Chapter 4. Custom Integrated Circuits

Academic and Research Staff

Professor Jonathan Allen, Professor Srinivas Devadas, Professor Bruce Musicus, Professor Jacob White, Professor John L. Wyatt, Jr., Professor Berthold Horn, Professor Hae-Seung Lee, Professor Tomaso Poggio, Professor Charles Sodini, Dr. Meir Feder

Visiting Scientist

Dr. Ehud Weinstein[1]

Graduate Students

Robert C. Armstrong, Donald G. Baltus, Cyrus S. Bamji, Ibrahim M. Elfadel, Dennis C.Y. Fogg, Mikko Hakkarainen, Craig Keast, Jennifer A. Lloyd, Andrew Lumsdaine, Lynne M. McCormick, Steven P. McCormick, Ignacio McQuirk, Hiroshi Miyanaga, Keith Nabors, James J. Olsen, Kevin Peterson, G.N. Srinivasa Prasanna, Mark Reichelt, Luis M. Silviera, William S. Song, David L. Standley, Ricardo Telichevsky, Christopher B. Umminger, Filip Van Aelten, Woodward Young

Undergraduate Student

Bennet Sikes

4.1 Custom Integrated Circuits

Sponsors

Analog Devices, Inc.
International Business Machines, Inc.
Joint Services Electronics Program
    (Contracts DAAL03-86-K-0002 and
    DAAL03-89-C-0001)
U.S. Air Force - Office of Scientific Research
    (Grant AFOSR 86-0164)

Project Staff

Professor Jonathan Allen, Robert C. Armstrong, Donald G. Baltus, Cyrus S. Bamji, Lynne M. McCormick, Steven P. McCormick, Hiroshi Miyanaga, Mark Reichelt, Filip Van Aelten

The overall goal of VLSI CAD research is to provide the means to produce custom integrated circuits correctly, quickly, and economically. In the past, correctness was applied at several different levels, such as layout (using design rule checking), and via the use of both circuit and logic simulation at higher levels. For a given functional specification, however, there are many possible realizations at lower levels of abstraction (e.g., architecture, logic, circuit, and layout). So, there is an increasing need to develop CAD tools that can enhance the performance of a particular design that is expressed in terms of speed, circuit area, and power. In this research group, the main focus is on the development of CAD tools for performance-directed synthesis, with particular emphasis on circuit representation as well as optimized architectures for digital signal processing applications. Central to this research direction is the specification of formal design representations at several levels of abstraction that are coupled to design exploration techniques which optimize the performance of the total design. These complete designs, however, require a unified approach to optimization. Traditionally, optimization techniques are applied within a single level of abstraction, but total system optimization implies the simultaneous specification of all

levels of representation, such that the desired performance goal is realized. It is not uncommon for optimizations at adjacent levels of representation to conflict (e.g., space-time trade-offs), so that isolated optimization techniques, which focus on a single level of representation, are not adequate. Furthermore, these representations must be coordinated so that each represents a consistent projection of a single overall design. This requirement guarantees that the distinct levels of abstraction can all be regarded as views of the same single abstract underlying design object.

From this perspective, the major research emphases in performance-directed synthesis have evolved in terms of the generation of both synthesis and analysis tools. This not only ensures that algorithms for design at several levels of representation are provided, but also that their well-formedness is ensured through verification techniques, and that efficient performance assessment tools are available to rapidly assess the goodness of a particular design alternative. The emphasis in much of this research has been on the development of formal techniques for design specification, verification, and analysis, thus providing a fundamental approach to the generation of high-performance circuits. Indeed, many of the contributions from this work are of a representational nature, where appropriateness for design synthesis, accuracy, computational efficiency, and insightfulness at the several levels of design are the goals.

The circuit specification is a particularly important level for design, since it is the most abstract level at which performance can be characterized explicitly. For this reason, it is a good intermediate level in design, and is the right place to specify many aspects of the design in order to achieve high performance. Nevertheless, parameters of circuit specification must be unified with those parameters that are strictly derivable from layout through a transformation between these two levels of representation. Baltus[2] has constructed an

algorithm for this purpose which converts any MOS circuit (NMOS or CMOS) to compact layouts by using a design hierarchy with gate matrix structures as basic layout forms that can be efficiently combined in an overall structure. Substantial experience with this program has shown that efficient layouts can be generated even when device sizes change substantially, and when there are unequal numbers of n and p devices. Dynamic programming techniques are used to minimize diffusion breaks, and the complete variety of MOS circuit styles can be readily accommodated. For this reason, the program is not only highly flexible, but also capable of providing efficient layouts in several design styles. For example, this program has recently been coupled to a high-level design procedure developed by J. Kaplan, which utilizes CMOS pass gates in order to achieve layout compactness. In this way, higher level tools can perform a transformation from a behavioral or architectural specification to the circuit representation, with the understanding that an efficient layout can always be generated from this point. The output of the program is in symbolic form, which can easily be compacted to final geometric specifications, thus allowing the ability to track technological advances. Recently, this work has been extended to high-level techniques for the efficient realization of architectures that have nonlocal connectivity. An interesting research issue is the degree to which highly efficient layouts and circuit representations can be achieved for architectures that specify a limited amount of nonlocal interconnections, as is the case in some two-dimensional systolic designs. This program is very successful, and it has been utilized to produce benchmark layouts for a forthcoming Physical Design Workshop.

A considerable amount of research is devoted to the characterization of waveform transitions in integrated circuits for delay modelling. These techniques have been used for both interconnect structures as well as nonlinear circuits involving active devices.

---

2 D.G. Baltus and J. Allen, "SOLO: A Generator of Efficient Layouts from Optimized MOS Circuit Schematics," *Proceedings of the 25th Design Automation Conference,* June 1988.

Recently, S. McCormick[3] has introduced the notion of modelling and simulation for these waveforms by using moments. The moment representation is derived from a Laplace transform representation of a single waveform transition by expanding the transform as a power series, and then truncating the expansion to include only the first n terms (n is usually no higher than four). The generation of the moment representation from the time domain is unique and straightforward, but the inverse problem of conversion from the moment representation to the time domain requires heuristics, which have been developed using a limited class of possible waveforms that occur in the circuits of interest. S. McCormick has also provided general linear network solutions for the moment representation, so that a wide class of circuits can be successfully and efficiently analyzed for characterizing time delay, charge sharing, delay through distributed RC lines, coupling between RC lines, and even two-stage coupled RLC lines. A fundamental theorem has been developed which specifies the number of moments needed in the linear network calculations in order to achieve a result accurate to a specified number of moments. This result guarantees the accuracy of these techniques, which are also highly efficient and provide a new level of generality for such analyses. In addition, great efficiency is achieved for transmission line solutions using the moment representation, which would otherwise have to be simulated (less accurately) by discretized circuits using SPICE. In fact, considerable efficiency is achieved by the moment representation, due to the fact that an entire node voltage transition can be characterized by a single moment representation, as opposed to the need to provide sampled values in direct-method simulation. Finally, S. McCormick has extended earlier work by L.M. McCormick (née Brocco) toward the development of new macromodels for non-linear networks. Whereas before, the basis for such macromodels was parameterization

in terms of input waveform slope and output capacitive loading, new techniques provide a more accurate input waveform parameter and output load parameter which have increased the accuracy of these macromodelling techniques. Since macromodelling converts non-linear circuits to an equivalent linear circuit representation, moment techniques can then solve these circuits by using the previously derived apparatus for the manipulation of moment representations. These macromodels have been used successfully for inverter chains with highly varying capacitive load, the drive of a highly resistive polysilicon line, multistage logic, conducting CMOS transmission gates, and switched CMOS transmission gates. Finally, the macromodelling and moment techniques can be used together to compile overall circuit behavior, leading to very high simulation speeds together with waveform accuracies in the order of 5 percent of values obtained by comprehensive direct methods.

The accuracy of a direct-method circuit simulator is limited by the inaccuracies of the device models it employs. Reichelt[4] is developing a new mixed circuit/device simulator aimed at avoiding these inaccuracies by solving the Poisson equation and the current continuity equations at the device level, while simultaneously solving the differential equations governing circuit operation. Effectively, this amounts to the incorporation of a semiconductor device simulator into the circuit simulator, and for very high-speed operation, increased accuracy is achieved through these means. Unfortunately, the computational expense of transient device simulation has limited the feasibility of mixed circuit/device simulation. Now, however, waveform relaxation iterative techniques are being used, which are both fast and parallelizable, so that semiconductor circuits can be accurately simulated in both two and three dimensions. Two theoretical results that indicate the method will be effective have been proved. One result suggests the

3 S.P. McCormick, *Modeling and Simulation of VLSI Interconnections with Moments,* Ph.D. thesis, Dept. of Electr. Eng. and Comp. Sci., MIT, June 1989.

4 M. Reichelt, J. White, J. Allen, and F. Odeh, "Waveform Relaxation Applied to Transient Device Simulation," *Proceedings of the International Symposium on Circuits and Systems,* 1988.

waveform relaxation method will converge in a uniform manner, independent of the time interval; and the other result implies the multirate integration involved in the simulation will be stable.

In custom design, it is important to guarantee that circuit style methodologies are rigorously followed, and that layouts provide the requisite topology to correspond to these circuits. While a variety of rule-based techniques have been introduced for this purpose, they are effectively open-ended, and do not provide a closed and concise characterization of the methodologies being applied. Instead, Bamji[5] has introduced the use of formal graph grammars which can concisely characterize a wide variety of MOS circuit styles. These can be efficiently and incrementally parsed, and a successful parse at the circuit level indicates that the circuit under investigation does adhere to the requisite methodology. Circuit methodologies can be mixed in this approach, and the overall parsing speeds are very fast. Since they are also incremental in nature, the parsing process can be run in the background of other CAD design programs. Grammars have also been specified for layout representations, and an important part of this work shows how to couple the circuit and layout grammars together to provide one unified representation of these two levels of design. This correspondence is established in a way that does not demand one-to-one coupling between the two levels, which is an important requirement for realistic designs.

Building on the work of Bamji, Van Aelten[6] has extended the grammatical rule base to build a semantic interpreter on top of the parsed circuit design representation. Because the grammar is effective for charac-

terizing structural relationships, the semantic tests are used for a variety of other purposes, including the analysis of charge sharing, device ratioing, and other constraints which establish the desired meaning correspondence between node waveform values and the more abstract logical representation. In this way, the verification techniques provided by the graph grammar for structural purposes are extended in a formally rigorous manner to embrace a wide variety of circuit-related specifications that must be verified.

A fundamental problem in digital design is the establishment and maintenance of a proper database that shows all different levels of representation consistently aligned between all views. Armstrong[7] is developing such a formal model, which may serve as the basis for incremental database consistency maintenance; guaranteeing that when any level of representation is edited, all other levels of representation are automatically adjusted to maintain consistency. If one level of representation is changed, then other levels which are lower in the level of abstraction will provide many possibilities corresponding to the higher level of representation. This work does not attempt to find the optimal lower level representation, but merely to find one exemplar that can be used as the basis for further optimization within one level of representation. Recent work has built on the exploitation of a variety of features in the X windows system, thus providing an attractive user interface with a minimum of additional code.

For some time, there has been considerable attention paid to the exploration of architectural styles for highly parallel digital signal processing tasks. Miyanaga[8] focused on a particular architectural class that provides a

[5] C.S. Bamji and J. Allen, "GRASP: A Grammar-based Schematic Parser," *Proceedings of the 26th Design Automation Conference,* June 1989.

[6] F. Van Aelten, *Constraining Mechanisms for Circuit and Logic Level VLSI Representations,* S.M. thesis proposal, Dept. of Electr. Eng. and Comp. Sci., MIT, 1988.

[7] R.C. Armstrong, *A Formal Approach to Incremental Consistency Maintenance in Multirepresentation VLSI Databases,* Ph.D. diss. proposal, Dept. of Electr. Eng. and Comp. Sci., MIT, 1987.

[8] H. Miyanaga, *A Systematic Approach to VLSI Implementation for Real-Time Digital Signal Processing,* S.M. thesis, Dept. of Electr. Eng. and Comp. Sci., MIT, January 1989.

systematic means for studying the effectiveness of space-time trade-offs for finite impulse response filters and the fast Fourier transform. This approach provides a systematic means of varying time multiplexing versus parallelism, as well as the assessment of the resulting performance for a variety of digital signal processing tasks. Although initially developed for a limited class of these algorithms, the results are being extended for other requirements, and the detailed circuit-level characterization of the basic cell has been obtained.

From the above descriptions, it is clear that several aspects of high-performance design, analysis, and verification have been addressed in the context of new and innovative design-level representations. The main direction of this group is now to extend these techniques to higher levels of representation, including behavioral and logic levels, and to couple the understanding of performance constraints at the lower levels of layout and circuit representation with the higher and more abstract logic and architectural levels.

4.2 Cellular Array for Image Processing

Sponsors

Rockwell International Corporation
OKI Semiconductor, Inc.
U.S. Navy - Office of Naval Research
  (Contract N00014-81-K-0742)

Project Staff

Professor Bruce R. Musicus, G.N. Srinivasa Prasanna, Bennet Sikes

Low-level image processing operations, such as contrast stretching, compensation for lighting variation, noise suppression, or edge enhancement, often rely on highly repetitive processing of the pixels in the image. In conventional image processing architectures, this characteristic is exploited by pipelining the image data through a computational pipeline which repetitively executes the same instruction on all the data flowing through it. An alternative approach, which we are exploring, is to build a large number of small

processors, and use these processors in parallel to execute the same instructions on different parts of the image. The Connection Machine is the best-known commercial implementation of this architectural idea. Our goal is to explore much simpler and cost effective implementations, which can be carefully matched to the algorithmic domain in order to achieve high performance at low cost.

To better understand hardware, software, and algorithmic issues involved in this approach, we have built a small 16 by 16 array of 256 single-bit processors, packaged on 2 VME boards with data memory, a horizontally microcoded sequencer, and a host interface. Combined with a frame grabber and a 68000 controller card, we have a very high performance machine capable of extremely high speed computation for a particular class of signal processing problems. The array is built from four AAP chips from OKI Semiconductor, and operates at a 6.5 MHz rate, performing 256 bit operations on every clock tick. Both bit-serial and bit-parallel arithmetic are supported. Data memory is specially designed to supply overlapping frames of bit-serial or bit-parallel data to the processor array.

The machine is programmed with a micro-assembler with high-level control constructs and expression evaluation. A linker/loader combines object code from multiple sources, and produces link files capable of being simulated or load files suitable for downloading. A simulator executes microcode programs, simulating the behavior of the entire processor array, sequencer, and much of the host interface. Many commands are available for displaying and/or changing the contents of memory, the array, or the sequencer, for single stepping the program or running with breakpoints, and so forth. The system tracks "undefined" values along with 0's and 1's, helping enormously with debugging. We also finished an operating system for the 68000 host capable of coordinating downloading of programs and data to the array, single stepping and running programs with breakpoints, coordinating DMA transfers of image data, displaying and changing the contents of memory, the array, and the sequencer, and providing various communi-

cation and other services to the array.[9] A microcode kernel cooperates with this operating system, maintaining circular communication buffers, and helping forward microcode service requests to the 68000 host. Finally, we finished several demonstration programs illustrating various filtering and image processing tasks on the hardware.[10]

4.3 Algorithmic Fault Tolerance in Digital Signal Processing

Sponsor

Charles Stark Draper Laboratory

Project Staff

Professor Bruce R. Musicus, William S. Song

Conventional methods for achieving fault tolerant computer architectures rely on triplicating computational resources, and using voter circuitry to reject incorrectly computed results. From an information theory viewpoint, however, a much better error control philosophy would be to use coding techniques to achieve a high level of error recovery with minimal overhead. Essentially, a coder distributes information across a noisy channel bandwidth in such a way that individual noise spikes may destroy a portion of many bits, but not an entire bit. A decoder at the receiver can combine information from the full channel bandwidth to reconstruct the original message, with a very high degree of reliability.

Coding techniques are used heavily in high performance memory systems and in communication networks. These techniques are excellent at protecting modules where data entering at one end is expected to arrive

intact and unchanged at the other end. However, coding techniques are not usually used to protect actual computation. Instead, high levels of fault tolerance within CPU's are traditionally achieved by duplicating or triplicating processor resources, and voting on the results. Another problem with coding is that the coding and decoding procedure adds to the latency of the channel, slowing down any machine using the protected component.

In this project, we are developing a new approach to fault tolerance, in which we can protect certain types of linear computation against processor failure by using a small number of redundant processors to protect each other and to protect the "real" processors. One application uses a bank of analog-to-digital converters operating in round-robin fashion to achieve an overall sampling rate somewhat above the Nyquist rate for the signal. A dither system and digital low-pass filter combine to reduce quantization errors in the front end. This same low-pass, however, can be used to detect and correct temporary or permanent errors in any of the converters, without substantially increasing the total amount of computation. The system is able to trade off additional hardware for greater accuracy and higher levels of fault protection. As converters fail, all that happens is that the effective quantization error increases.

Another application is to protect repetitive linear computation such as the FFT's used in range and velocity doppler sonar processing.[11] Here we use a stack of processors to process multiple scans of sonar data from a phased-array antenna. Each processor does the same linear FFT processing, but on different sets of range cells. Adding extra processors working on linear combinations of the inputs to the other

[9] B. Sikes, *An Operating System for the OKI Array Processor*, S.B. Thesis, Dept. Electr. Eng. and Comp. Sci., MIT, 1988.

[10] B. Musicus et al., *OKI Advanced Array Processor (AAP) - Development Software Manual*, RLE Technical Report No. 539. Cambridge, MA: MIT, 1988. 242 pp.

[11] Song, W.S., *A Fault-Tolerant Multiprocessor Architecture for Digital Signal Processing Applications*, Ph.D. dissertation, Dept. Electr. Eng. and Comp. Sci., MIT, 1989; Musicus, B.R., and W.S. Song, *Fault-Tolerant Signal Processing Machine and Method*, U.S. patent applied, No. 251,572.

processors' inputs allows simple fault detection and correction. Regardless of the number of processors in the system, detecting K simultaneous failures requires only K extra processors; detecting and correcting K simultaneous failures requires only 2K extra processors. When conventional truncation or rounding arithmetic is used, however, then the error checking can only be approximate. In this case, adding more processors improves the accuracy of the fault checking and correction. Generalized likelihood ratio tests are used to select the most likely failure hypothesis, and to perform the most likely fault correction. Realistic systems result which use comparatively small computational overhead (<50 percent) to achieve 100 percent single fault detection and correction. We are presently working with Draper Labs on the design of a sonar system incorporating these concepts. We are also currently trying to extend the idea to cover nonlinear computation.

4.4 Compiling Signal Processing Algorithms into Architectures

Sponsors

Project Staff

Professor Bruce R. Musicus, Dennis C.Y. Fogg, James A. Olsen, Kevin Peterson, G.N. Srinivasa Prasanna

We are currently studying several important and difficult problems which arise when compiling signal processing algorithms into either prespecified architectures, or into custom hardware designs. In general, the problem of optimized high level mapping of algorithms into flexibly defined architectures is extraordinarily difficult, requiring a search over an enormous algorithmic and architectural design space, and requiring matching arbitrarily structured algorithms onto arbitrary architectures. Focusing on signal processing has several advantages, however. There is a large literature on efficient algorithms for many signal processing problems, and there are often clear strategies for converting a signal processing calculation into a variety of alternative forms that may be better suited for particular implementations. These algorithms also typically rely on a large amount of matrix algebra, and are therefore composed of high level modules with regular computation, data access, and control structure. Often much of the branch control inside these high level modules can be precompiled or otherwise anticipated, thus allowing highly efficient pipelining and parallelism. Because we can rely on static analysis of the program, effective optimizing compilers and optimizing Computer-Aided Design packages can be developed. Special techniques which exploit knowledge of signal processing can be used to reorganize the computation into appropriate large scale modules, and then map the large regular computation modules onto specialized hardware. Finally, the real-time constraints and huge computational burden often associated with signal processing applications such as radar, sonar, speech or image processing, often justify investment in special purpose, parallel and pipelined systems.

Our efforts this year have focussed on a small set of issues whose solution is critical to the development of high level algorithm compilers. In one project, we are working on writing a parallel algorithm expert for doing matrix computations and Fast Fourier Transforms (FFT) on linear or general arrays of processors. Because these computations are so highly structured, we can prove theorems setting lower bounds on the computation time for a given architecture with fixed computation, communication, and memory bandwidths. Furthermore, we can derive optimal algorithmic transformations of these problems for a given ratio of computation speed to memory speed to communication speed. Systematic sweep strategies based on Critical Path Method (CPM) scheduling can achieve execution times very close to the optimal bounds.

Our second focus has been on the problem of mapping a given dataflow graph into

custom VLSI.[12] Large numbers of possible designs are attempted for solving the given problem, and each is rated according to multiple performance objectives. The result is a scatter plot illustrating the achievable system design tradeoffs between speed and cost. The human designer can then choose a particular performance range for further exploration. The architectural exploration proceeds in two phases. In the first, the system proposes a variety of possible hardware block diagrams for solving the problem. In the second phase, pipelining is assigned, and the modules are instantiated from a library of possible parts. Decoupled design techniques are combined with sample search methods and heuristics for pruning the search space, in order to reduce the complexity of searching through the design space. We have also found a fast linear programming algorithm for solving approximately for the optimal system tradeoff of speed versus area. Current efforts are focussed on extending this algorithm to handle pipelining, and developing good techniques for efficiently exploring a large enough hardware design space.

4.5 Iterative Algorithms for Stochastic Estimation

Sponsors

Project Staff

Professor Bruce R. Musicus, Dr. Meir Feder, Dr. Ehud Weinstein

Maximum Likelihood (ML) is a well-known technique for generating asymptotically efficient estimates of unknown system parameters using noisy and incomplete observations. Classic ML algorithms include Wiener and Kalman filtering for estimating signals from noisy observations, and auto-regressive, moving-average algorithms for estimating pole-zero models from clean data. Unfortunately, when the signal model has unknown parameters, and the observations are corrupted by noise, the optimal ML algorithm is often difficult to compute. We have been working on a variety of iterative algorithms for solving these difficult ML problems which are similar to the Estimate-Maximize (EM) algorithm. The key idea is to decouple the estimation of the unknown internal signal from the estimation of the unknown parameters. We iteratively estimate the signal, use the signal to estimate the parameters, then use the parameters to build a better filter for estimating the signal. Each step is quite similar to a classical ML filter or parameter estimation calculation, and convergence can be guaranteed to a stationary point of the likelihood function.

One application we have examined has been recursive estimation of pole-zero models from noisy observations. These algorithms resemble Widrow's LMS algorithm, except they can properly handle noisy observations. Each iteration alternates between filtering the next observation sample, and adaptively updating the signal model using this latest signal sample estimate. Another set of applications we have explored is time delay estimation for arrays of receivers. Here each iteration estimates the unknown source signal, then cross-correlates this signal against each receiver to estimate the relative time delay between the source and that receiver. Careful analysis of the convergence behavior has led to highly improved hybrid EM-ML algorithms with super-linear convergence rates for the delay estimates. Future work will continue to focus on the ARMA and beam forming array problems.

---

12 Fogg, D.C., *Assisting Design Given Multiple Performance Criteria.*, MIT Laboratory for Computer Science, Technical Report, MIT/LCS/TM-363. Cambridge, MA: 1988; Fogg, D.C., and R.S. Patil, *Decoupled Design: A Technique for Designing Given Multiple Performance Criteria.* Submitted to IJCAI-89.

4.6   The Vision Chip Project

Sponsors

National Science Foundation
   (Grant MIP 88-14612)
DuPont Corporation

Project Staff

Professor John L. Wyatt, Jr., Professor Berthold Horn, Professor Hae-Seung Lee, Professor Tomaso Poggio, Professor Charles Sodini, Craig Keast, Ibrahim M. Elfadel, Mikko Hakkarainen, David L. Standley, Ignacio McQuirk, Christopher B. Umminger, Woodward Young

4.6.1   Overall Project Description

Problem Definition and Methods

Computational Demands of Vision and the Smart Sensor Paradigm: A major problem in machine vision is the sheer quantity of input data to be acquired, managed and processed. The extraordinary volume of data in real-time grey-level images leads to communication bottlenecks between imager, memory and processors, while the computational demands remain high once data have reached the processors. The result is that conventional machines are incapable of any but the most rudimentary forms of real-time vision processing.

A new approach to this problem -- the *smart sensor paradigm* -- is presently emerging and shows great promise. It is potentially applicable to a variety of sensor types, e.g., tactile and acoustic arrays, but our attention is restricted in this project to vision. The key idea is to incorporate signal processing as early as possible into a system's signal flow path to reduce demands for transmission bandwidth and subsequent computation. In this paradigm the major themes are:

• sensors and processing circuitry integrated on a single chip,

• parallel computation,

• processing circuits distributed throughout the array to do spatially local operations *in situ* with minimal wiring overhead,

• the use of analog circuits for early processing tasks to avoid the speed and area costs of analog-to-digital conversion on high-bandwidth data streams,

• selection of tasks and algorithms requiring low to moderate precision, and special emphasis on computations that map naturally to physical processes in silicon, e.g., to relaxation processes or to resistive grids.

Analog Integrated Circuits: Both continuous-time and sampled-data analog integrated circuits play a key role in this proposal. They offer substantial advantages in speed, functionality per unit silicon area, and ease of interfacing to inherently analog input data. (Digital circuits are superior for ease of design and the unlimited precision obtainable by using progressively longer wordlengths.) The Gilbert 2-quadrant multiplier is a dramatic example of analog capability. It requires only four transistors, and similar but slightly more complex circuits achieve 0.02 percent precision multiplication in under 200 nanoseconds with a dynamic range of about five orders of magnitude![13] More generally, analog circuits that perform such basic operations as addition, subtraction, multiplication, thresholding, absolute value, logarithm, exponentiation, time integration and simple linear filtering are much more compact and, in most instances, faster than corresponding digital implementations. They can be fabricated in array form on the same chip as transducers and operate directly on transducer outputs.

The price to be paid is in the inevitable drift and imprecision of analog integrated circuits, which necessitate careful work by experienced designers. They render analog methods inappropriate for certain tasks but

---

[13] B. Gilbert, "Four Quadrant Analog Divider/Multiplier with 0.01 Percent Distortion," *ISSCC Digest of Technical Papers,* pp. 248-249, February 1983.

are not major drawbacks in early vision applications, where the input data (pixel intensities) are rarely accurate to much better than 1 percent (about 7 bits equivalent). The demand is for a *massive volume of computation* rather than for high precision.

This proposal is built in part around the speed and area advantages of analog *systems,* as distinct from single components like multipliers. For example, the CCD 2-D *binomial convolver* we propose to build is conceptually an analog systolic array, capable of massively parallel computation at an estimated 10MHz rate using minimal area. The *parallel spatial correlator* we are proposing achieves a similar level of parallelism in a small space using the special properties of resistive grids.

Advances in parallel computation and VLSI technology will also enable continued growth in general purpose digital computer capabilities. We expect that advanced digital machines will long remain the tool of choice for high-level, i.e., *late,* vision tasks such as object recognition, while the analog implementation of the smart sensor paradigm will enable enormous further reductions in the machine size, power consumption, speed and cost required for low-level or *early* vision processing.

Project Direction and Goals

The goal of this project is to develop novel, high-performance silicon systems for several real-time processing tasks in early vision. We will pursue two parallel routes. The first leads to *single-chip sensor and processor systems,* while the second leads to a *modular early vision system.*

As used here, the term "single-chip sensor and processor system" refers to any chip that acquires image data through a photoreceptor array, has processors spatially distributed throughout the array, and performs a specialized form of parallel computation resulting in low-bandwidth output. Most systems designed in Carver Mead's Laboratory are of this form.[14] These systems are physically small, cheap to manufacture and specialized in application. The term "modular early vision system" refers to a larger, much less specialized system that will require board-level implementation or, perhaps eventually, wafer-scale integration. Its primary output will consist of arrays of intensity, depth and velocity information, and it will be useful as a front-end processor for a variety of vision applications.

These two routes occupy complementary positions in our program, and we believe it is necessary to follow both. The former will yield a working system quickly and enable us to evaluate our approach at an early stage. The latter allows for continued leverage off past work for future system designs, but the cost is a major, time-consuming initial investment in developing the early vision modules.

This project has distinct goals that lie both within the three-year funding period of this proposal and beyond. In three years, we will make components that are useful for a variety of early vision tasks. In five years we will make a more general early vision system using these components and a more advanced specialized component useful for robot navigation. Our long-term goal is an advanced early vision system with highly processed, low-bandwidth output representing robustly segmented images.

---

[14] The silicon retina only partially fits this category as defined above. That system passes entire images off chips at a high data rate and may not be considered very specialized, within the vision domain; J. Tanner, and C. Mead, "An Integrated Optical Motion Sensor," *VLSI Signal Processing II,* (Proc. ASSP Conf. VLSI Signal Processing, November 5-7, 1986, University of California, Los Angeles), S.Y. Kung, R.E. Owen and J.G. Nash, eds., IEEE Press, 59-76, 1987; M.A. Sivilotti, M.A. Mahowald and C.A. Mead, "Real Time Visual Computations Using Analog CMOS Processing Arrays," in *Advanced Research in VLSI: Proceedings of the 1987 Stanford Conference,* ed., P. Losleben, 295-312. Cambridge, Massachusetts, MIT Press, 1987; C.A. Mead, and M.A. Mahowald, "A Silicon Model of Early Visual Processing," *Neural Networks* 1 (1): 91-97 (1988); S. DeWeerth and C.A. Mead, "A Two-Dimensional Visual Tracking Array," to appear in *Advanced Research in VLSI: Proceedings of the 1988 MIT Conference.* Cambridge, Massachusetts, MIT Press, 1988; C.A. Mead, *Analog VLSI and Neural Systems.* Reading, Massachusetts, Addison-Wesley, to be published in 1989.

Three-Year Goals: For the modular early vision system we will build three components that will be useful for a wide range of image processing tasks: an analog charged-coupled-device (CCD) restoring *image memory*, a parallel pipelined CCD 2-D *binomial convolver*, and a parallel *spatial correlator* using resistive grid averagers. These three systems are the foundation for our longer-term plans outlined on the following pages.

MOSIS presently has no CCD capability. Therefore we will develop a CCD enhancement to MIT's baseline CMOS process to fabricate the CCD binomial convolver and image memory. We plan to explore both MOS and bipolar designs for the resistor circuits and arithmetic circuits used in the correlator: the fairly rapid turnaround CMOS process at MOSIS and the BiCMOS process on campus give us valuable flexibility in developing this system.

Design and simulation of the modular early vision system for which these components are intended will also be completed in the third year. These simulations will be based in part on experimental data from test circuits and will be used to modify the specifications for these circuits as a function of simulated overall system performance.

For the single-chip sensor and processor part of the project we will fabricate a working single-chip *image sensor and moment extractor* through MOSIS in the second year. Experimental data from testing and evaluating it will play a role in our development of a more general performance evaluation methodology for early vision systems.

Five-Year Goals: The work in this project is organized around two five-year goals. The first is a *single-chip motion sensor and analyzer* useful for robot navigation. It is based on a new algorithm described in this proposal. This is a complex system that will use many of the circuits developed for the simpler image sensor and moment extractor.

The second goal is a *modular early vision system:* a relatively inexpensive, low-power, single-board system that will operate in real time. Its sensor arrays can be placed at the focal plane of a pair of lenses to acquire binocular image input directly from the environment, and its outputs will consist of the intensity field, velocity field and depth map of the scene. Edge detector circuits will detect and enhance discontinuities in the individual fields.

The CCD imager, image memory, spatial correlator, and binomial convolver with edge detector circuits are key components in this system. Our current plans call for discrete- and continuous-time circuits with continuous voltage levels used to represent pixel values: only the edge detector outputs will be binary. This system will be of real industrial value as a front-end processor for such vision applications as robot navigation and object recognition.

4.6.2 Progress in 1988

The initial funding for this project was received from the National Science Foundation in September 1988 and additional funding from the DuPont Corporation was received in December 1988. We are just starting up.

Here's what we accomplished in 1988:

1. Mr. Craig Keast and Professor Charles Sodini have designed the additional process steps that will enable the MIT CMOS fabrication line to produce CMOS/CCD chips. Craig has designed a test chip with various 4-phase shift register structures and is taking it through each of the process steps by hand to debug the process. The first set of test chips is expected about March 1, 1989.

2. Mr. David Standley and Professor John Wyatt have developed a new stability theory for nonlinear, active resistive grid systems of the type we expect to be useful in a variety of parallel vision processing tasks. The theory yields novel, explicit design criteria that guarantee overall system stability.

3. Mr. Woodward Yang, working with Professor Tomaso Poggio at MIT and Dr. Alice Chiang at Lincoln Laboratories, has completed the architecture definition and circuit design for a parallel, pipelined 2D CCD binomial convolver, which is a very

257

high speed form of image low pass fil-
tering device. The chip layout should be
completed and ready for fabrication
about March 1, 1989.

Publications

Journal Articles to be Published

Standley, D.L., and J.L. Wyatt, Jr., "Stability
Criterion for Lateral Inhibition and Related
Networks that is Robust in the Presence
of Integrated Circuit Parasitics," *IEEE
Trans. Circuits and Systems,* Special Issue
on Neural Networks (1989).

Wyatt, J.L., Jr., and D.L. Standley, "Criteria
for Robust Stability in a Class of Lateral
Inhibition Networks Coupled Through
Resistive Grids," *Neural Computation* 1
(1): (1989).

Papers Presented at Meetings

Standley, D.L. and J.L. Wyatt, Jr., "Circuit
Design Criteria For Stability In a Class of
Lateral Inhibition Neural Networks," *Proc.
IEEE Conf. Decision and Control,* Austin,
Texas, December 1988.

Standley, D.L. and J.L. Wyatt, Jr., "Stability
Theorem for Lateral Inhibition Networks
that is Robust in the Presence of Circuit
Parasitics," *Proc. IEEE Int. Conf. on
Neural Networks,* San Diego, California,
July 1988.

Wyatt, J.L., Jr., and D.L. Standley, "Circuit
Design Criteria for Stable Lateral
Inhibition Neural Networks," *Proc. 1988
IEEE Int. Symp. on Circuits and Systems,*
Helsinki, Finland, June 1988.

Technical Reports

Horn, B.K.P.H., *Parallel Networks for
Machine Vision,* Artificial Intelligence
Laboratory Memo No. 1071. MIT, 1988.

Standley, D.L., and J.L. Wyatt, Jr., *Circuit
Design Criteria For Stability In a Class of
Lateral Inhibition Neural Networks,* VLSI
Memo No. 88-477. MIT, 1988.

Wyatt J.L., and D.L. Standley, *A Method for
the Design of Stable Lateral Inhibition
Networks that is Robust in the Presence
of Circuit Parasitics,* VLSI Memo No.
88-429. MIT, 1988.

4.7 Techniques for Logic Synthesis, Testing and Design-for-Testability

Sponsors

Defense Advanced Research Projects
Agency/U.S. Navy - Office of Naval
Research (Contract N00014-87-K-0825)

Project Staff

Professor Srinivas Devadas

This research focuses on the optimization of
combinational and sequential circuits speci-
fied at the register-transfer or logic levels
with area, performance and testability of the
synthesized circuit as design parameters. The
research problems being addressed are:

1. Area and performance optimization of
   general sequential circuits composed of
   interacting finite state machines,

2. Test generation for general sequential cir-
   cuits without the restrictions of Scan
   Design rules and

3. The exploration of relationships between
   combinational/sequential logic synthesis
   and testability with a view to the devel-
   opment of techniques for the automatic
   synthesis of fully and easily testable cir-
   cuits.

Interacting finite state machines (FSMs) are
common in industrial chip designs. While
optimization techniques for single FSMs are

---

[15] S. Devadas, "Approaches to Multi-Level Sequential Logic Synthesis," *Proceedings of the 26th Design Automation
Conference,* Las Vegas, Nevada, June 1989.

relatively well developed, the problem of optimization across latch boundaries has received much less attention. Techniques to optimize pipelined combinational logic to improve area/throughput have been proposed. However, logic *cannot* be straightforwardly migrated across latch boundaries when the basic blocks are sequential rather than combinational circuits. We have addressed the problem of logic migration across state machine boundaries to make particular machines less complex at the possible expense of making others more complex.[15] This can be useful from both an area and performance point of view. Optimization algorithms, based on automata-theoretic decomposition techniques, that *incrementally* modify state machine structures across latch boundaries to improve area or throughput of a sequential circuit, have been developed. We are now looking toward developing more global techniques for logic migration in sequential circuits.

Interacting sequential circuits can be optimized by specifying and exploiting the don't care conditions that occur at the boundaries of the different machines. While the specification of don't care conditions for interconnected combinational circuits is a well-understood problem, the corresponding sequential circuit problem has received very little attention. We have defined a *complete* set of don't cares associated with arbitrary, interconnected sequential machines. These *sequential don't cares* represent both single vectors and sequences of vectors that never occur at latch boundaries. Exploiting these don't cares can result in a significant reduction in the number of states and complexities of the individual FSMs in a distributed specification. We have developed algorithms for the systematic exploitation of these don't

cares[16] and are currently improving the performance of these algorithms.

Optimization of single or lumped FSMs has been the subject of a great deal of research. Optimal state assignment and FSM decomposition are critical to the synthesis of area-efficient logic circuits.

The problem of FSM decomposition entails decomposing a machine into interacting sub-machines to improve area or performance of the circuit. We have developed new decomposition techniques based on *factorization* of sequential machines.[17] This form of optimization involves identifying *subroutines* or *factors* in the original machine and *extracting* these factors to produce factored and factoring machines. Factorization can result in submachines which are smaller and faster than the original machine. Experimental results indicate that factorization compares favorably to other techniques for FSM decomposition. We are also currently exploring the relationships between factorization and the optimal state assignment problem.

The problem of optimal state assignment entails finding an optimal binary encoding of the states in a FSM, so the encoded and minimized FSM has minimum area. All previous automatic approaches to state encoding and assignment have involved the use of heuristic techniques. Other than the straightforward, exhaustive search procedure, no exact solution methods have been proposed. A straightforward, exhaustive search procedure requires $O(N!)$ exact Boolean minimizations, where N is the number of symbolic states. We have discovered a new minimization procedure[18] for multiple-valued input and multiple-valued output functions that represents an exact state assignment algorithm. The present state and next state

[16] Ibid.

[17] S. Devadas and A.R. Newton,"Decomposition and Factorization of Sequential Finite State Machines," *Proceedings of the International Conference on Computer-Aided Design,* Santa Clara, California, November 1988; S. Devadas, "General Decomposition of Sequential Machines: Relationships to State Assignment," *Proceedings of the 26th Design Automation Conference,* Las Vegas, Nevada, June 1989.

[18] S. Devadas and A.R. Newton, *Exact Algorithms for Output Encoding, State Assignment and Four-Level Boolean Minimization.* University of California, Berkeley, Electronics Research Laboratory Memorandum M89/8, February 1989.

spaces of the State Transition Graph of a FSM are treated as multiple-valued variables, taking on as many values are there are states in the machine. The minimization procedure involves constrained prime implicant generation and covering and operates on multiple-valued input, multiple-valued output functions. If a minimum set of prime implicants is selected, an minimum solution to the state assignment problem is obtained. While our covering problem is more complex than the classic unate covering problem of two-level Boolean minimization, a single logic minimization step replaces $O(N!)$ minimizations. We are currently evaluating the performance of this exact algorithm and developing computationally-efficient heuristic state assignment strategies based on the exact algorithm.

The problem of four-level Boolean minimization or the problem of finding a cascaded pair of two-level logic functions that implement another logic function, such that the sum of the product terms in the two cascaded functions or truth-tables is minimum, can also be mapped onto an encoding problem, similar to state assignment. We have extended the exact state encoding algorithm to the four-level Boolean minimization case.

After chip fabrication, a chip has to be tested for correct functionality. Logic testing is a very difficult problem and has traditionally been a post-design step; however, the impact of the design or synthesis process on the testability of the circuit is very profound.

Our research in the testing area involves test pattern generation for sequential circuits as well as the development of synthesis-for-testability approaches for combinational and sequential circuits. Highly sequential circuits, like datapaths, are not amenable to standard test pattern generation techniques. We are attempting to develop algorithms that are efficient in generating tests for datapath-like circuits, by exploiting knowledge of both the sequential behavior and the logic structure of the logic circuit.

Recently, there has been an explosion of interest in incorporating testability measures in logic synthesis techniques. Our research follows the paradigm that redundancy in a circuit, which renders a circuit untestable, is the result of a sub-optimal logic synthesis step. Thus, optimal logic synthesis can, in principle, ensure fully testable combinational or sequential logic designs.

The relationships between redundant logic and don't care conditions in combinational circuits are well known. Redundancies in a combinational circuit can be explicitly identified using test generation algorithms or implicitly eliminated by specifying don't cares for each gate in the combinational network and minimizing the gates, subject to the don't care conditions. We have explored the relationships between redundant logic and don't care conditions in arbitrary, interacting sequential circuits.[19] Stuck-at faults in a sequential circuit may be testable in the combinational sense, but may be redundant because they do not alter the *terminal behavior* of a non-scan sequential machine. These *sequential redundancies* result in a faulty State Transition Graph (STG) that is equivalent to the STG of the true machine. We have classified all possible kinds of redundant faults in sequential circuits, composed of single or interacting finite state machines. For each of the different classes of redundancies, we define don't care sets which if optimally exploited will result in the implicit elimination of any such redundancies in a given circuit. We have shown that the exploitation of *sequential don't cares* that correspond to sequences of vectors that never appear in cascaded or interacting sequential circuits, is critically necessary in the synthesis of irredundant circuits. Using a complete don't care set in an optimal sequential synthesis procedure of state minimization, state assignment and combinational logic optimization results in fully testable, lumped or interacting finite state

[19] S. Devadas et al., *Irredundant Sequential Machines Via Optimal Logic Synthesis.* University of California, Berkeley, Electronics Research Laboratory Memorandum M88/52, August 1988; S. Devadas et al., *Redundancies and Don't Cares in Sequential Logic Synthesis,* manuscript in preparation, February 1989.

machines. Preliminary experimental results indicate that irredundant sequential circuits can be synthesized with *no area overhead* and within reasonable CPU times by exploiting these don't cares.

Procedures that guarantee *easy testability* of sequential machines via constraints on the optimization steps are also a subject of research. These procedures address both the testability of circuits under the stuck-at fault and the crosspoint fault model. These procedures may result in circuits that are larger than area-minimal circuits, but which are more easily testable.

The different pieces of the research described above are all focused on an algorithmic approach for the optimal synthesis of custom integrated circuit chips with area, performance and testability as design parameters. The various techniques can be incorporated into an ASIC synthesis system.

Publications

Journal Publications

Devadas, S., and A.R. Newton, "Algorithms for Hardware Allocation in Datapath Synthesis," *IEEE Trans. Computer-Aided Des. Integr. Circuits,* to appear in 8 (6)(1989).

Devadas, S., A.R. Wang, A.R. Newton, and A. Sangiovanni-Vincentelli, "Boolean Decomposition in Multi-Level Logic Optimization," *J. Solid State Circuits,* to appear in April 1989.

Devadas, S., H.-K.T. Ma, A.R. Newton, and A. Sangiovanni-Vincentelli, "MUSTANG: State Assignment of Finite State Machines Targeting Multi-Level Logic Implementations," *IEEE Trans. Computer-Aided Des. Integr. Circuits,* 7 (12):1290-1300 (1989).

Devadas, S., H.-K.T. Ma, A.R. Newton, and A. Sangiovanni-Vincentelli, "A Synthesis and Optimization Procedure for Fully and Easily Testable Sequential Machines," accepted by *IEEE Trans. Computer-Aided Des. Integr. Circuits,* to appear in 1989.

Ma, H.-K.T., S. Devadas, R.-S. Wei, and A. Sangiovanni-Vincentelli, "Logic Verification Algorithms and Their Parallel Implementation," *IEEE Trans. Computer-Aided Des. Integr. Circuits,* 8 (1):181-189 (1989).

Ma, H.-K.T., S. Devadas, A.R. Newton, and A. Sangiovanni-Vincentelli, "Test Generation for Sequential Circuits," *IEEE Trans. Computer-Aided Des. Integr. Circuits,* 7 (10):1081-1093 (1988).

Proceedings

Devadas, S., "Approaches to Multi-Level Sequential Logic Synthesis," *Proceedings of the 26th Design Automation Conference,* Las Vegas, Nevada, June 1989.

Devadas, S., A.R. Wang, A.R. Newton, and A. Sangiovanni-Vincentelli, "Boolean Decomposition in Multi-Level Logic Optimization," *Proceedings of the International Conference on Computer-Aided Design,* Santa Clara, California, November 1988.

Devadas, S., and A.R. Newton, "Decomposition and Factorization of Sequential Finite State Machines," *Proceedings of the International Conference on Computer-Aided Design of Integrated Circuits,* Santa Clara, California, November 1988.

Devadas, S., H.-K.T. Ma, and A.R. Newton, "Easily Testable PLA-Based Finite State Machines," *Proceedings of the Fault Tolerant Computing Symposium,* Chicago, Illinois, June 1989.

Devadas, S., H.-K.T. Ma, A.R. Newton, and A. Sangiovanni-Vincentelli, "Optimal Logic Synthesis and Testability: Two Faces of the Same Coin," *Proceedings of the International Test Conference,* September 1988.

Devadas, S. "General Decomposition of Sequential Machines: Relationships to State Assignment," *Proceedings of the 26th Design Automation Conference,* Las Vegas, Nevada, June 1989.

Devadas, S., H.-K.T. Ma, A.R. Newton, and A. Sangiovanni-Vincentelli, "Synthesis and Optimization Procedures for Fully and Easily Testable Sequential Machines," *Proceedings of the International Test Conference,* September 1988.

Ma, H.-K.T., S. Devadas, A.R. Newton, and A. Sangiovanni-Vincentelli, "An Incomplete Scan Design Approach to Test Generation for Sequential Circuits," *Proceedings of the International Test Conference,* September 1988.

4.8 Mixed Circuit/Device Simulation

Sponsors

International Business Machines
Defense Advanced Research Projects
   Agency/U.S. Navy - Office of Naval
   Research (Contract N00014-87-K-0825)

Project Staff

Mark Reichelt, Professor Jacob White, Professor Jonathan Allen

For critical applications, the four-terminal lumped models for MOS devices used in programs like SPICE are not sufficiently accurate. Also, it is difficult to relate circuit performance to process changes using lumped models. Sufficiently accurate transient simulations can be performed if, instead of using a lumped model for each transistor, some of the transistor terminal currents and charges are computed by numerically solving the drift-diffusion based partial differential equation approximation for electron transport in the device. However, simulating a circuit with even a few of the transistors treated by solving the drift-diffusion equations is very computationally expensive, because the accurate solution of the transport equations of an MOS device requires a two-

dimensional mesh with more than a thousand points.

One approach to accelerating this kind of mixed device and circuit simulation is to use waveform relaxation to perform the transient simulation, not just at the circuit level, but inside the devices being simulated with a drift-diffusion description. In the present investigation, the WR algorithm is being applied to the sparsely-connected system of algebraic and ordinary differential equations in time, generated by standard spatial discretization of the drift-diffusion equations that describe MOS devices. Two theoretical results that indicate the method will be effective have been proved. One result suggests that the WR algorithm will converge in a uniform manner independent of the time interval[20] and the other implies that the multirate integration will be stable[21]. Present work is on accelerating a WR-based 2-D device simulation algorithm.

4.9 Circuit Simulation Algorithms for Specialized Applications

Sponsors

American Telephone and Telegraph
Digital Equipment Corporation
Defense Advanced Research Projects
   Agency/U.S. Navy - Office of Naval
   Research (Contract N00014-87-K-0825)
National Science Foundation
   (Grant MIP-88-58764)

Project Staff

Andrew Lumsdaine, Professor Jacob White, Professor John L. Wyatt, Jr.

For all practical purposes, the general circuit simulation problem has been solved. Given enough time, programs like SPICE or ASTAP are capable of simulating virtually any circuit,

---

[20] M. Reichelt, J. White, J. Allen, and F. Odeh, "Waveform Relaxation Applied to Transient Device Simulation," Paper presented at the International Symposium on Circuits and Systems, Espoo, Finland, 1988.

[21] M. Crow, J. White, and M. Ilic, "Stability and Convergence Aspects of Waveform Relaxation Applied to Power System Simulation," Proceedings of the International Symposium on Circuits and Systems, Portland, Oregon, 1989.

given enough time. Unfortunately, for some types of circuits, "enough time" is too much time for simulation to be a practical part of a circuit design cycle. To obtain large performance increases over a program such as SPICE, one must exploit special properties of the specific problem to be solved. In particular, we are developing fast and accurate simulation algorithms for two applications areas: the simulation of clocked analog circuits like switching filters, switching power supplies, or phase-locked loops; and the simulation of analog signal processing circuits used for early vision.

The first of these, clocked analog circuits like switching filters, switching power supplies, and phase-locked loops, are computationally expensive circuits to simulate using conventional techniques because these kinds of circuits are clocked at a frequency whose period is orders of magnitude smaller than the time interval of interest to the designer. To construct such a long time solution, a program like SPICE or ASTAP must calculate the behavior of the circuit for many high frequency clock cycles.

The basic approach to simulating these circuits more efficiently is to exploit only the property that the behavior of such a circuit in a given high frequency clock cycle is similar, but not identical, to the behavior in the preceding and following cycles. Therefore, by accurately computing the solution over a few selected cycles, an accurate long time solution can be constructed.

Simulating clocked analog systems is an old problem, but this novel approach has led to a very efficient algorithm for the distortion analysis of switched-capacitor filters. The idea is based on simulating selected cycles of the high-frequency clock accurately with a standard discretization method, and pasting together the selected cycles by computing the low-frequency behavior with a truncated Fourier series. If carefully constructed, the nonlinear system that must be solved for the Fourier coefficients is almost linear and can be solved rapidly with Newton's method[22] Transient behavior, important for switching power supply designers, has also been accelerated using similar techniques.[23] In particular, the "envelope" of the high-frequency clock can be followed by accurately computing the circuit behavior over occasional cycles.

The second application area is the simulation of analog signal processing circuits used for early vision. These circuits are expensive to simulate with classical methods because they usually contain large grids of components which must be simulated at an analog level (i.e., one cannot perform simulations at a switch or gate level as is commonly done with very large digital circuits).

Several properties of the analog signal processing circuits can be exploited to improve the efficiency of simulation algorithms. As most of these circuits are arranged in large regular grids, the computation involved is like the computations used to solve certain types of partial differential equations. We expect this research direction will lead us to generalizations of certain types of fast partial differential equation methods.

4.10 Numerical Simulation of Short Channel MOS Devices

Sponsors

Analog Devices, Inc.
Defense Advanced Research Projects
  Agency/U.S. Navy - Office of Naval
  Research (Contract N00014-87-K-0825)

[22] K. Kundert, J. White, and A. Sangiovanni-Vincentelli, "A Mixed Frequency-Time Approach for Finding the Steady State Solution of Clocked Analog Circuits," Paper presented at the Custom Integrated Circuits Conference, Rochester, New York, 1988; K. Kundert, J. White, and A. Sangiovanni-Vincentelli, "A Mixed Frequency-Time Approach for Distortion Analysis of Switched Capacitor Filters," to appear in *IEEE J. Solid State Circuits* 24 (2) (1989).

[23] K. Kundert, J. White, and A. Sangiovanni-Vincentelli, "An Envelope-Following Method for the Efficient Transient Simulation of Switching Power Converters," *Proceedings of the International Conference on Computer-Aided Design,* Santa Clara, California, October 1988.

Project Staff

Jennifer A. Lloyd, Professor Dimitri A. Antoniadas, Professor Jacob White

The model used in conventional device simulation programs is based on the drift-diffusion model of electron transport, but this model does not accurately predict the field distribution near the drain in small geometry devices. This accuracy is of particular importance for predicting oxide breakdown due to penetration by "hot" electrons. There are two approaches for computing the electric fields in MOS devices more accurately. One is based on adding an energy equation to the drift-diffusion model and the second is based on particle or Monte-Carlo simulations.

In the first approach, an energy balance equation is solved along with the drift-diffusion equations so that the electron temperatures are computed accurately. This combined system is numerically less tame than the standard approach, and must be solved carefully. Implementations of the energy balance equation in simulators either circumvent this problem by ignoring difficult terms, or they occasionally produce oscillatory results. Research in this area is developing a simulation program based on the drift-diffusion plus energy equations which is both efficient and robust. A stable numerical method for 1-D simulation has been implemented, and present work is to carry this forward to a a 2-D simulator.

Work on the second approach, solving the Boltzman equation with Monte-Carlo algorithms, is just beginning. We are focusing on issues of the numerical interaction between the computation of the self-consistent electric fields and the simulation timesteps. In addition, we are investigating approaches which parallelize efficiently.

4.11 Efficient Capacitance Extraction Algorithms

Sponsors

Project Staff

Keith Nabors, Professor Jacob White

A fast algorithm for computing the capacitance of a complicated 3-D geometry of ideal conductors in a uniform dielectric has been developed.[24] The method we are using is an acceleration of the standard integral equation approach for multiconductor capacitance extraction. These integral equation methods are slow because they lead to dense matrix problems which are typically solved with some form of Gaussian elimination. This implies the computation grows like n^3 where n is the number of tiles needed to accurately discretize the conductor surface charges. We have developed a preconditioned conjugate-gradient iterative algorithm with a multipole approximation to compute the iterates. This reduces the complexity of the multiconductor capacitance calculations to grow as nm where m is the number of conductors.

4.12 Parallel Numerical Simulation Algorithms

Sponsors

Project Staff

Andrew Lumsdaine, Luis M. Silviera, Ricardo Telichevsky, Professor Jacob White

The key problem in parallelizing many of the numerical algorithms used in circuit and device simulators is finding efficient techniques for solving large sparse linear systems

---

[24] K. Nabors and J. White, "A Fast Multipole Algorithm for Capacitance Extraction of Complex 3-D Geometries," *Proceedings of the Conference on Custom Integrated Circuits,* San Diego, California, 1989.

in parallel. Most parallel matrix solution algorithms fall into one of two general categories, the direct (Gaussian-elimination based) and the iterative, and each presents quite different problems. The computation in the direct approach is not very structured, and is therefore difficult to parallelize. Iterative methods are easily parallelized, but are not as numerically robust.

The direct solution of circuit simulation matrices is particularly difficult to parallelize, in part because methods like parallel nested dissection are ineffective due to the difficulty of finding good separators. For that reason, general sparse matrix techniques are being studied. In particular, the interaction between sparse matrix data structures, computer memory structure, and multiprocessor communication is being investigated (with Professor William Dally). One interesting recent result from simulations is that communication throughput, and not latency, is more crucial to final performance.

To improve the convergence of relaxation methods for circuit simulation, an algorithm is being investigated that is based on extracting bands from a given sparse matrix, solving the bands directly, and relaxing on the rest of the matrix. This approach is effi-

cient because band matrices can be solved in order $log(n)$ time on order n processors, and this approach is more reliable than standard relaxation, because "less" relaxation is being used. A banded relaxation scheme has been developed that automatically selects the ordering of the matrix to best exploit the direct solution of the band, and to automatically select the band size.[25] In addition, an implemention of the method on the Connection Machine is in progress.

As mentioned above, relaxation algorithms for solving matrices are easily parallelized. It is also possible to apply relaxation methods directly to the differential equations. This method referred to as waveform relaxation (WR), and easily develop a parallel algorithm in which different differential equations are solved on different processors. A recently developed variant of the WR algorithm, referred to as waveform-relaxation Newton (WRN), allows for additional parallelism because most of the computation for each of the discretization timepoints for a single differential equation can be computed in parallel. In recent theoretical work, it has been proved that WRN converges globally even when applied to circuits with nonlinear capacitors.[26]

---

[25] A. Lumsdaine, D. Webber, J. White, and A. Sangiovanni-Vincentelli, "A Band Relaxation Algorithm for Reliable and Parallelizable Circuit Simulation," *Proceedings of the International Conference on Computer-Aided Design,* Santa Clara, California, October 1988.

[26] R. Saleh, J. White, A. Sangiovanni-Vincentelli, and A.R. Newton, "Accelerating Relaxation Algorithms for Circuit Simulation Using Waveform-Newton and Step-Size Refinement," submitted to *IEEE Trans. Computer-Aided Des. Integr. Circuits*