Part III  Systems and Signals

Section 1  Computer-Aided Design
Section 2  Digital Signal Processing
Section 1  Computer-Aided Design

Chapter 1  Custom Integrated Circuits

Chapter 2  Computer-Integrated Design and Manufacture of Integrated Circuits
Chapter 1. Custom Integrated Circuits

Academic and Research Staff
Professor Jonathan Allen, Professor John L. Wyatt, Jr., Professor Srinivas Devadas, Professor Anantha P. Chandrakasan, Professor Jacob K. White, Professor Berthold K.P. Horn, Professor Hae-Seung Lee, Professor J. Nicholas Newman, Professor Terry P. Orlando, Professor Martin A. Schmidt, Professor Stephen D. Senturia, Professor Chales G. Sodini, Dr. Narayana R. Aluru, Dr. Ali Beskok, Dr. F. Thomas Korsmeyer, Dr. Scott B. Dynes, Dr. Ichiro Masaki, Dr. Christopher J. Terman

Visiting Scientists and Research Affiliates
Daniel W. Engels, Dr. Herre S.J. van der Zant

Graduate Students

Technical and Support Staff
Susan E. Chafe, Dorothy A. Fleischer

1.1 Custom Integrated Circuit Design

Sponsors
IBM Corporation
MIT School of Engineering

Project Staff
Professor Jonathan Allen, Dr. Scott B. Dynes, Michael S. Ehrlich, Dr. Christopher J. Terman

Since last year, considerable progress has been made toward the implementation of an interactive learning environment for VLSI design. The major goal of this research is to design a computer-based learning environment that includes a text with active links to other programs; and to couple this text to a computer-aided design environment that will allow interactive exploration of circuit behavior and the design of substantial MOS integrated circuits. The text is now in hypertext form, and links to the computer-aided design environment are being developed.

The core of the computer-aided design capability is a SPICE-like simulator in which the essential circuit simulation capability has been preserved without retaining other SPICE features not needed in this environment. MOS device models have been added to this simulator, enabling it to broadly simulate all styles of MOS digital circuits. In particular, the use of a SPICE-like simulator avoids convergence difficulties found with some simulators, in connection with pass-gate circuits and other structures that can provide difficulties in look-up table simulators. For each figure in the text, a corresponding SPICE input deck has been developed. The intent is for students to write and read these listings. In this way, it is easy for a student to modify any aspect of a circuit, such as width-to-length ratios of transistors. In the future, the plan is to provide a schematic editor to graphically edit such circuits. However, for the present, a schematic viewer based on the SPICE deck is provided.

For integrated-circuit layout, a reduced version of the MAGIC layout program is utilized. In addition, an extraction program is provided so users can convert the layout to a net list appropriate for SPICE simulation. Basic plotting capability is provided, as well as waveform arithmetic, which is exceedingly useful in understanding the behavior of circuits.

The computer-aided design tools described above enable design entry, as well as simulation, at both the circuit and layout levels, based on these representations. This will be the core of the initial set of computer-aided design resources provided in the interactive learning environment. Earlier, substantial effort was made to utilize commercial computer-aided design tools, but the project is now focused on the utilization of modified public domain software and the creation of new software. The entire interactive learning environment will be capable of being used not only in a stand-alone manner on either a workstation, PC, or laptop computer, but it will also
Chapter 1. Custom Integrated Circuits

be capable of being used as a network resource utilizing JAVA.

It is anticipated that the first version of the interactive learning environment will be used in MIT subject 6.371 (Introduction to VLSI Systems) in the fall of 1997. Now that the project is self-contained and no longer dependent on arrangements with outside vendors, it has progressed rapidly with the project team in control of all aspects of the software.

1.2 Cost-Effective Hybrid Vision Systems for Intelligent Highway Applications

Sponsor
National Science Foundation
Grant MIP 94-23221

Project Staff
Professor Berthold K.P. Horn, Professor Hae-Seung Lee, Dr. Ichiro Masaki, Professor Martin A. Schmidt, Professor Charles G. Sodini, Professor John L. Wyatt, Jr., Jason R. Bergendahl, Steven J. Decker, Stanislav E. Frumkin, Jeffrey C. Gealow, David A. Martin, Chig-Chun Wang

The cost of machine vision systems is limiting their applications. The goal of this project is to develop new cost-effective architectures for vision systems and to evaluate them for intelligent highway applications. We propose an advanced modular architecture as a way of improving the cost and performance of vision systems. We call the architecture the heterogeneous nanocomputer network. It is heterogeneous because the network consists of a variety of functional modules. Some modules are application-specific integrated circuits (ASIC) chips for straightforward early-vision processing while the others are highly programmable. The modules can be either analog, digital or mixed signal. The heterogeneity provides a significant opportunity to lower the cost of each module by tailoring its architecture to a particular function with minimum constraints. Each functional module is called a nanocomputer because its computational silicon area is significantly smaller than a conventional microcomputer. Reducing module size is important to reducing total system cost. A small-grain architecture also provides greater flexibility for integrating modules. A network architecture is essential for integrating multiple modules without communication bottlenecks.

1.2.1 Brightness Adaptive CMOS Imager with Digital Output

Graduate student Steven J. Decker, working with Professor Charles Sodini, has been designing and testing an imager to act as the front end of a modular vision system. The imager consists of a 256 X 256 wide-dynamic-range pixel array, 256 correlated double-sampling circuits, 128 cyclic analog-to-digital converters, and test structures. The chip was fabricated in a 0.8 $\mu$m CMOS process. Die area is approximately 10.4 mm X 8.3 mm. Digital output is column-parallel, with a maximum design speed of 1000 frames/second at 370 mW of static power dissipation.

Each pixel consists of a photodiode, a sense capacitor, a lateral overflow gate, a charge spill gate, a row select device, and a source follower. Over the integration period, photodiode current passes through the charge spill gate into the sense capacitor. At the end of the integration period, the row select device is turned on, connecting the source follower to a bias current source. The output (source) voltage follows the voltage on the sense capacitor, less a threshold voltage. The pixel is then reset through the lateral overflow gate.

The pixel is designed for high sensitivity and operation over a wide dynamic range. Sensitivity is enhanced by having a large fill factor (47 percent). Sensitivity is also improved by using a charge spill gate, which allows the photocharge collected over a large area to be sensed using a small capacitance. The conversion gain is approximately 12 $\mu$V/e, and the dark current is less than 10 fA.

Wide dynamic range operation is achieved by applying a time-varying voltage to the lateral overflow gate. For very low illumination, all of the photocurrent is collected in the photodiode, and the effective integration time is identical to the integration period. As illumination increases, the effective integration time decreases, decreasing the sensitivity of the pixel. The compression characteristic is user-adjustable. No additional pixel area is required.

The lateral overflow gate waveform is generated by a 256-stage, 8b wide digital shift register, eight dc analog voltage lines, and 256 DACs. Each row in the imaging array has a common lateral overflow gate voltage, generated by one stage of the shift register and one DAC. The user inputs digital words at the start of the shift register, each of which contains exactly one high bit. At each row, the DAC connects the lateral overflow gates to the analog voltage selected by the high bit. As the shift register is clocked forward, a stepped voltage waveform is generated at each row, and the waveforms are
staggered between rows so that the pixels can be read out using a progressive scan technique.

Correlated double-sampling (CDS) is used to reduce the fixed-pattern noise. Two measurements of the pixel output voltage are taken, one at the end of the integration period, and one after the pixel is reset. By subtracting the second sample from the first sample, offset errors are removed to first order. In particular, the effect of threshold voltage offsets in the source follower and lateral overflow devices are removed. Fixed-pattern noise caused by geometry mismatch in the source follower device and current source bias mismatch is also removed. The CDS circuit also converts the signal from single-ended to fully differential for the ADC.

Analog-to-digital conversion is performed by a switched-capacitor cyclic converter. The algorithm can be operated at varying levels of accuracy, depending on the number of conversion cycles performed, up to the inherent accuracy of the converter. Ultimate accuracy is limited by capacitor matching, which is expected to be in the 8b to 10b range. The converter uses a four-phase clock, and produces bits on the odd phases. For 8b output, the conversion time is 1.6 µs.

A surface-mount printed circuit board has been designed and fabricated to test the chip. The operation of all individual test cells (pixel, CDS circuit, and ADC) has been verified. Work is proceeding on downloading complete image frames.

1.2.2 Silicon IC Process Compatible Uncooled Infrared Imager Design

Infrared imagers are widely adopted in satellite remote sensing, night vision systems and military target acquisition because infrared has a high transmission coefficient in air and objects emit strong infrared radiation around room temperature. Compared to a typical imager design, an uncooled one is favored since an expensive cooling system for reducing thermal noise is avoided. Moreover, the imager has the potential to be low cost if the fabrication process is fully silicon IC process compatible.

The uncooled imager in our study utilizes a thermal detector structure. Through sealed-cavity wafer bonding technology, this structure incorporates a thermally isolated silicon nitride plate and a single crystal Si-bolometer transducer above it. The membrane, suspended in air and connected to the substrate by several thin tethers, provides mechanical support for the Si-bolometer and acts as an absorption layer since nitride has an specific infrared absorption band. The Si-bolometer is constructed by patterning bonded silicon wafers and is connected electrically to outside circuits by titanium wires. Titanium serves as the contacting metal in this structure due to its lower thermal conductivity compared to aluminum. In our design, all the necessary materials are currently IC foundry compatible. From a process point of view, the extra work other than typical IC fabrication procedures are the wafer bonding pre-processing and bolometer forming post-processing.

The uncooled imager works as follows: first, infrared photons are absorbed by absorption layers on the pixels, which correspond to the nitride membranes. The photon energy is converted into heat. Consequently, the temperature of the pixels rise above that of the substrate due to the thermally-isolated structure. The silicon bolometer, a temperature sensitive resistor, will convert this temperature rise into voltage signal output under constant current bias. As the infrared image is projected and focused on the imager chip, the signal on each pixel can be picked up sequentially by utilizing a x-y multiplexer on chip. In order to achieve high signal output and reasonable frequency response, the heat capacity and thermal conductance for each pixel should be made as small as possible. This means the size of the pixel should be minimized and the tethers should be long, thin, and narrow. Minimization of the pixel size increases imager resolution. However, photolithography technology poses a limit on the pixel size. Besides that, the membranes are suspended above the substrate by 10 microns to relieve the impact of thermal conductance caused by air.

This research will be separated into two main phases. In the first phase, the preliminary theoretical calculation will be performed by assuming a first order lumped thermal model and the optimal pixel structure will be determined. Furthermore, optical and electrical parameters will be extracted to predict the imager performance. The integrity of the nitride membrane through high temperature process and metallization with the required step coverage will also be investigated. In the second phase, the focus will be on the design and fabrication of the imager chip which incorporates signal processing circuits. System testing will be completed and the prototype infrared imager will be demonstrated.

1.2.3 Real-time Three-dimensional Vision System

Graduate student Jason R. Bergendahl, under the supervision of Professor Berthold Horn and Dr. Ichiro Masaki, is developing a real-time machine vision system for stereo distance measurements in automotive applications. The basic approach differs
substantially from those used by other stereo vision systems under development. First, the stereo distance measurements rely on edge correlation between stereo pairs, rather than area correlation. The algorithm is greatly simplified to allow real-time performance at frame rate (30 frames/second) to be realized. Another important difference is that the distance measurements are based on sub-pixel disparity measurements between the stereo pairs, allowing for a compact camera arrangement that does not preclude accurate distance measurement.

In the first phase of the project, completed in December 1995, the proposed stereo vision algorithm was implemented on a workstation. This effort confirmed the functionality of the basic algorithm, and contributed to a better understanding of the capabilities of a PC for performing image processing without the aid of additional hardware. A standard Pentium-based desktop PC with PCI bus was used. To meet our real-time goal, a hardware system capable of processing images in real-time was added to the PC. We use a PCI plug-in board from Imaging Technology, Inc. (ITI) as our hardware image processing platform.

The completed system has three major components: the camera assembly, the PC-based image processing hardware, and the PC itself. The PC is present to both serve as host to the image processing hardware and to perform the high level functions of the algorithm. The camera assembly consists of three miniature CCD cameras, which are equally-spaced with optically axes aligned, and secured to a metal baseplate. In our current experimental setup, the left and right cameras are 25 cm apart, with the center camera located midway between them. A small LCD display for viewing the results of the algorithm on-road was also added. The complete system has been installed in a test vehicle for on-road trials. Power for the system is obtained exclusively from the electrical supply of the test vehicle (through the cigarette lighter).

The software has undergone a number of revisions since the first implementation on the workstation. The software now controls the simultaneous image acquisition from the three cameras, and the image processing steps performed both in the image processing hardware and the PC. Our basic approach is to recover the depth to features in the 3-D environment using a stereo triple acquired from the three cameras. We find point correspondences among features in the images, and compute depth from disparity. The specific steps in our algorithm are: edge detection, localizing edges, finding edge correspondences, localizing edges to subpixel accuracy, and generating the depth map. The depth map is then used to locate vehicles and lane markings on the roadway.

On-road trials have been conducted on a local highway to evaluate our approach and algorithm under a number of different driving conditions. Although the system has not yet been tested against ground truth data, the results are qualitatively reasonable. The current effort underway aims to reliably separate the vehicles on the roadway from the lane markings. Once successfully accomplished, this will greatly simplify the task of vehicle identification and lane tracking.

1.2.4 Time-to-Collision Chip

The purpose of this project is to compute the time it takes an object to hit a camera in real time. To date, the least squares based time to collision algorithm has been verified and the IC implementing a simplified version has been designed. The remaining work includes testing of the chip logic design and its layout. The chip will then be manufactured though MOSIS IC fabrication service.

The major application of the algorithm as seen at this time is a warning system that makes one object aware of a potential collision with another object in its vicinity. One example of this would be a system consisting of an imager and a chip implementing the algorithm placed on a car. The system (imager) would be directed in one of the driver's blind spots and provide an early warning about approaching vehicles in terms of time to impact.

The algorithm is based on the following equations, assumptions and mathematical techniques. The brightness constraint equation is used in conjunction with the rigid body constant velocity motion. The brightness constraint equation provides an equation constraining the values of spatial and temporal derivatives, object normal and velocity vectors for each pixel in the image. Since pixel brightness is the input from the imager, precision is not well controlled. To make an accurate estimate of object normal and velocity (therefore time to collision), a least squares mathematical technique is employed. The brightness constraint equation is formulated in terms of spatial and temporal derivatives which are approximated using an appropriate derivative kernel. The error introduced in this process is another reason for the use of the least squares approach.

As the first phase of the project, the algorithm has been coded in C and evaluated based on synthetic (generated in the computer) and real data taken in the lab under controlled conditions. The algorithm simulations in both settings have shown that the algorithm estimates the time to impact very accurately.
Chapter 1. Custom Integrated Circuits

The main version of the algorithm (based on a 2-D array of pixels) has the advantage that the camera does not require calibration. That is X and Y coordinates of a pixel can be specified to within a constant without affecting the output. However, the algorithm is still rather complex and requires significant hardware resources for real time implementation.

The simplified version of the algorithm takes advantage of the observation that the motion of a car can be described by a 2-D motion vector (in the plane of the road) rather than 3-D vector. This allows us to use only one line of data with Y=0 to compute the time to collision. Therefore, the simplified algorithm is based on a 1-D array of pixels. The computational structure of the algorithm becomes very symmetrical and is simplified considerably. These improvements make the algorithm ready for the real time hardware implementation.

The simplified algorithm has been programmed and verified with good results. The arithmetic operations (add, multiply, divide) have been coded using finite arithmetic. The precision limitations have been investigated. The only drawback of this simplified algorithm is that it requires camera calibration. (The line of data with Y=0 is required.)

The IC implementation of the algorithm is based on the fact, that the core of the algorithm computation adds and multiplies with real numbers. Due to the large dynamic range exhibited by the spacial and temporal derivatives, the floating point implementation is used for the arithmetic. The simulations with the data available have shown that 12 bits for the mantissa and 6 bits for exponent produce adequate results. However, to be conservative the mantissa is chosen to be 16 bits.

The IC design represents programmable floating point units (floating point multiplier and adder operating on 16b/6b floating point data). The relatively small amount of computation allows us to multiplex all the arithmetic operations onto a single floating point unit. The sequence of the arithmetic operations is determined by a program that is downloaded on the chip at the initialization time. At present, the major part of the logic design is complete.

Some minor modifications to the computational core as well as the initialization logic are yet to be designed. After a comprehensive simulation of the chip logic, the chip will be layed out and manufactured through the MOSIS fabrication service.

1.2.5 Integrated Computing Structure for Pixel-Parallel Image Processing

Graduate student Jeffrey C. Gealow, working with Professor Charles G. Sodini, has developed a pixel-parallel image processor chip. Typical low-level image processing tasks are performed by applying a uniform set of operations for each pixel in each input image. Thus, they may be efficiently handled by an array of processing elements, one per pixel, sharing instructions issued by a single controller. Using logic pitch-matched to DRAM cells, a single chip provides a 64 X 64 processing element array suitable for real-time applications.

A processing element combines a 128-bit DRAM column with one-bit-wide logic. Processing elements are interconnected to create a two-dimensional rectangular array matching the structure of image data. With the interconnection network extended across chip boundaries, multiple chips may be used to form processing element arrays matching the size of large images.

The chip was fabricated in a 3.3 V ASIC process available through the MOSIS Service. The 80 square mm-chip has eight blocks of 512 processing elements. At the center of each block is a twin cell DRAM array with 128 rows and 512 columns. The layout pitch of processing element logic is double the memory column pitch. Arrays of 256 logic units are placed above and below the DRAM array. There are no column decoders—logic circuits are connected directly to the bitlines. The pitch-matched processor implementation maximizes the bandwidth between memory and logic and minimizes processing element area.

Fabricated chips are fully functional. Operating with a 60 ns clock cycle, the chips dissipate 300 mW. A demonstration system employs four chips to form a 128 X 128 processing element array. Several low-level image processing tasks have been implemented: smoothing and segmentation, median filtering, edge detection, and optical flow computation. All have been successfully performed in real time with input images provided at standard video frame rate.

1.2.6 Mixed-Signal Array Processor for Multi-Instruction Multi-Data Operations

Graduate student David A. Martin, working with Professor Hae-Seung Lee and Dr. Ichiro Masaki, has developed a mixed-signal array processor chip. Many low level vision algorithms require only low to moderate precision (6 to 8 bits). For these applications, a special purpose analog circuit is often smaller, faster, and lower power solution than a
general purpose digital processors. However, because these analog chips are only suitable for one very specific function, they are often expensive, low volume products. This sub-project presents a more general purpose programmable mixed-signal array processor that combines the flexibility of digital processor with the small area and low power of an analog circuit. It achieves a processing efficiency in terms of power and area superior to that of comparable digital processor. Each processor in the array has a digital control unit, an analog switch fabric, an analog storage unit, and an analog arithmetic unit with an accuracy of 7 bits. The analog arithmetic unit utilizes a unique circuit that combines a cyclic switched capacitor A/D and D/A to perform addition, subtraction, multiplication, and division. Each processor cell performs, in parallel with all of the other processor cells, 0.8 million operations per second, consumes 1.825 mW of power, and uses 700 μm by 270 μm of silicon area. The chip was fabricated in Hewlett Packard's 0.8 μm triple metal CMOS process. The multi-instruction, multidata operation scheme of the chip made it possible to implement image convolution operations with a relatively small number of processing cells. An array of these processors was used to successfully perform an edge detection algorithm, and a sub-pixel resolution algorithm executed on the array increased the resolution of the edge locations by a factor of four.

1.3 Computer-Aided Design Techniques for Embedded System Design

Sponsor
Defense Advanced Research Projects Agency/
U.S. Army Intelligence Center
Contract DABT63-94-C-0053

Project Staff
Professor Srinivas Devadas, Daniel W. Engels,
George I. Hadjiiannis, Silvina Z. Hanono

1.3.1 Embedded Systems

The high level of integration afforded by advances in processing technology has brought new challenges to the design of digital systems. Designing an entire system with a custom integrated circuit is now neither economical nor practical. As time-to-market requirements a place greater burden on designers for fast design cycles, programmable components are being introduced into the system and an increasing amount of system functionality is implemented in software relative to hardware. These programmable components, called embedded processors or embedded controllers, can be general-purpose microprocessors, off-the-shelf digital signal processors (DSPs), in-house application-specific instruction-set processors (ASIPs), or microcontrollers. Systems containing programmable processors that are employed for applications other than general-purpose computing are called embedded systems.

The advantages of incorporating software components are twofold: (1) faster turnaround time due to the use of predefined processors, and (2) increased flexibility due to field programmability. Moreover, various applications of the same genre may share the same hardware structure, with differences reflected in the software component, and it is possible to use the same mask for all applications. Since the aggregate volume of a number of applications is higher than the individual volumes, using software components substantially reduces the cost of manufacturing.

1.3.2 Hardware-Software Co-Design

As a result of the advantages of software, modern integrated systems are often composed of a heterogeneous mixture of hardware and software components. A system may consist of a digital signal processor (DSP) core, a program ROM, RAM, application-specific circuitry (ASIC), and other interface and peripheral circuits. Because of the trend towards this design style, developers of computer-aided design (CAD) tools are faced with the challenge of providing circuit designers with tools that can support the design of such systems.

In the design flow, the designer first determines which parts of the functionality of the system will be implemented in hardware and which parts in software. The designer then proceeds to design each of the hardware and software components. The system is simulated and evaluated with a hardware-software co-simulator. If the results of the simulation meet design specifications (e.g., correctness and timing constraints), then the design is accepted. Otherwise, the designer may re-partition the original algorithmic specification and reiterate the same process. Under this methodology, tools for code generation and hardware-software co-simulation have become essential parts of the designer's tool-box.

The methodology required for code generation will be most useful if it can be easily adapted to generating code for different processors. This property is commonly called retargetability. The methodology will require, in addition to traditional code optimizations, new techniques specifically for the classes of
processors that are commonly found in embedded systems: off-the-shelf fixed-point DSP architectures and application-specific instruction processors used in DSP applications.

1.3.3 Retargetable Compilation

The term retargetable compilation has been widely used and describes a broad range of capabilities. We divide the spectrum into three categories and briefly describe the extent to which each of these categories applies.

Automatically retargetable. The ideal automatically retargetable compiler takes in a description of the target architecture in its structural form and generates code for the target. By structural form we mean a form that is suitable also for the synthesis of hardware. Thus, the code generation process requires no user intervention. In practice, however, this level of retargetability seems to be applicable to a limited family of architectures within which variations are well characterized, e.g., number of registers in a register file, bit-widths, and available functions of an execution unit.

User retargetable. Here, the user specifies the target architecture to a compiler generator by describing the instruction set and the actions of each instruction. Such a description is usually called behavioral. The compiler generator takes the description as input, and outputs a compiler for the target architecture. This method has been used with some success for the code selection phase of existing compilers.

Developer retargetable. This level of retargetability is sometimes simply termed portability. One way to handle machine specific optimizations that go beyond instruction selection is to permit the developer to modify the compiler to target the given architecture. Clearly the dividing line between retargeting and essentially writing a new compiler for each architecture is rather thin.

1.3.4 Retargetability and a Machine-Description Language

Given an application or a set of applications, a system designer is required to develop a high-performance system that runs the applications. Once the partitioning between hardware and software has been determined, the tasks of hardware synthesis and software synthesis remain.

The target architecture model has to be determined in the software synthesis trajectory. The designer may opt to design or redesign an application-specific instruction processor (ASIP) to run the software. The design then needs to specify the functionality of the ASIP which involves deciding the hardware structure of the ASIP. In order to determine how fast the application runs on the given hardware specification of the ASIP, one has to compile the application program to ASIP code. Hand-coding is not an option since it eliminates the capability of iterating over the hardware specification of the ASIP optimizing for code speed and/or size. We argue that in order to be able to explore the design space, an efficient and versatile retargetable compilation strategy is required.

A retargetable compiler receives as input the program corresponding to the application as well as the machine description of the ASIP. The machine description includes a microarchitecture specification and an instruction set specification. The code generator produces code that can run on the ASIP optimized for speed and size. By varying the machine description and evaluating the resulting object code, one can effectively explore the design space of both hardware and software components. The ASIP design flow scenario we have developed is illustrated in figure 1.

The machine description language ISDL is a driver for various templates which correspond to parameterizable and scalable datapaths. Our machine description has five different components:

1. Instruction set: number of fields and field formats,
2. Operations available in each unit,
3. Registers and topology,
4. Conflicts (in the form of rules) between operations that cannot be executed simultaneously or have to be a certain number of clock cycles apart, and
5. Hints for optimization (optional).

To make retargetable compilation easier, we focus on restricted addressing modes, uniform interfaces to each functional unit, and assume source operand
commutativity in instructions\(^1\) and that complications such as interlock handling are done in hardware.

Traditional compiler optimizations such as scheduling, register allocation, loop unrolling and address assignment can be used in the retargetable compiler and are driven by the parameters provided by the machine description. The key step in the preliminary which produces the SUIF intermediate form, is the intermediate form of the Stanford SUIF compiler. The preliminary code generated is produced from a rule-based machine description written in OLIVE, a tree-based pattern matcher. SUIF then serves as an input to the instruction selector/scheduler. Depending on the number and type of available functional units, matches are generated on SUIF operations, where each match corresponds to a legal combination of SUIF operations that can be executed in parallel on the given machine. Optimal instruction selection/scheduling involves choosing the appropriate subset of matches, minimizing the cost of instructions; we have investigated the use of binate covering to solve this problem.

Handling data transfers is the next major step. The various operations in the instructions require data to be in prescribed locations. Register allocation is first performed where it is decided which values will reside in memory and which will reside in each register file. We then generate the data transfers required to implement the chosen instructions. Data transfers include register-to-register, memory-to-register, and register-to-memory transfers. Next, the data transfers are optimally scheduled in parallel with the instructions without violating dependence constraints; we will investigate the use of compaction algorithms to solve this problem. This step is parameterized by the number of available address generators, the ports available at each register file, and the topology of connections between register files, memory and the functional units.

The optimizations of storage assignment and code compression are performed on the generated code, if the optimizations are enabled in the target architecture. Peephole optimization to exploit irregular features of the target architecture is the final step.

\(^1\) We do not assume that each operation is commutative, merely that each operand can come from the same set of sources, e.g., register file A, memory, etc.
1.4 Functional Verification of VLSI Systems

Sponsors
Mitsubishi Corporation
National Science Foundation
Young Investigator Award
Fellowship MIP 92-58376

Project Staff
Professor Srinivas Devadas, Farzan Fallah

Functional simulation is the most widely used method for design verification. At various levels of abstraction, e.g., behavioral, register-transfer level and gate level, the designer simulates the design using a large number of vectors attempting to debug and verify the design. A major problem with functional simulation is the lack of good metrics or tools to evaluate the quality of a set of functional vectors. Metrics used currently are based on instruction counts and are quite simplistic. Designers are forced to use ad-hoc methods to terminate functional simulation, e.g., CPU time limitations.

We have proposed a new metric for measuring the extent of design verification provided by a set of functional simulation vectors. This metric is universal and can be used uniformly for all designs. Our metric computes controllability information corresponding to instructions being executed, as well as observability information corresponding to whether effects of errors at inputs to modules can be observed at module outputs.

The proposed metric is inspired by notions of fault coverage used in manufacturing test and notions of coverage certification used in software testing. In order to compute observability information, we tag variable values in behavioral or register-transfer level simulation and use a simulation calculus to efficiently calculate the coverage provided by an arbitrary set of functional vectors. Tags on variable values are not tied to particular design errors; they serve as a mechanism for extending standard coverage metrics to include observability requirements.

Our initial experiments support the validity of the proposed metric. Using this metric in design verification can result in higher-quality functional tests and improved correctness checking.

1.4.1 Publications

Journal Articles


Conference Proceedings


Continuing our efforts in the time-domain analysis of ship and body motions, we released TIMIT 2.1 to approximately 14 organizations including the David Taylor Research Center, major international oil companies and contractors, and classification societies. While current research (including our own) is advancing into quasi-linear and completely non-linear analyses of ship motion in waves, efficient, primarily linearized, computational tools are needed. We have also completed or nearly completed three extensions of the theory and implementation: finite depth analysis, generalized modal analysis, and complete second-order mean force analysis.2 In the area of accelerated algorithms for potential flow problems, the release of FastLap 2.0 probably marks the furthest advancement to which we will take the fast-multipole approach. Version 2.0 extends multipole acceleration to new integral formulations such as the so-called source formulation and the desingularized approach. The latter requires a generalization of the algorithm to include separate upward and downward pass trees. This feature also allows the computation of quantities in the field, such as the gradient of the potential.3 Also, we have extended the precorrected-FFT algorithm to problems in ocean engineering, particularly those with oscillatory kernels.4

1.5 Offshore Structure Analysis - WAMIT Development

Sponsor
Joint Industry Program on Offshore Structure Analysis

Project Staff
Dr. Ali Beskok, Dr. F. Thomas Korsmeyer, Chang Ho Lee, Professor J. Nicholas Newman, Professor Jacob K. White

In order to reduce the number of prototype failures, designers of these devices need to make frequent use of simulation tools. To efficiently predict the performance of microelectromechanical systems, these simulation tools need to account for the interaction between electrical, mechanical, and fluidic forces. Simulating this coupled problem is made more difficult by the fact that most MEMS devices are innately three-dimensional and geometrically complicated. It is possible to simulate efficiently these devices using domain-specific solvers, provided the coupling between domains can be handled effectively. In this work we have developed several new approaches and tools for efficient computer aided design and analysis of MEMS.5

References


1.7 Simulation Algorithms for RF Circuits

**Sponsors**
Cadence Design Systems  
MAFET Consortium

**Project Staff**
Ogden J. Nastov, Professor Jacob K. White

Designers of RF integrated circuits make extensive use of simulation tools which perform nonlinear periodic steady-state analysis and its extensions. However, the computational cost of these simulation tools have restricted users from examining the detailed behavior of complete RF subsystems. Recent algorithmic developments, based on matrix-implicit iterative methods, are rapidly changing this situation and providing new faster tools which can easily analyze circuits with hundreds of devices. We have investigated these new methods by describing how they can be used to accelerate finite-difference, shooting-Newton, and harmonic-balance based algorithms for periodic steady-state analysis.6

We developed a preconditioned recycled Krylov-subspace method to accelerate a recently developed approach for ac and noise analysis of linear periodically-varying communication circuits. Examples show that the combined method can be used to analyze switching filter frequency response, mixer noise frequency translation, and amplifier intermodulation distortion. In addition, it can be shown that for large circuits the preconditioned recycled Krylov-subspace method is up to 40 times faster than the standard optimized direct methods.7

1.8 Numerical Techniques for Simulating Josephson Junction Arrays

**Sponsors**
Consortium for Superconducting Electronics  
National Defense Science and Engineering Graduate Fellowship

**Project Staff**
Joel R. Phillips, Dr. Herre S.J. van der Zant, Professor Terry P. Orlando, Professor Jacob K. White

Vortices play a central role in determining the static and dynamic properties of two-dimensional (2-D) superconductors. Artificially fabricated networks of superconducting islands weakly coupled by Josephson junctions are model systems to study the behavior of vortices. Through simulation, we have discovered that the static properties of vortices in an array of Josephson junctions can be significantly influenced by magnetic fields induced by the vortex currents. The energy barrier for vortex motion is enhanced, nearly doubling for penetration depths on the order of a cell size. Moreover, we have found that correct calculation of the vortex current distribution, the magnetic moment, and the lower critical field require modeling mutual inductance interactions between all cell pairs in the array. To make numerical simulation of the system with all inductive effects computationally feasible, a novel FFT-accelerated integral equation solver was derived. This algorithm is sufficiently efficient to allow Shapiro steps and the dynamics of row-switched states in large (500X500 cells) arrays.8

1.9 Efficient 3-D Interconnect Analysis

**Sponsors**
Consortium for Superconducting Electronics  
Defense Advanced Research Projects Agency  
Digital Equipment Corporation  
IBM Corporation  
MIT Lincoln Laboratory  
National Defense Science and Engineering

---


Graduate Fellowship  
Semiconductor Research Corporation

**Project Staff**

Michael T. Chou, Mattan Kamon, Yehia M. Massoud, Joel R. Phillips, Johannes Tausch, Professor Jacob K. White

We have developed multipole-accelerated algorithms for computing capacitances and inductances of complicated 3-D geometries, and have implemented these algorithms in the programs FASTCAP and FASTHENRY. The methods are accelerations of the boundary-element or method-of-moments techniques for solving the integral equations associated with the multiconductor capacitance or inductance extraction problem. Boundary-element methods become slow when a large number of elements are used because they lead to dense matrix problems which are typically solved with some form of Gaussian elimination. This implies that the computation grows as $n^3$, where $n$ is the number of panels or tiles needed to accurately discretize the conductor surface charges. Our new algorithms, which use Krylov subspace iterative algorithms with a multipole approximation to compute the iterates, reduces the complexity. Accurate multiconductor capacitance and inductance calculations grow nearly as $nm$ where $m$ is the number of conductors. For practical problems which require as many as 10,000 panels or filaments, FASTCAP and FASTHENRY are more than two orders of magnitude faster than standard boundary-element based programs. Manuals and source code for FASTCAP and FASTHENRY are available directly from MIT.

In more recent work, we have been developing an alternative to the fast-multipole approach to potential calculation. The new approach uses an approximate representation of charge density by point charges lying on a uniform grid instead of by multipole expansions. For engineering accuracies, the grid-charge representation has been shown to be a more efficient charge representation than the multipole expansions. Numerical experiments on a variety of engineering examples arising indicate that algorithms based on the resulting "precorrected-FFT" method are comparable in computational efficiency to multipole-accelerated iterative schemes, and superior in terms of memory utilization.

The precorrected-FFT method has another significant advantage over the multipole-based schemes: because it can be easily generalized to some other common kernels. Preliminary results indicate that the precorrected-FFT method can easily incorporate kernels arising from the problem of capacitance extraction in layered media. More importantly, problems with a Helmholtz equation kernel have been solved at moderate frequencies with only a modest increase in computational resources over the zero-frequency case. An algorithm based on the precorrected-FFT method which efficiently solves the Helmholtz equation could form the basis for a rapid yet accurate full-wave electromagnetic analysis tool.

Our latest work has been on extending the applicability of precorrected-FFT and fast multipole schemes. We have combined these fast techniques with new numerically stable integral formulations for problems in electroquasistatics (distributed RC) and for problems with high dielectric permittivity ratios. In addition, we have extended the algorithms in FASTHENRY to include finite substrate conductivity and used it to analyze on-chip inductance. We have also developed parallel versions of the precorrected-FFT algorithm, as

---


well as preconditioners which improve the convergence of the iterative solver used in FASTHENRY.\(^\text{13}\)

### 1.10 Model Order Reduction for Interconnect Analysis

**Sponsors**
Defense Advanced Research Projects Agency  
Semiconductor Research Corporation

**Project Staff**
Ibrahim M. Elfadel, Mattan Kamon, Joel R. Phillips,  
Michael T. Chou, Professor Jacob K. White

Reduced-order modeling techniques are now commonly used to efficiently simulate circuits combined with interconnect. Generating reduced-order models from realistic 3-D structures, however, has received less attention. Recently we have been studying an accurate approach to using the iterative method in the 3-D magnetoquasistatic analysis program FASTHENRY to compute reduced-order models of frequency-dependent inductance matrices associated with complicated 3-D structures. This method, based on a Krylov-subspace technique, namely the Arnoldi iteration, reformulates the system of linear ODEs resulting from the FASTHENRY equation into a state-space form and directly produces a reduced-order model in state-space form. The key advantage of this method is that it is no more expensive than computing the inductance matrix at a single frequency. The method compares well with the standard Padé approaches and may present some advantages because in the Arnoldi-based algorithm, each set of iterations produces an entire column of the inductance matrix rather than a single entry, and if matrix-vector product costs dominate then the Arnoldi-based algorithm produces a better approximation for a given amount of work. Finally, we have shown that the Arnoldi method generates guaranteed stable reduced order models, even for RLC problems.\(^\text{14}\)

### 1.11 Adaptive Techniques for Fast Integral Equation Solvers

**Sponsors**
Defense Advanced Research Projects Agency  
Multiuniversity Research Initiative  
National Science Foundation  
Semiconductor Research Corporation

**Project Staff**
Jing-Rebecca Li, Johannes Tausch, Dr. F. Thomas Korsmeyer, Professor Jacob K. White

Finding computationally efficient numerical techniques for simulation of three-dimensional structures has been an important research topic in almost every engineering domain. Surprisingly, the most numerically intractable problem across these various disciplines can be reduced to the problem of solving a three-dimensional Laplace problem. Such problems are often referred to as potential problems, and application examples include electrostatic analysis of sensors and actuators, electro- and magneto-quasistatic analysis of integrated circuit interconnect and packaging, and potential flow based analysis of wave-ocean structure interaction.

The development of extremely fast-multipole- and precorrected-FFT-accelerated iterative algorithms solving potential problems has made it possible to analyze far more complicated structures than ever before. This has led to a new problem. Since the structures being analyzed are so complicated, they can not be discretized by hand nor can the algorithms be tuned by hand. Instead, algorithms must be developed which are completely automatic, and insure accurate results. In order to develop such codes, we are investigating alternative formulations.

---


and mesh refinement strategies for integral equations.\textsuperscript{15}

1.11.1 Publications


