Frank Luk, and TRT filters, being investigated by G. Bilardi. They may serve as important ingredients in signal processing computations.

Our task investigates the design, programming and implementation of real-time systems through instruction-level concurrent processing. We have been working on problems that computer architects at various organizations have just begun to investigate.

DISCUSSION OF STATE-OF-THE-ART

Concurrent processing is an approach to meeting the "real-time challenge" in signal processing; it takes many forms: parallel processors, systems with multiple functional units (SMFU's), and various combinations of the previous two.

Parallel Processors

Researchers have been investigating parallel processors intensively. A well connected set of processors constitutes an inviting scheme to reducing computation latency. Many applications have been identified for such structures; problems, however, also remain: task partition and scheduling, task synchronization, interprocessor interconnection and communication, system flexibility with respect to task variations, incorporation of specialized functional units, and under-utilization of processors.

Systems with Multiple Functional Units (SMFU)

A block diagram of a system with multiple functional units (SMFU) is given in Figure 1. Each functional unit is designed to perform a specific set of arithmetic or logical operations in an optimum fashion. High-speed registers serve as a buffer between the memory and the functional units. These registers supply operands to the functional units and receive results from them; they also load from and write into the main memory. This constitutes a register-register architecture, with the main memory being accessed through load/store operations.

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

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

SMFU's, such as CRAY [1] and Floating Point Systems [2] machines, are designed for general purpose applications. Recently announced micro-processors such as Motorola 88000, Intel 80960, and AMD 29000 also implement this configuration; furthermore, these processors represent the new wave of Reduced Instruction Set Computers (RISC). SMFU's for specific military applications have been explored and developed at many organizations.

SMFU's achieves high throughput and low latency without incurring the cost of performing task partition, task scheduling and task synchronization; the system reacts easily to task variations.

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

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

Two problems remain: The bandwidth between the main memory and the execution complex has to be high enough so that an adequate supply of instructions and data is maintained. This problem is addressed by among other things splitting the main memory into an instruction memory and a data memory.

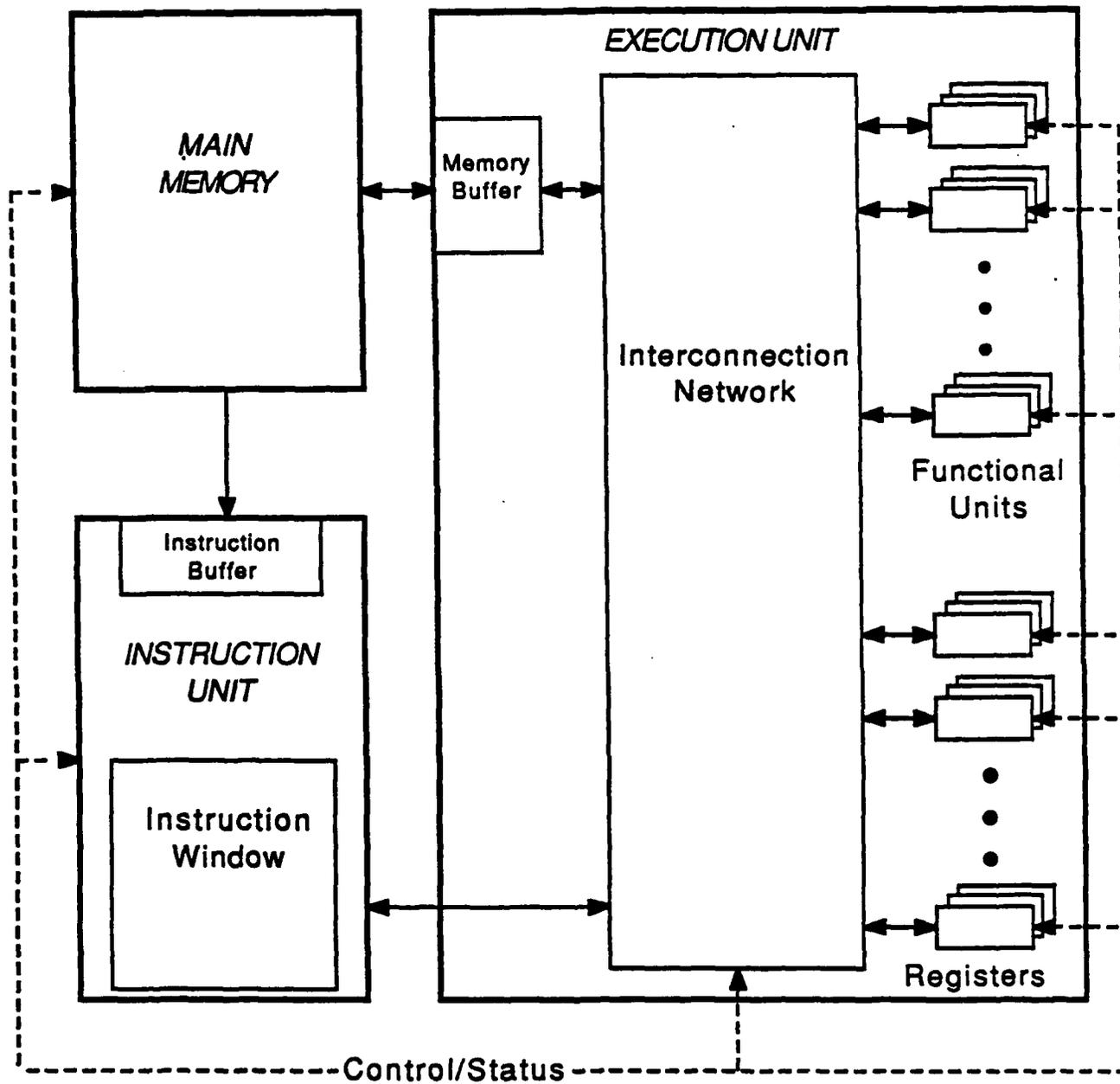


FIGURE 1: Multiple Functional Unit Processor:
General Architecture

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

To elevate SMFU's to a higher level of performance, as we must, we have to address the instruction issuance issue. We do that in this task.

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

Dispatch Stack

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

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

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

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

The DS is conceptually a "device", which displays and checks dependencies among instructions in an instruction stream; it can be viewed as a window on the instruction stream with certain capabilities. The dispatch stack can be either realized in software or hardware and it will be exploited in the programming, design and implementation of SMFU's for real-time signal processing.

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

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

Interrupt handling mechanisms are evaluated by the following three factors:

Precise processor state: When an interrupt request is received, the processor must have the capability to save its processor state precisely;

Latency: An interrupt handling mechanism should be judged by the latency between the receipt of an interrupt request and the completion of saving the processor state.

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

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

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

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

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

PROGRESS

We have received a patent: "Instruction Issuing Mechanism for Processors with Multiple Functional Units", U. S. Patent 4,807,115, February 21, 1989. This patent essentially covers the Dispatch Stack that we have developed. We feel that we have been addressing the instruction issuing question that processor designers, especially those concerned with RISC processors, have just begun to explore; this is confirmed by our conversations with several companies.

The Dispatch Stack (DS) that we have formulated performs dynamic instruction scheduling at run time. It is critical that the requisite data dependency checking among the instructions that reside in the DS, the fetching of instructions into the open slots in the DS, and the identification of instructions that can be issued at a given machine cycle must be carried out successfully without elongating the machine cycle time, which can be in the range of 10 nanoseconds for certain processors. We have completed a study on the implementation of the Dispatch Stack. And have formulated ways to realize time-critical circuits so that all functions specified for the Dispatch Stack be performed even in an extremely short machine cycle.

Multiple and out-of-order instruction issuances for a modern processor, as implemented with the proposed DS, brings about situations when several instructions are being executed concurrently and an instruction may be completed before instructions that appear before it in the instruction stream. Precise and prompt interrupt handling is essential if these high-performance processors are to be successfully integrated into real-time computing systems. It is also important that the installation of interrupt handling mechanism do not degrade system performance in the absence of interrupt requests. We have formulated a preliminary solution to this problem. The solution makes use of the presence of the DS in the processor. It does not call for additional components; it responds to interrupt requests promptly; and it imposes no performance penalty when no interrupt requests are present. Much work remains to be done.

We have initiated investigations on the handling of conditional branches. Test runs have verified the correctness of simulator.

With the delivery of 4 HP 370 Workstations in the next few weeks, we will be able to perform extensive simulations to evaluate various schemes for interrupt and branch handling. Both Livermore Loops and Dhrystone benchmarks will be used.

On another front, we have been investigating the relationship between processors with multiple functional units (SMFU) and systolic arrays, being investigated by Frank Luk and others. We are implementing some of the basic and important matrix operations, which have been efficiently carried out with systolic arrays, on SMFU's. It should be pointed out that SMFU's are general purpose systems, while systolic arrays are designed for specific operations.

SCIENTIFIC IMPACT OF RESEARCH

Processors with multiple functional units, such as the CRAY machines and the emerging RISC microprocessors, represent an important architecture that can be exploited for the implementation of real-time signal processing

systems. Our task addresses a vexing problem for such structures: the under-utilization of functional units.

It is expected that the Dispatch Stack that we have formulated and tested will enable such processors to issue possibly more than one instruction for each machine cycle and instructions can be issued out-of-order. These features enhance significantly the performances of such processors without raising the clock rate. And we believe that our study provides timely and much needed understanding in defining the processors that follow the recently announced Motorola 88000, Intel 80960 and other RISC processors. We have received several inquiries on the Dispatch Stack.

Furthermore, our continuing work on interrupt handling will remove yet another obstacle and make such processors dominate computer structures for real-time signal processing and other applications.

DEGREES

1. Martin F. R. Day,
Master of Engineering, May 1989.

REFERENCES

- [1] R. M. Russell, "The CRAY-1 Computer System", ACM Communications, 21, 1, January 1978, pp. 63-72.
- [2] A. E. Charlesworth, "An Approach to Scientific Array Processing: The architectural design of the AP-120B/FPS 164 family", Computer, 14, 9, September 1981, pp. 18-27.
- [3] J. E. Thornton, DESIGN OF A COMPUTER: The Control Data 6600, Scott Foresman, 1970.
- [4] R. M. Tomasulo, "An Efficient Algorithms for Exploiting Multiple Arithmetic Units", IBM Journal, vol 11, January 1967, pp. 25-33.
- [5] R. D. Acosta, J. Kjelstrup and H. C. Torng, "An Instruction Issuing Approach to Enhancing Performance in Multiple Functional Unit Processors", IEEE Trans. on Computers, C-35, 9, September 1986, pp. 815-828.
- [6] Hwu, Y. N. Patt, "Checkpoint Repair for High Performance Out-of-Order Execution Machines", IEEE Trans. Computer., vol C-36, no. 12, Dec. 1987, pp. 1496-1514, .
- [7] J. E. Smith and A. R. Pleszkun, "Implementing Precise Interrupts in Pipelined Processors", IEEE Trans. Computer., vol C-37, no. 5, pp. May 1988, 562-573, .
- [8] J. K. F. Lee and A. J. Smith, " Branch Prediction Strategies and Branch Target Buffer Design", IEEE COMPUTER, vol. 17, no. 1, January 1984, pp. 6-22, .

ISEP PUBLICATIONS

- [1] H. C. Torng, "Instruction Issuing Mechanism for Processors with Multiple Functional Units", U. S. Patent 4,807,115, February 21, 1989.
- [2] H. C. Torng, "Beyond the Present RISC -- Execution Concurrency and Branch Prediction", in preparation.
- [3] H. Dwyer and H. C. Torng, "A Fast Instruction Dispatch Unit for Multiple and Out-of-Sequence Issuances", Technical Report EE-CEG-88-09, July 1988.
- [4] H. C. Torng, "Innovations in Computer Architecture for Real-Time Signal Processing and Other Applications", Cornell Engineering Quarterly, 23, 1, pp. 18-21, Autumn 1988.
- [5] Martin Day and H. C. Torng, "A New Interrupt Handling Approach for Out-of-Order Execution Machines", in manuscript form.