Cost-Effective Vision Systems for Intelligent Highway Applications

Personnel

B. K. P. Horn, H.-S. Lee, I. Masaki, C. G. Sodini, and J. L. Wyatt, Jr.

Sponsorship

NSF

Demands for cost-effective vision systems are increasing for multimedia, surveillance, automation, and other related applications. We chose intelligent vehicle applications for our case study because of various social, industrial, and academic reasons. About 40,000 people are being killed by car accidents annually in the US and our society needs new transportation paradigms for increased safety, efficiency, convenience, and environmental friendliness. Over 8 million vehicles are sold annually in the US and the size of market will compensate for research investment on developing new vision chip schemes. From academic point of view, vision chips for vehicles are challenging because of varying weather conditions, high reliability and processing speed requirements, and strict cost requirement.

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 variety of functional modules. Some modules are ASIC (Application-Specific Integrated Circuits) chips for straight-forward early-vision processing while 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 for 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. ❑

Cost-Effective Hybrid Vision System for Intelligent Cruise Control

Personnel
M. Spaeth
(H.-S. Lee and I. Masaki)

Sponsorship
NSF

An essential component of an intelligent cruise control system is a module that calculates the distance to objects in the vehicle's field of view. This module must be able to operate in real time at a relatively high frame rate. In order to achieve these high frame rates the algorithm used to generate the distance map is simplified using a prescribed camera geometry and processing units specially suited to the application.

The system being designed calculates the distance map using a distance from disparity algorithm. Initially, each image is converted into an edge map, for correlation with the other images. Next, the edge positions are computed to sub-pixel accuracy, to increase the resolution of the disparity measurement used to calculate the distance map. Finally, edge positions are correlated between the images, and disparities and distances are calculated. To expedite the search for edge correspondences, the three imagers are mounted on an equidistant baseline with optical axes aligned, constraining corresponding edge positions to a horizontal line in the other images.

In order to compute the edge map and the subpixel edge positions efficiently, the ADAP (Analog/Digital Array Processor) MIMD programmable analog array processor is used. Properly programmed, the ADAP is fully pipelined and can output an edge map value or a sub-pixel resolved edge position at 1 MIPS. The positive or negative edge map can be computed with a 9 sample latency, while a composite map with both positive and negative edges has an 13 sample latency. The sub-pixel edge resolution has an 11 sample latency.

In the current system, sub-pixel edge data is transferred to a PC for the correlation and distance map calculations, but the algorithm is sufficiently simple so that it could later be implemented in an ASIC, facilitating a stand alone system. ❑

High Speed, Low Skew Distributed Clocking

Personnel
V. Gutnik
(A. Chandrakasan)

Sponsorship
Intel Fellowship

The problem of distributing the clock on microprocessor chips is well known, and there are certainly solutions that work well for the current 200 - 500 MHz clock speeds. However, even now the skew and jitter in the clock signal accounts for a noticeable fraction of the total cycle time. As improvements in process technology allow scaling to even smaller dimensions, wire delay is likely to increase just as clock speeds are expected to go up. This combination will make even the current distribution schemes unacceptable.

Most of the work this year has focused on development and theoretical verification of a distributed clock distribution model where the chip clock is generated at multiple points on the chip. The stability of such a network has been verified using high-level analytical models and simulations. Circuits that implement the functions needed are currently being developed. ❑

Applications of a Pixel-Parallel Image Processing Chip for Intelligent Vehicle Control

Personnel

Z. A. Talib and J. C. Gealow
(I. Masaki and C. G. Sodini)

Sponsorship

ONR and NSF

The implementation of several intelligent vehicle applications requires the real-time performance of multiple low-level image processing tasks. In applications such as lane following, obstacle avoidance, and adaptive cruise control, a large amount of the computation resources are spent on the low-level image processing tasks such as template matching, optical flow, and stereo vision. A processor-per-pixel scheme employing a Single Instruction stream, Multiple Data stream (SIMD) design demonstrates a system fast enough to perform real-time image processing, yet flexible enough to be programmed for a variety of image processing applications.

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. Each processing element combines a 128-bit DRAM column with a 256 function one-bit-wide arithmetic logic unit.

Figure 1 illustrates the processor-per-pixel system based on the integrated array. The image processing system is comprised of an image data path and a control path for the transmission of pixel data and instructions, respectively. The processing element array receives instructions from the controller, which is managed by the host computer. Analog images from a video camera or other source are converted to digital signals, then reformatted for processing using the processing element array. Output images from the array are converted to a format appropriate for subsequent use.

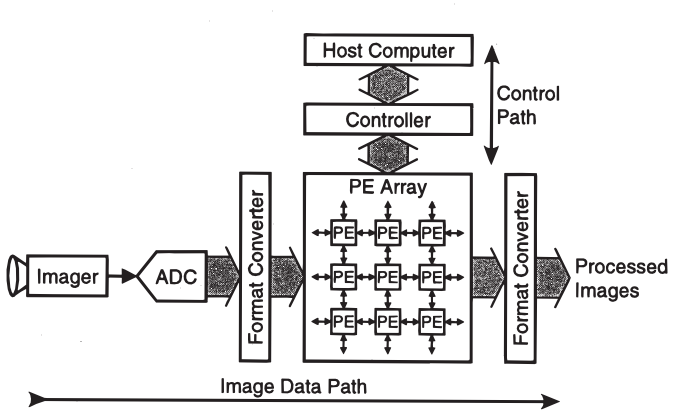

Fig. 1: Pixel Parallel Processor Architecture.

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: median filtering, smoothing and segmentation, edge detection, and optical flow computation. All have been successfully performed in real time with input images provided at standard video frame rate. Experimental results are summarized in Table 1.

Table 1:

Application	Execution Time	Maximum Number of Frames per Second
5 x 5 Median Filter	274 μs	3649
Edge Detection	277 μs	3610
Smoothing and Segmentation	4.80 ms	207
Optical Flow	7.36 ms	135

continued

continued

Current hardware development includes the expansion of the demonstration system from four chips (processing 128 x 128 pixel images) to sixteen chips (processing 256 x 256 pixel images). Current application development includes demonstrating real-time generation of a depth-map using a stereo vision algorithm. Utilizing three cameras separated by a known fixed distance, the stereo vision algorithm determines the absolute distance to an object.

The future plan is to expand and improve the system so that its usefulness can be demonstrated in a real intelligent vehicle control application. While the current system takes raw images as input and returns the full processed images as output, in order to efficiently determine control instructions for a vehicle, it is necessary to extract only the pertinent features (e.g. the location of the edges in the case of edge detection). The low-level information determined by the pixel-parallel processor system would therefore serve as input to a conventional serial microprocessor which would, in turn, determine the appropriate control instructions to deliver to the vehicle. ❑

Block Matching Video Chip

Personnel

Y. Fang
(I. Masaki and B. K. P. Horn)

Sponsorship

Sharp Corporation

Efficient coding of image sequences is important for telephone conferences, portable video phones, and other related applications. Video coding tasks are categorized into the following four categories: intraframe coding, intraframe decoding, interframe coding, and interframe decoding. With current industrial standards such as MPEG-2 and H.261, the interframe coding part is the most computationally intense. The most popular method for interframe coding is motion estimation by block matching. Block matching involves partitioning the current image frame into blocks. For every block in the current image frame, there is a corresponding search window in the previous image frame. The search window size is determined by the size of the current frame block and its maximum displacement in the horizontal and vertical directions. The most commonly used criterion for conventional block matching is mean-absolute error in 8-bit intensity values. The computational load is 6.8 Giga operations per second assuming that the block size is 16 x 16-pixels, the image size is 512 x 512-pixels, the searching range is +/- 8-pixels, and the frame rate is 30 frames/sec. We first proposed a novel criteria function to simplify the algorithm, and then developed a custom chip for the new algorithm.

The algorithm we proposed is binary block matching. With this algorithm, we calculate binary edge images of the current and previous images and correlate binary image blocks in the current image with corresponding binary searching windows in the previous image. The operation load is reduced by replacing conventional 8-bit operations with single-bit operations. The performance of motion estimation is not degraded significantly because reliable motion estimation data can be obtained only in small regions near edges where the intensity changes spatially. Edge detection is added as a new overhead operations but its computational load is much smaller compared to the correlation part. In our computer simulation, the motion estimation obtained with binary block matching increases the mean absolute

continued

continued

error by approximately only 10% in most cases while reducing the computational load by a factor of more than four including the edge detection overhead.

A test chip for binary block matching was developed and is being measured. The input of the chip is binary edge images and the output is correlation values in analog voltage. The chip stores a 16 x 16-pixel current binary edge image and a 32 x 32-pixel previous window edge image, and correlates them in parallel with 16 x 16 processing elements. The silicon area was reduced by a factor of 4.4 by replacing conventional 8-bit correlation with single-bit one. The speed may be doubled because 16-bit adders which are a speed bottleneck with the conventional scheme can be replaced with 8-bit adders with our binary correlation scheme. With the test chip, the 8-bit adders were replaced with analog circuits which add unit-currents from processing elements. The summed current is converted into analog voltage for the output. Preliminary evaluation results of the chip are encouraging and further evaluations are underway. ❑

Brightness Adaptive CCD Imager

Personnel

K. Fife and S. Decker
(C. G. Sodini)

Sponsorship

NSF and DARPA

Many chips designed for early vision are integrated with the imaging array to avoid a communication bottleneck between the imager and the processing array. As more complex algorithms are developed, however, it becomes increasingly difficult to put both imager and processor on the same chip. A dedicated imaging chip with high throughput line parallel outputs can provide some useful features currently unavailable. Basic specifications include a 256 x 256 array of pixels, progressive scan operation, column-parallel output, and 10-b digital output. The maximum operating speed is
1000 frames/sec.

Natural scenes can have image intensities which vary over six orders of magnitude. Typical imaging elements tend to saturate in high intensity regions and produce signals below the noise level in the low intensity regions. Conventional cameras frequently employ a mechanical or electronic shutter to adjust the global integration time. This allows good performance over wide variations in the global illumination, but does not improve imaging of scenes which contain but very bright and very dark regions. The wide dynamic range imager employs a pixel with a lateral overflow drain which can be used to extend the dynamic range. Effectively, it uses a long integration period for regions of low illumination and a short integration period for regions of high illumination. This technique requires little or no additional pixel area, provides a user-adjustable compression characteristic, and can be used with other techniques to enhance the performance at low illumination.

A chip has been fabricated in a 0.8 um 3-metal 1-poly CMOS process offered through National Semiconductor. The increased pixel dynamic range works as expected, producing a dynamic range of at least 80 dB. A printed circuit board has also been produced to display valid frame data at a reduced rate. A real-time camera is currently under development which will allow further demonstration of the wide dynamic range pixel and convert the digital data to any format necessary for signal processing. The first version of the real-time camera will convert the digital data to the standard NTSC signal for a television monitor, while later versions of the camera will integrate multiple imagers onto one circuit board for stereo vision applications. ❑

Three-dimensional Vision System for Traffic Monitoring

Personnel

N. S. Love
(I. Masaki and B. K. P. Horn)

Sponsorship

US Dept. of Transportation

Traffic surveillance cameras are important for controlling traffic flow intelligently. Conventional two-dimensional vision systems for traffic monitoring have two problems: overlapping vehicles and shadows. When two vehicles are overlapping each other, it is difficult to separate them reliably in a two-dimensional image. It is also difficult to discriminate dark-color vehicles from their shadows. Three-dimensional vision systems are superior for those problems because of their capability of measuring distances to objects. Conventional three-dimensional vision systems, however, were too expensive or too slow for traffic monitoring.

We have developed cost-effective real-time three dimensional vision systems. At the processing speed of 30 image frames per second, the distance resolution has been improved from 5-bit with the conventional leading system to 11-bit. This became possible by concentrating the computational power on small "edge" regions in which reliable distance information is available. In those limited regions, sub-pixel operations are carried out for a higher distance resolution while they were conventionally available only for slower operations because of their large computational load. The system was implemented with a Pentium PC and two off-the-shelf plug-in boards. Applications such as vehicle counting and speed measurements are being developed.

The research is also being extended toward "recognition in compressed image domain". Image compression and image recognition were different technologies which were independent of each other. In digital image network, for example, images are transmitted in a compressed format and compressed images are decompressed whenever image processing is required. It is especially time consuming to find some particular object from stored image database because every image must be decompressed before processing. We compress the image data by using attributes which are relevant both for compression and recognition. Examples of common attributes are binary edge locations and the color, intensity, and distance information on pixels only along the binary edges. In preliminary experiments, sufficient compression and recognition performances were achieved by using those attributes. Further investigation is underway. ❑

A High Sensitivity CMOS Imager

Personnel

C.-C. Wang
(C. G. Sodini)

Sponsorship

NSF

A CMOS imager is superior to a traditional CCD imager in the aspects of low power consumption, circuit integration and the ability for pixel random access. A significant drawback to current CMOS imagers is lower sensitivity than CCD's. To enhance the sensitivity of the CMOS imager, two approaches are taken in the research. The first approach is to enhance the quantum efficiency of the photo-sensors, and the second is to reduce the noise introduced into the CMOS imager or remove the noise by circuit techniques.

Quantum Efficiency (QE) is a key parameter for optical sensors. It is defined as the number of incident photons at a specific frequency versus the number of carriers generated and detected by the sensor. It directly determines a sensor's transduction coefficient from an optical signal into an electrical quantity. A high quantum efficiency photo-sensor is desired to enhance the signal-to-noise ratio of the whole system.

Silicon photodiodes and photogates are two sensing structures often used for visible-light solid-state imagers. The QE of the photogate is seriously attenuated by its capped poly gate. The peak QE in visible-light range is typically less than 40%. A photodiode generally has a peak QE around 80-90% but the QE is not optimized at all frequencies. There are two kinds of photodiode structures often used in a standard CMOS process. The traditional n^+-diffusion/p-substrate (or p^+-diffusion/n-substrate) photodiode has a short minority-carrier diffusion length of about 0.25 μm. Electron-hole pairs generated in the n^+ diffusion region will most likely recombine before they are captured by the depletion region. The shallow diffusion-substrate junction is also harmful for red-light absorption because red light can penetrate beyond this depletion region. Both effects reduce the QE. A second photodiode structure can be implemented with a well-substrate junction. This photodiode has a longer diffusion length in the well, but its blue light response is poor. Blue light will generate e-h pairs near the Si surface. Since the well-substrate junction is deeply buried, electron-hole pairs will most likely recombine at the surface. To address the problems mentioned above, a high quantum efficiency photodiode is under research.

To reduce the noise of the CMOS imager, a deep understanding of all the noise sources is essential. A significant noise source in CMOS imagers is the reset noise. Reset noise is the kT/C noise due to the voltage reset at the photo-sensing node. Because the capacitance of the photo-sensing node is typically tenths of fento-farads, a special circuit technique is required to remove the reset noise. Dark current due to randomly generated electron-hole pairs is another noise source. The noise introduced from the signal amplification stages should also be controlled. Low-noise pixel structures and noise reduction circuit techniques are currently under investigation. ❑

Image Processing System-on-a-Chip

Personnel
P. M. Acosta Serafini
(I. Masaki and C. G. Sodini)

Sponsorship
NSF

An image processing system is typically made of an image sensor (imager) and a processing unit. Recent advances in the imager field indicate that the analog processing is well suited for focal-plane applications like dynamic range expansion, or for very specific tasks such as whitespot compensation. Conversely, complex image processing is typically done with programmable digital systems. It has been demonstrated that image coprocessors featuring hundreds of pixel processors are the key for real time operation.

Other recent advances indicate that CMOS imagers have become a competitive alternative for producing digital cameras, in part because analog-to-digital converters can be integrated in the same die as the sensing array. As the desire to have products with better spatial and intensity resolution continues, timing requirements will move imager architectures toward a column-parallel design, where there is one ADC per pixel column.

Up to now column-parallel imagers and image processors have had dissimilar requirements due to timing and architecture differences. A flow-through processor would address and eliminate most of the redundancies and incompatibilities present in the two systems. This new processor receives the data in parallel format from the imager ADCs and then manipulates it in a parallel fashion according to some programmed functionality. The kernels of typical low-level image processing operations are not much larger than 3*3, so it is plausible to think that there is a efficient mapping of these convolution operations onto a reduced number of processing element rows. If this is achieved, it would be possible to have a complete, inexpensive, high performance image processing system-on-a-chip, where the input of the system would be the analog illumination and the output would be the digital, processed data. ❑

A Low Power Image Sensor

Personnel
I. L. Fujimori
(C. G. Sodini)

Sponsorship
DARPA and Lucent Technologies Fellowship

Over the past decade, CMOS image sensors have received much attention in the electronics industry. Compared to its CCD counterpart, CMOS imagers consume lower power, allow random access and can be integrated with analog and digital functional blocks in a standard CMOS process. The cost for these advantages is a reduction in the image quality, namely an increase in sources of noise, such as dark current and Fixed Pattern Noise (FPN). A large portion of the cost and effort of a CCD fabrication process is dedicated to minimizing the pixel dark current and improving the efficiency of the light-to-voltage conversion. In a standard CMOS process, however, the imager designer has little to no control over the fabrication steps. The challenge of a CMOS imager designer therefore becomes utilizing innovative circuit techniques to achieve CCD quality images in a standard CMOS technology.

A CMOS passive pixel (single transistor) cell is a promising implementation that could potentially reduce the effects of the pixel dark current, and FPN normally observed in a CMOS active pixel. A differential architecture (shown in Figure 2) in which the output of a sensing pixel is compared to that of a dummy pixel kept in the dark, is used to reject any common-mode signals such as ground bounce and temperature variations. This differential readout scheme can also be used to subtract the sensing pixel dark current from the dummy cell's dark current, thus reducing the effects of the dark current.

The passive pixel consists of a high-efficiency n-well photodiode and a transistor for row select. The output of the pixel, which is in the form of charge, is converted to a voltage with a sense amplifier at the bottom of every column. Consistent with the low number of transistors per cell, the passive pixel has few sources for fixed pattern noise.

continued

There is one inherent weakness for passive pixels, however. Long wavelength radiation (red and near IR) is absorbed very deep in the substrate of the photodiode. Some of these photogenerated charges will find their way to the depletion region of the photodiode while others may be swept up by the reverse-biased diffusion of the column line. The combined effect of the charge leakage from 256 cells can be significant and will appear as a parasitic current at every column line. Though this parasitic current is also present in active pixels, its effect is more pronounced in passive pixels because charge amplification does not occur within the cell. Fortunately, this signal dependent parasitic current can be removed with correlated double sampling, making the passive pixel a competitive choice in the implementation of CMOS imagers.

While CMOS imagers have not yet achieved the superior imaging quality of CCD's, they offer many advantages for applications where high image quality is not essential, but where low power, low cost and high integration is desired. Some examples include surveillance, biomedical and videoconferencing applications. One of the proposed methods to save power in the wireless camera resembles a feature commonly found in pagers. ❑

Fig. 2

Time-to-Collision Chip

Personnel
H. Cheng
(B. K. P. Horn and I. Masaki)

Sponsorship
NSF

Motion vision methods cannot provide absolute velocities or absolute distances, but can be used to give the ratio of distance to velocity, which is the time to collision between the camera and the surface being viewed. When determining the time to collision, the actual direction of motion is of secondary interest. Hence the solution for this problem is complementary to that of finding the Focus Of Expansion (FOE), where the direction of motion is determined by the position of the FOE in the image, while distances and velocities are not determined. Our approach to finding the FOE is based on information at so-called "stationary points", where the instantaneous rate of change of brightness is zero. These are the very points that provide no constraint on the time-to-collision. Instead, we have to work with the remaining image points, where the rate of change of brightness depends on the distance to the corresponding point in the scene and the velocity. We can use least squares methods to attack this problem if we assume that a portion of the surface being viewed is nearly planar. Then one has to recover the dot-product of the surface normal with the velocity vector. This can be accomplished by solving a non-linear optimization problem. There are still some questions about the computational work required to find the optimum, as well as the existence of local minima. Stas Frumkin explored this approach in simulation and verified that it works both for synthetic image sequences where the data was perfect and the ground truth accurately known and for a few real sequences where the images are of limited quality and the exact motion was unknown. He also designed part of a digital processor to perform this computation. We are looking into extending this method to second order polynomial surface models. We are also now investigating implementation of the existing algorithm in hardware. ❑

Ultra Low Power Wireless Sensor Project

Personnel
A. Chandrakasan, H.-S. Lee, C. G. Sodini, M. D. Trott, and T. Barber - Analog Devices

Sponsorship
DARPA

This program is developing a prototype wireless image sensor system capable of transmitting a wide dynamic range of data rates (1 bit/s - 1 Mbit/s) over a wide range of average transmission output power levels (10 μW - 10 mW). The prototype will dissipate approximately 50 mW and is based on research IC's designed at MIT. A preliminary block diagram of the prototype wireless camera system is shown in Figure 3.

The imager is a 256 x 256-pixel sensor being designed in a 0.6 μm CMOS technology. The power dissipation of the imager is reduced by decreasing the power supply voltage and incorporating random-access readout. The key issue in reducing the power supply voltage is the minimization of noise.

To optimize power consumption for different signal frequencies and dynamic ranges, a reconfigurable Analog-to-Digital Converter (ADC) is being designed. The concept is based on the fact that the pipeline and sigma-delta ADCs share the same basic analog building blocks such as switched capacitor integrators, subtractors, and voltage comparators. A switch matrix will configure these components into one of the two converter types, depending on the signal frequency and dynamic range requirements. In addition, the converter will be reconfigured within each type (e. g. # of stages, order,) for each performance range.

The digital signal processing aspects of this system focus on low-power image compression, encryption, and the digital control of the entire system. Efficient power-down techniques are used where only the most significant bits representing the intensity of the image are processed when higher resolution images are not required.

continued

continued

The circuit design of the RF front end is focused on the implementation of narrowband transmit and receive functions that are optimized for low-power operation. The transmitter architecture eliminates the need for power-intensive mixers and D/A converters. The VCO employs bond wire inductors placed on chip and noise-matching techniques. These chips have been used to demonstrate a DECT transmitter capable of 1.25 Mb/s at 1.8 GHz using only 27 mW.

Various techniques are employed to achieve an efficient power supply at low voltage and power levels, including delay-line PWM, discontinuous-mode operation, and shared control for multiple outputs. The circuit is programmable to accommodate adaptive voltages and power levels. ❑

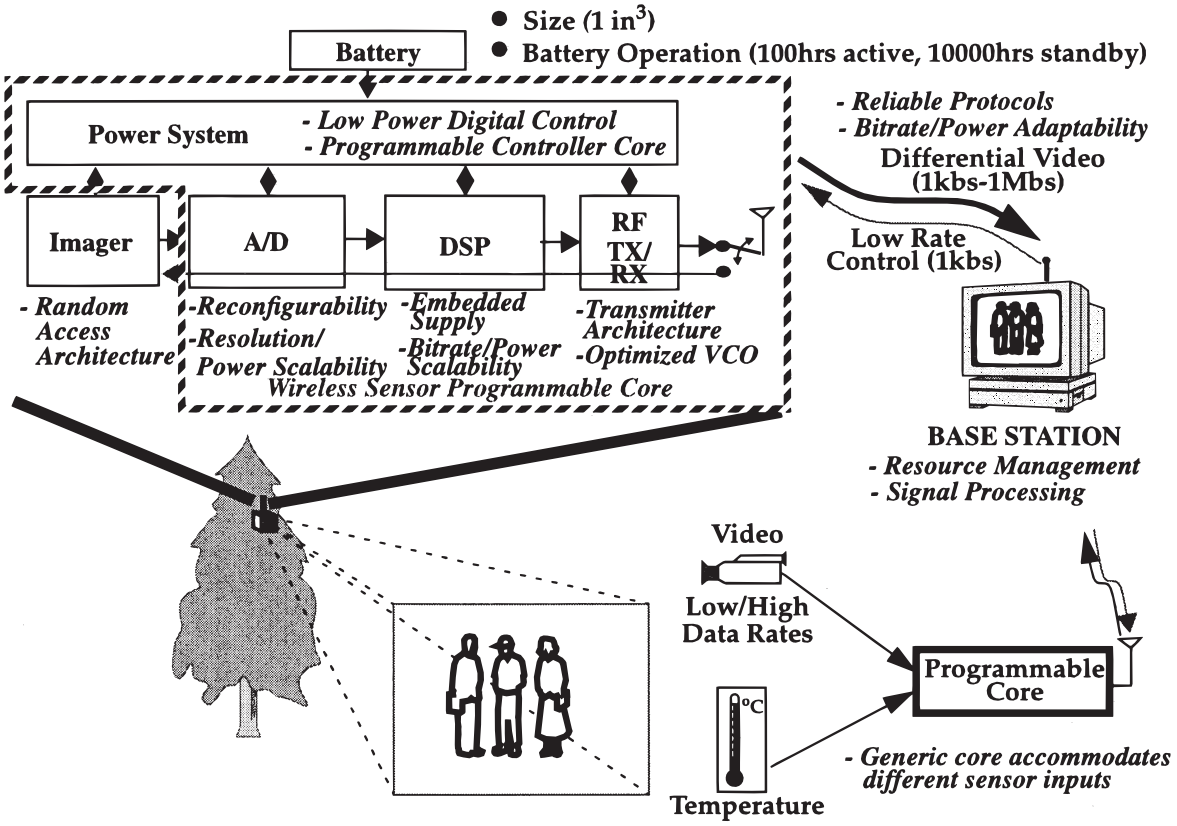

Fig. 3

Low Power Reconfigurable Analog-to-Digital Converter

Personnel
K. Gulati
(H.-S. Lee)

Sponsorship
ARPA

The objective of this project is to create a low power A/D converter to be employed in a battery powered wireless sensor that can handle a variety of signals ranging from high data-rate medium-resolution video to low-frequency high-resolution temperature signals. The key requirement of the converter, therefore, is that it be able to cater to a wide range of input data-rate, (1sample/s to > 50 Msamples/s) and produce a resolution ranging from 8 bits to as high as 20 bits. This large size of data-rate versus resolution space warrants the need for several different A/D converter architectures. However, such a converter implementation would require prohibitively large area and power. To be able to cover the entire rate-resolution space, a reconfigurable A/D converter is proposed. Another requirement of the ADC is that it consume the lowest power possible. To achieve this a host of techniques are being employed, one of them being a new low power.

The opamp is the most power-hungry device in the A/D and consequently significant work has been done to reduce the power consumption of this device. Given the observation that SNR, speed and power consumption of an opamp are all interrelated and, specifically, that the power consumption of the opamp is inversely proportional to the square of the SNR (or swing of the amplifier, for a fixed input referred noise), a new operational amplifier topology was created.

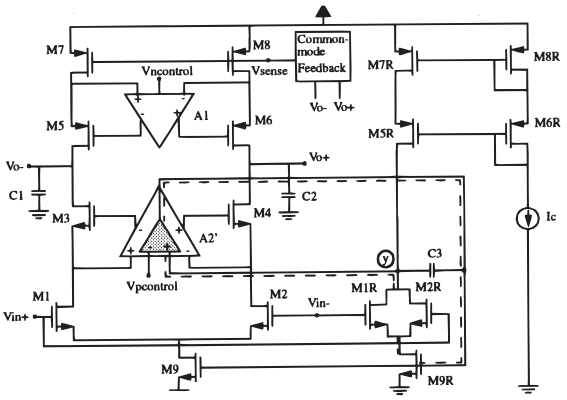

Fig. 4: Implementation of Operational Amplifier.

A telescopic cascode op amp typically has a higher frequency capability and consumes less power than other topologies. The disadvantage of a telescopic op amp is the severely limited output swing.

The opamp developed in this project (Figure 4) uses the telescopic architecture as its core for reasons of high speed and low power consumption and offers much higher output swing than a conventional telescopic amplifier while maintaining high CMRR and supply rejection (PSRR), and ensuring constant performance parameters. Transistors M7-M8, and M9 are deliberately driven deep into the linear region. In this case, the output swing is improved by 0.7V from a telescopic amplifier and becomes slightly better than that of a folded cascode amplifier. The reduction of gain and CMRR due to the low output resistance in the linear region is compensated by gain enhancement and replica tail feedback, respectively. The gain enhancement employs the well-known differential regulated cascode structure except the control voltage Vncontrol is chosen to bias M7-M8 in the linear region. The gain enhancement amplifier A2' incorporates the replica tail feedback to keep the drain current of M9 constant despite input common-mode voltage variation.

The basic goal of the replica tail feedback (Figure 5) is to keep the tail current constant. It accomplishes this by modulating the gate voltage of the tail transistor. Transistors M1, M2 and M9 represent the input devices and

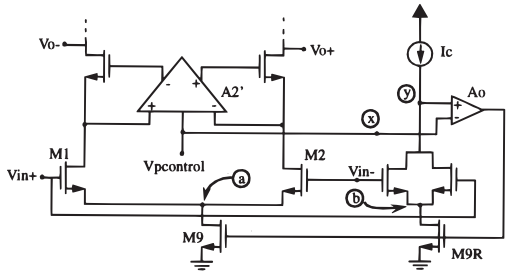

Fig. 5: Basic concept of replica tail feedback.

continued

tail current source of a differential amplifier, while M1R, M2R and M9R form their corresponding replicas. Amplifier Ao is placed in negative feedback across the replica circuitry which forces the voltage at node y to equal the voltage at pcontrol. Also, A2' forces the common-mode component of the drain voltages of M1 and M2 to equal the voltage at pcontrol. Under these conditions, the voltage at the drain of the tail transistor (node a) always equals the voltage at the drain of the replica tail transistor (node b). Since current through M9R is fixed by Ic, current through M9 remains fixed even if it is biased in the deep linear region.

A prototype of the operational amplifier was fabricated (Figure 6) using a single-poly, three-metal 0.8 um HP process. The measured/simulated specifications for the opamp are detailed in Tables 1 and 2. As part of continued work on this project, the ADC design will be completed, fabricated, and characterized. ❑

Technology	0.8 µm CMOS
Power Supply	3.3V
Die Size	600 µm x 630 µm
Power Consumption	4.8 mW

Table 1: Chip specifications.

Fig. 6: Op amp microphotograph.

	Simulation	**Measurement**
Differential Open Loop Gain (500 Hz)	105 dB	90 dB
Differential Output Swing	+/- 2.4V	+/- 2.45V
Unity Gain Frequency (Output Load = 3.52 pf)	93 MHz	90 MHz
Phase Margin	78 degs	--
Settling Time (1%; noise gain > 2, Output Load ~3.65pF)	17.3 ns	26 ns
Slew Rate	133 MV/ns	125 mV/ns
Offset (typical)	9.5 mV	1-2 mV
CMRR (500 Hz)	57 dB	>50 dB

Table 2: Op amp measured and simulated performance.

Low Voltage, Low Power CMOS Operational Amplifier Design for Switched Capacitor Circuits

Personnel

P. M. Naik
(H.-S. Lee)

Sponsorship

NSF Fellowship and DARPA

The increasing demand for circuits with low voltage operation and low power consumption are being driven by smaller submicron technologies and the move towards more portable electronic systems. Lower supply rails are necessary due to smaller device breakdown voltages resulting from smaller feature sizes. In addition, low voltage is desirable so fewer batteries are required, thus reducing system size and weight. Low power consumption is necessary to ensure reasonable battery lifetime.

These demands present challenges in circuit design, particularly in the analog domain. Unlike digital circuits where power consumption reduces proportionally to the square of the supply voltage, in analog circuits lower voltage actually increases power consumption. Another design challenge that arises is lower drive availability, which decreases device speed. Reduced dynamic range is yet another problem since the supply voltage is reduced while the device threshold voltage remains the same. The objective of this research is to design a low voltage, low power CMOS operational amplifier with high performance for use in switched capacitor circuits.

Performance Parameter	Design Goal	Simulation Results
Total Power	< 100 µW	95.29 uW
0.1% Settling Time	< 100 ns	65.Ins
Gain (DC)	> 10K (80 dB)	85.44K (98.6 dB)
Swing (>20% inc. gain)	max	+/- 1.54V
Unity Gain Frequency	> 15 MHz	41.48 MHz
CMRR	max	69.0 dB
PSRR+	max	46.92 dB
PSRR-	max	45.65 dB

Table 3: Opamp Results.

Table 3 shows some key design goals and Hspice simulation results of the proposed opamp, shown in Figure 7. The power supply is +/-0.9V, the capacitive load is 200fF and the process used was the HP 0.5 um minimum feature size. The design uses a simple two-stage topology with added features to improve performance. These features include cascoding to improve gain; a replica tail to provide constant current which improves common mode input range, supply and common mode rejection ratios; a local switched capacitor common mode feedback circuit; and a bias circuit tracking the opamp. A layout of this chip using a triple metal, single poly n-well process has been completed and sent for fabrication through Mosis. Testing will be completed and results will be analyzed. ❑

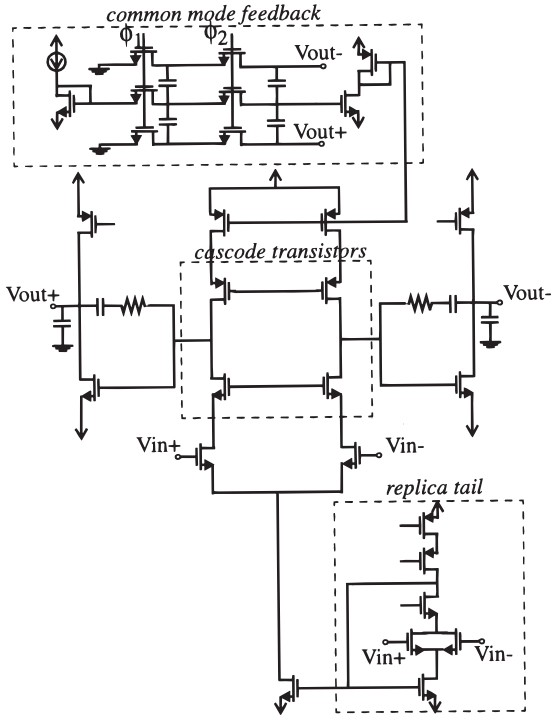

Fig. 7: Proposed Opamp Design (bias circuit not shown).

Compression for Wireless Video Transmission

Personnel
T. Simon
(A. Chandrakasan)

Sponsorship
DARPA

Video compression is an integral part of the digital signal processing of the Ultra Low Power Wireless Sensor Project. Image compression, or more generally compression of any source data, is used to minimize system power by trading local computation for transmitted bandwidth. The goal of this project is to optimize the power efficiency of compression at the algorithmic, architecture, and circuit levels.

The image compression work must produce an algorithm/architecture capable of optimizing system wide power for widely differing output bit rates. During, possibly long, periods of no motion in the image source, the system power is determined by the computational cost of determining there is no movement, the operation of the sensor and converter, and the standby losses of all modules and power supply. Besides minimizing standby losses in individual modules through architectural and circuit techniques, system power may be reduced by the image compression module by adaptive control of frame rate, bit resolution, and feedback regarding optimal operating voltages to the power supply.

During periods of motion in the image source, transmission power becomes a significant part of system power. The compression module must achieve good compression performance without using undue computational power. The baseline algorithm uses hierarchical wavelet transformation and zero-tree based significance coding, with compromises designed to localize communication and limit the scope of filtering in the time dimension to reduce frame memory requirements.

The architectural and circuit work is centered around a massively parallel SIMD array of very simple computational/memory cells. The SIMD array is architecturally well suited to running the compression algorithm with great area and power efficiency. The parallel array enables a very low clock rate and hence a low supply voltage. A Verilog simulation has been performed to verify chip functionality. The circuits (logic and DRAM) are currently being designed, with emphasis on low-voltage CMOS operation. ❑

Low Power Video Decoding

Personnel
T. Xanthopoulos
(A. Chandrakasan)

Sponsorship
SHARP Corporation

This work describes the implementation of a low power IDCT chip targetted to medium and low bitrate applications. Our strategy for reducing the chip power was two-fold: first, we selected an IDCT algorithm that minimizes activity by exploiting the relative occurrence of zero-valued DCT coefficients in compressed video. Previous IDCT implementations have relied on conventional fast IDCT algorithms that perform a constant number of operations per block independent of the data distribution. Our approach performs a variable number of operations that depends on the statistical properties of the input data. Second, we minimized the energy through aggressive voltage scaling using deep pipelining and appropriate circuit techniques so that the chip could produce 14 Msamples/sec (640 x 480, 30 fps, 4:2:0) at 1.3V in a standard 3.3V process (VTP = -0.9V, VTN = 0.7V) and meet the requirement for MPEG2 MP@ML.

A novel row-column algorithm is used which treats each incoming DCT coefficient individually and saves operations when the bitstream contains zero-valued coefficients. The IDCT chip consists of two 1-D IDCT units performing multiply-accumulate operations and a transposition memory structure (TRAM) implemented as a 2D shift register. A block diagram is shown in Figure 8. Every 8 cycles, the 1-D IDCT of a block row has converged within the 8 accumulators of the first stage. The intermediate result is transposed on-the fly by the transposition structure. The second 1-D stage performs the same operations on the columns. The multiply-accumulate operations must be pipelined by a factor of 4 to meet the bandwidth requirement at V_{DD} = 1.3V.

A microphotograph of the chip is shown in Figure 9. It is functional over a wide range of frequencies and power supplies. Over 2.8 million separate power measurements on an 8 x 8 block basis have been taken during several weeks while the chip was stimulated from 6 different MPEG2 sequences. The results of these measurements are plotted in Figure 10 vs. the number of non-zero coefficients within each DCT block. The block non-zero content histogram is also shown. Power dissipation shows strong correlation vs. the non-zero coefficients because of the data-dependent processing algorithm. Conventional architectures on the other hand exhibit a virtually flat power dissipation profile vs. non-zero block content.

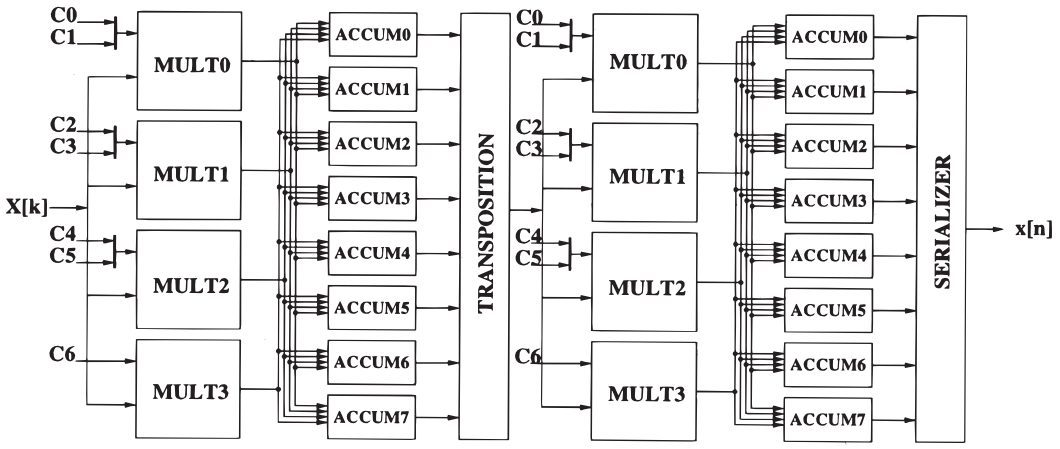

Fig. 8: IDCT Chip Block Diagram.

continued

This property enables the IDCT chip to tradeoff image quality and power. The average power dissipation for this data set is 4.65 mWatts at 1.3V, 14 MHz (5.68 NZ coefficients on average per block). The IDCT chip exhibits lower switched-capacitance per sample (factor of 2) from past IDCT chips when normalized for equal process feature size. The power performance of the chip improves significantly at lower bitrates (coarser quantization). This makes our approach ideal for emerging quality-on-demand compression protocols that trade-off power dissipation and video quality. ❑

Fig. 9: IDCT Chip Microphotograph.

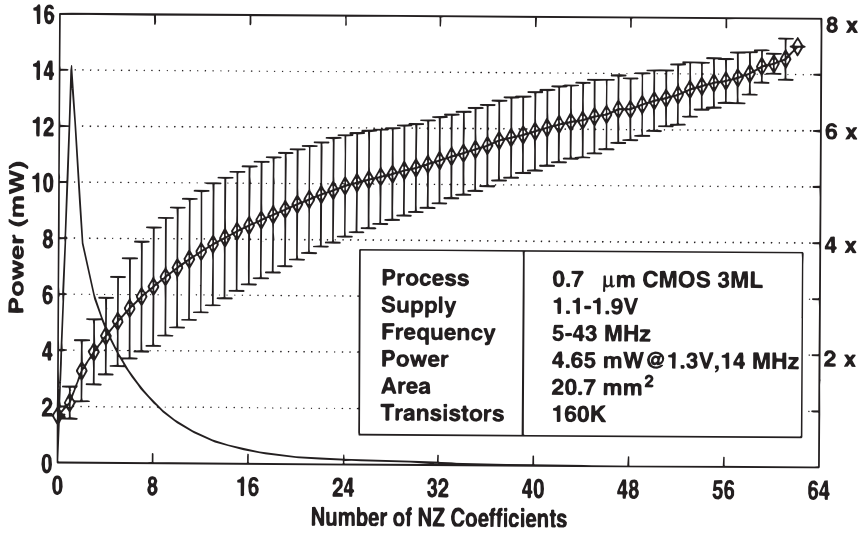

Fig. 10: IDCT Chip Measured Power Results at 1.32V, 14 MHz. This plot represent over 2.8 million separate power measurements.

An Ultra Low Power Variable Length Decoder for MPEG-2 Exploiting Codeword Statistics

Personnel

S.-H. Cho and T. Xanthopoulos
(A. Chandrakasan)

Sponsorship

SHARP Corporation

Variable length coding is a lossless compression method usually applied in video and image compression systems to further compress the data without degrading the image quality. It assigns frequent symbols to short codewords and infrequent symbols to long words and therefore achieves minimum average codeword length. In MPEG-2, variable length decoder is used to decompress the incoming bit stream so the data can be further decoded in the following modules such as inverse discrete cosine transform and motion compensation. As a part of the low power MPEG-2 project we have designed a variable length decoder which exploits the signal statistics of Variable Length Codes (VLCs) to reduce power. It uses fine grain table partitioning and VLC detector size reduction based on codeword frequency.

A variable length decoder can be decomposed into two parts, the VLC detector and the lookup table. The VLC detector discerns the variable length code from the bit stream and provides it to the lookup table where the output is produced. In conventional methods the lookup table is often implemented in a single table to achieve high throughput, which results in the majority of power dissipation. To reduce this power we decomposed the table into several non-uniform tables so that the energy required to decode a codeword depends on the probability. Frequent codewords are decoded in small lookup table (hence small energy), while less frequent codewords use larger tables. The average power consumption is minimized by selecting the optimum number of partitions. The size of the VLC detector is also reduced for least power consumption. By exploiting the codeword statistics that short codewords are frequent, the energy consumption is minimized. The optimized VLC detector achieves a factor of two power reduction while maintaining the same throughput compared to the conventional VLC detector whose size is determined by the maximum length VLC.

A chip was fabricated which decodes all 15 different variable length code tables for MPEG-2. It provides the IDCT module with 48 million samples/sec while consuming 0.53 mW. An order of magnitude power reduction was demonstrated compared to the previous approaches. ❑

Power	530 µW @ 1.35V
Technology	0.6 µm CMOS (Vtn = 0.67V, Vtp = -0.93V)
Area	5.24 mm^2 (core), 16.54 mm^2 (full)
Frequency	12.5 MHz
Throughput	7.57 M codewords/sec
Video Rate	48.75 M samples/sec

Table 4: Summary of Chip Test Results.

Fig. 11: Die Photo of Low Power VLD.

A Reconfigurable Dual Output Low Power Digital PWM Power Converter

Personnel
A. Dancy and J. Bretz
(A. Chandrakasan)

Sponsorship
DARPA and NSF Fellowship

Electronic circuits for mobile systems can be designed to operate over the range of the voltages supplied by the battery over its discharge cycle. However, adding some form of power regulation can significantly increase battery life, since it allows circuitry to operate at the optimal supply voltage from a power perspective. As the power dissipation of electronic circuits drop (exploiting low-voltage process technology and other power management techniques), there is a need for high-efficiency DC-DC conversion circuits to deliver low power levels.

This versatile power converter controller provides dual outputs at a fixed switching frequency and can regulate either output voltage or target system delay, using an external L-C filter. In the voltage regulation mode, the output voltage is monitored with an A/D converter, and the feedback compensation network is implemented digitally. The generation of the PWM signal is done with a hybrid delay line/counter approach, which saves power and area relative to previous implementations. Power devices are included on chip to create the two independently regulated output PWM signals. The key features of this design are its low power dissipation, reconfigurability, the delay or voltage feedback, and multiple outputs. Figure 12 gives an overview of the architecture.

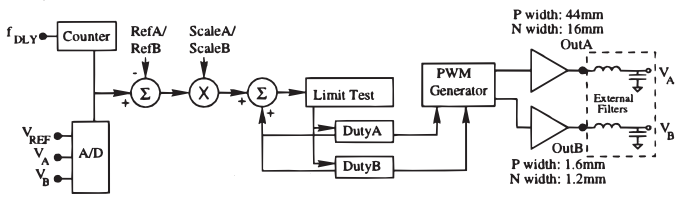

Fig. 12: Block diagram of converter.

Fig. 13: Die Photograph.

Table 5 gives the characteristics of the controller in two different configurations. Measured efficiencies were between 89% and 80% over a range of output currents, for the particular output filter selected. There is a trade off between the size and cost of the output filter and the achievable efficiency. The filter selected here represents a low cost, small area selection. The losses are dominated by a 9.5 Ω resistance in the output inductor at high output powers. At 500 kHz, with 256 levels of duty cycle resolution, the control circuit draws less than 45 μA.

continued

continued

At low output loads, efficiency goes down due to three sources of loss which are independent of load: control logic, output transistor gate drive, and ripple current through the series resistance of the filter inductor and transistor channel region. Work is currently being done to operate the converter in discontinuous mode, whereby the ripple current is reduced by turning off both output transistors when the current in the inductor reverses direction. This happens when the average output current is lower than the ripple current amplitude. Also, given this limit on the output current, the size of the output transistors can be optimized to reduce gate drive losses. ❑

Die Size	3.2mm x 2.8mm (test chip is pad limited)
Technology	$0.6\mu m$ DPDM
A/D INL	± 0.5 LSB
A/D DNL	+0.3, -0.4 LSB
Power: 1024 Taps, 500kHz	
Min. supply voltage	$2.05V$
PLL & Logic Current	$199.3\mu A$
Analog Circuits Current	$1.5\mu A$
Power: 256 Taps, 500kHz	
Min. supply voltage	$1.35V$
PLL & Logic Current	$42.8\mu A$
Analog Circuits Current	$1.5\mu A$
Output A Efficiency, f_{SW}=500kHz, V_{out} = 2.0V	
I_{out} =10mA	η =89%
I_{out} =45mA	η =80%
Output B Efficiency, f_{SW}=500kHz, V_{out} = 1.0V	
I_{out} = 750μA	η =89%
I_{out} =10mA	η =80%
Inductor Value	$220\mu H$
Capacitor Value	$0.22\mu F$
Filter Area	$0.024in^2$
Output Ripple	$< 40mV$ for $f_{SW} \geq$500kHz
Operating Frequency	< 2.5MHz

Table 5: Summary of Chip Data.

Reconfigurable Architectures for Power Efficient Data Encryption

Personnel

J. Goodman and A. Dancy
(A. Chandrakasan)

Sponsorship

DARPA

The secure transmission of multimedia information (e.g., voice and video) is critical in many wireless network applications. Wireless transmission imposes constraints not found in typical wired systems such as low power consumption, tolerance to high bit error rates, scalability, and limited data bandwidth.

The high bit error rates of wireless channels (e.g., ~10^{-2}) requires us to consider the effects of error propagation introduced by various encryption/decryption algorithms. Block ciphers multiply single bit errors to yield a higher effective BER as compared to stream ciphers. The increased BER would require the use of more complex error correction coding algorithms. This would introduce additional redundancy into the transmitted bit stream, which would in turn reduce the already-limited data bandwidth of the system.

We propose a scalable encryption approach which can be utilized to dynamically trade-off the amount of security provided vs. the energy expended to encrypt a bit. This enables us to exploit the fact that transmitted data streams have an inherent structure consisting of both high and low priority information. As a result, we can dynamically allocate the strength of the encryption used based on the underlying data priority, allowing energy to be conserved when strong encryption is not required.

Two different encryption engines have been proposed and implemented. The first is based on scalable Linear Feedback Shift Registers (LFSRs) structures. LFSR-based stream ciphers provide a very simple, and power efficient (on the order of 10's of μW) implementation at the cost of reduced security. This can be improved using such techniques as constantly reinitializing the cipher with pseudo-random seed values, and varying the feedback polynomials of the LFSRs. The second utilizes a much more secure (and computationally complex) algorithm known as the Quadratic Residue Generator (QRG). The QRG performs repeated modular squaring

continued

continued

operations to produce a cryptographically-secure pseudo random keystream that can then be XORed with a data stream to form an encrypted data stream. The security of such a system relies on the difficulty of factoring the modulus used into its constituent prime factors — a problem whose complexity grows sub-exponentially with the size of the modulus used.

We've fabricated a reconfigurable, variable width QRG has been fabricated and tested (Figure 14 and Table 6). The width, and thus the security, of the QRG can be reconfigured on the fly to vary from between 64 and 512 bits, in 64 bit increments. The QRG makes extensive use of clock gating and shutdown techniques, as well as self-

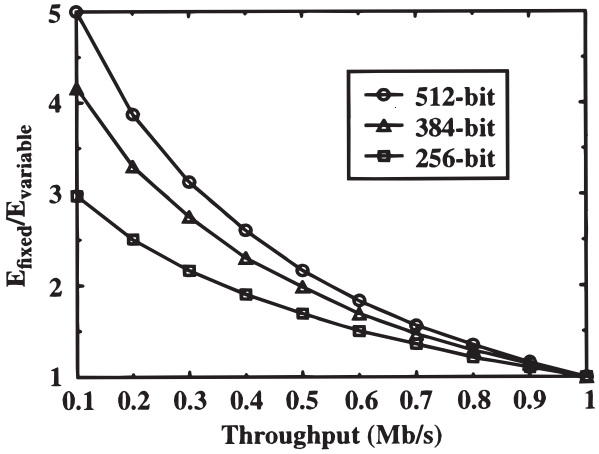

Fig. 15: Variable supply energy reduction for varying QRG width and throughput.

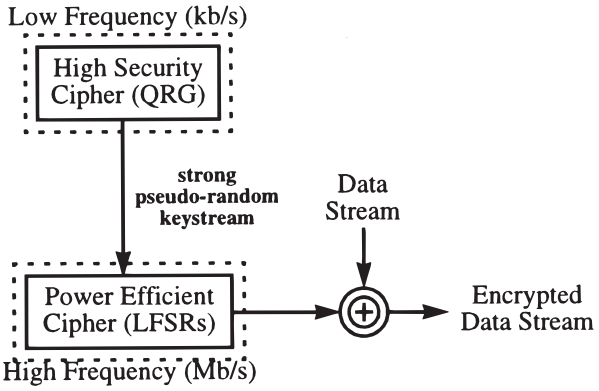

Fig. 16: Proposed hybrid encryption scheme.

Fig. 14: Reconfigurable QRG with embedded power supply die photo.

continued

timed gating in order to minimize the energy consumption of the circuitry. The power consumption of the QRG has been measured to be 134 mW at a width of 512 bits and data rate of 1 Mb/s.

The QRG also features an embedded variable output power supply that enables us to exploit time varying data rates and computation requirements to minimize the supply voltage and hence the energy dissipation. Variable supply techniques enable us to achieve significant energy reductions over conventional fixed-supply implementations (Figure 15).

Dimensions	6.2mm x 7mm
Device Count	260k
Process	0.6 mm DPDM
Threshold Voltages	V_{tP} = -0.88V, V_{tN} = 0.75V
Minimum Operating Voltage	1V
$P_{QRG-512}$ @ 1 Mbps (V_{dd} = 2.5V)	134mW
$P_{QRG-512}$ @ 18 kbps (V_{dd} = 1V)	360µW

Table 6: Implementation details of Reconfigurable QRG w/ embedded power supply.

A hybrid scheme (Figure 16) using a simple output LFSR cipher coupled with a much more secure and power intensive pseudo-random seed generator (QRG) has been proposed for ultra low power applications such as a wireless image sensor. In the hybrid system, only the low-power LFSR-based stream cipher operates at the data rate. The high energy dissipation of the QRG is offset by operating it in the background at very low supply voltages (~1V) and data rates (~3 kb/s) as a seed generator and polynomial selector. As a result the normally prohibitive power demands of the QRG can be reduced to just 60 µW. Based on these estimates, the power consumption of our hybrid encryption module has been estimated to be on the order of 100 µW operating at an encryption rate of 1 MB/s. ❑

Low Power Silicon Monolithic 1.8 GHz Voltage-Controlled Oscillator

Personnel
D. A. Hitko
(C. G. Sodini)

Sponsorship
DARPA

As highlighted in the MIT Wireless Transceiver Demo system, a frequency synthesizer capable of high data rate transmission at 1.8 GHz had be developed to operate within the low power confines of the MIT Wireless Sensor Project. The synthesizer architecture had been constructed to minimize the reliance upon power-intensive circuit components such as the RF mixers and high-rate digital-to-analog converters found in many digital transmitters. However, one important element that remains within this synthesizer is a Voltage-Controlled Oscillator (VCO) which generates the RF signal for transmission. As the VCO output spectrum is the transmitted spectrum, effort needs to be placed into minimizing the (phase) noise of the oscillator, from which the noise spectrum results, in addition to designing for operation within the low power constraints imposed by the sensor.

For implementation of this 1.8 GHz VCO, a production 25 GHz silicon bipolar process was chosen to be consistent with commercial wireless applications demanding low cost, high yield technology, and to provide the capability for integrating the entire frequency synthesizer into a single-chip BiCMOS solution. Bipolar devices are preferred for this low power application because silicon BJTs provide a lower noise figure at the frequencies of operation for a given level of power consumption than can be achieved in today's CMOS processes. As a result, an oscillator with the desired noise performance can be realized with less power using bipolar transistors.

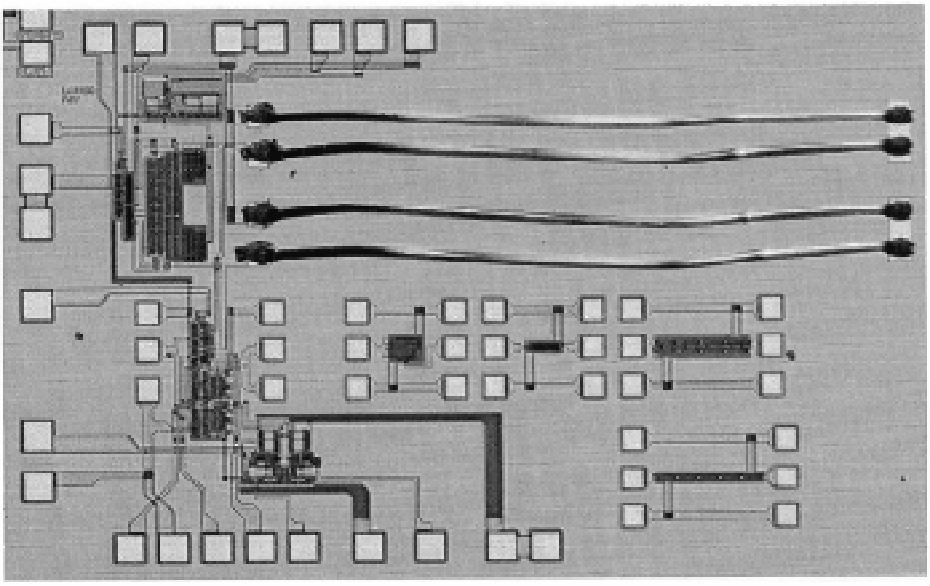

Fig. 17: Die Photo.

continued

continued

On-chip bond wire inductors exhibit less loss than planar spirals, and are used within the VCO to incorporate 2 nH inductances. Polysilicon to n+ capacitors complement varactors forged from the collector-base junctions of lateral PNP devices to complete the fully integrated resonator. The varactors provide for an oscillator tuning range of about 50 MHz.

Balanced VCO topologies were investigated to work with supply voltages down to 1V while maintaining good noise performance and DC to RF conversion efficiency. In addition to the improvements realized with the bond wire inductors, the oscillator phase noise spectrum is further reduced by optimizing the feedback loop around the VCO gain stage to provide for impedance matching. Circuits have been fabricated, and the oscillator has been verified to operate in fully packaged form with 5 mA drawn from a 1V supply, providing a measured phase noise spectral density of –94 dBc/Hz at an offset of 100 kHz from a –15 dBm carrier. A die photo of this VCO chip, with the bond wire inductors in place, is shown in Figure 17. ❑

Self-Powered Signal Processing

Personnel

R. Amirtharajah
(A. Chandrakasan)

Sponsorship

ARL Advanced Sensors Federated Lab and
NSF Fellowship

Portable systems that depend on batteries have a limited operating life and are prone to failure at inconvenient times. We propose a system that uses ambient energy as a power source for a DSP that processes sensor data.

The sources of ambient energy available to the portable or embedded system depend on the application. Examples include light or electromagnetic fields, thermal gradients, and fluid flow. In this study, we focus on a generator that transduces mechanical vibrations.

Figure 18 shows detailed block diagram of the system. The generator consists of a mass connected to a spring, the whole mounted within a rigid housing. As the housing is vibrated, the mass moves relative to the housing and energy is stored in the mass-spring system. A coil is mounted on the mass and moves through a radially directed magnetic field. The time-varying field creates a voltage on the coil and power can then be extracted from it. The system is damped by mechanical losses and the action of the electromechanical transducer. The output of the generator was transformed to a higher voltage that could be rectified by a diode onto a capacitor.

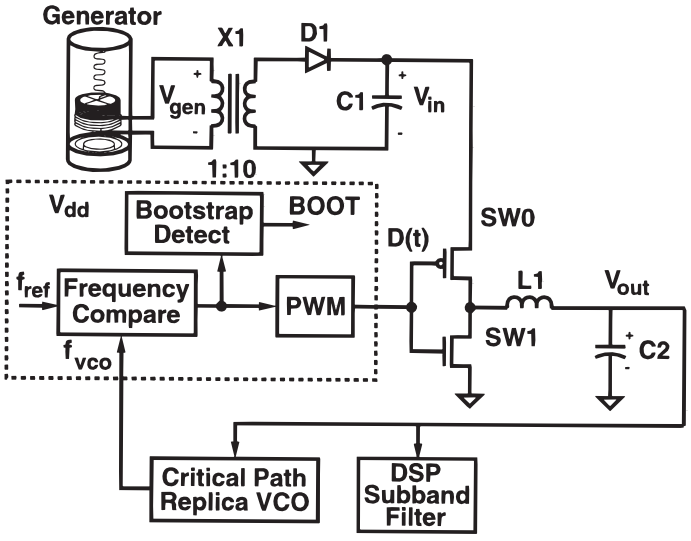

Fig. 18: System Feedback Loop.

continued

The DSP rate command, f_{clk}, is compared to the VCO output f_{vco}. The VCO is a ring consisting of the DSP adder critical path and is supplied by the regulated output voltage. The converter is a Buck converter with very small P and N FETs (1200 µm and 300 µm) controlled by a PWM waveform with 6 bit resolution. A new duty cycle is determined from the 2 bit frequency error. The modulated Vin signal is then passed through an LC lowpass filter, external to the chip, to produce Vout, the power supply for the DSP. The FIR filter is a well-known subband filter.

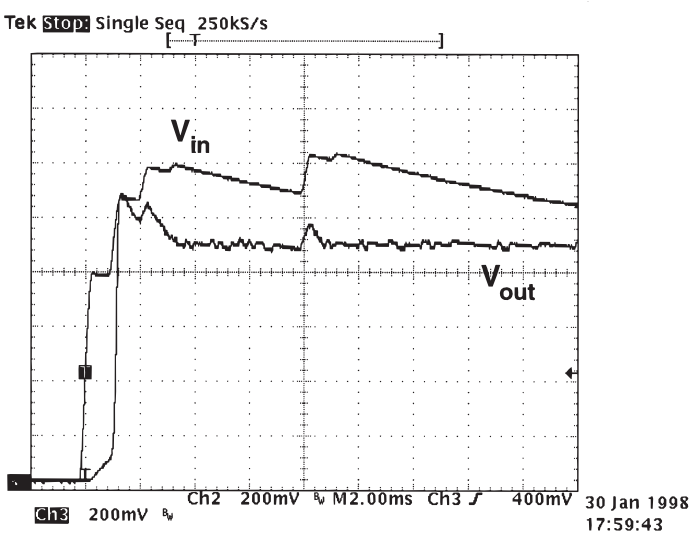

Fig. 19: Voltage Regulation.

Figure 19 shows the regulator loop response to the changing generator output voltage for a shock vibration. A die photo of the regulator is shown in Figure 20.

We are continuing this research by designing an ultra-low power DSP chip for biomedical sensor applications. This chip will demonstrate the feasibility of doing practical signal processing using self-powered techniques. ❑

Fig. 20: Chip Die Photo.

Low Power 1.8 GHz Frequency Synthesizer Capable of 2.5 Mbit/s Modulation

Personnel

D. A. Hitko, D. R. McMahill, N. R. Shnidman, and M. H. Perrott
(C. G. Sodini)

Sponsorship

DARPA

This research effort focuses on the development of a low power, wireless transmitter capable of digital modulation at data rates greater than 1 Mbit/s at a carrier frequency of 1.8 GHz. Our approach is architectural in nature, with the goal of achieving a design that allows complete integration of the transmitter except for the RF bandpass filter.

To achieve a low power solution, we have chosen a topology having a minimal number of components. We argue that any narrowband transmitter with good spectral characteristics must contain a frequency synthesizer to select the desired output frequency, and a transmit filter to shape the modulation spectrum. Our approach, therefore, is to directly modulate a frequency synthesizer with data that has been shaped by a digital transmit filter. This design removes the need for mixers and I/Q D/A converters found in other transmitter architectures.

The idea of directly modulating a frequency synthesizer is not new, and has been successfully achieved by other researchers using fractional-N synthesizers with noise shaping. Unfortunately, the bandwidth constraints on the modulation data imposed by the synthesizer dynamics are a strong deterrent to achieving data rates above 100 Kbit/s.

The contribution of this research is the proposal and verification of a compensation technique that allows over an order of magnitude increase in data rate when directly modulating a frequency synthesizer. This technique is quite simple – the digital transmit filter is modified by convolving it with the inverse of the PLL transfer function seen by the data. To verify the method, a custom 0.6 µm CMOS frequency synthesizer was built and shown to successfully perform GFSK modulation at data rates greater than 2.5 Mbit/s at a 1.8 GHz carrier frequency. Key points to the implementation are that the bandwidth of the synthesizer is 84 KHz, and its power consumption 27 mW.

A demonstration system has been constructed to showcase the CMOS synthesizer IC and the bipolar VCO IC. The demo system is made up of two self contained systems. The first system is a transmitter which incorporates the two IC's designed at MIT. The transmitter is able to accept an external 1.25 Mbps data stream or generate a PRBS-9 test sequence. In addition, an on board oversampled A/D converter provides a means for digitizing an analog signal for transmission. The pulse shaping and precompensation FIR filter has been implemented in a programmable logic device. A pair of digital LCD panel meters provide a real time display of the power consumption of the VCO and synthesizer IC's. The RF output frequency may be tuned to any of the 10 DECT channels in the 1.88 - 1.9 GHz range.

The other half of the demo system is a basestation which receives the transmitted RF signal and provides a serial data and clock at its output. In addition, an on board oversampled D/A converter is used to drive an audio amplifier and speaker. The audio output portion combined with the A/D in the transmitter allow the system to be easily showcased using only a portable CD player. ❑

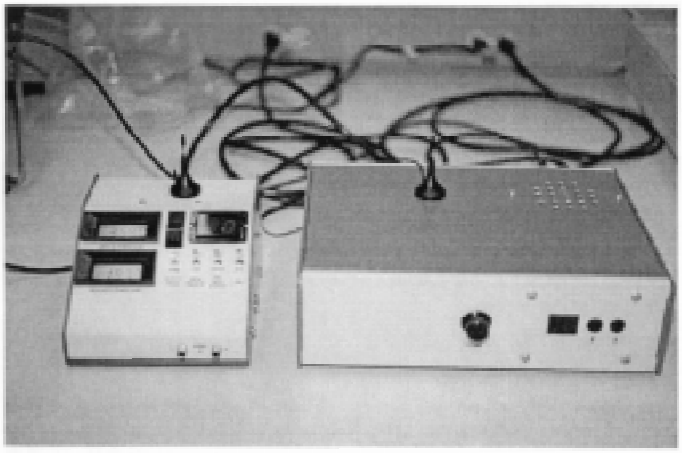

Fig. 21

Synthesis of Delta-Sigma Converters for MEMS Applications

Personnel

M. Shane Peng
(H.-S. Lee)

Sponsorship

Heinle Memorial Fund

With the increasing importance of quick generation of analog circuits, many analog circuit synthesis CAD tools have been designed and are in current development. However, most attempts have been very general, rendering the designs generated highly impractical. Instead of trying to encompass all applications possible, we have chosen to synthesize specific analog circuits for specific applications, thereby decreasing complexity and improving robustness.

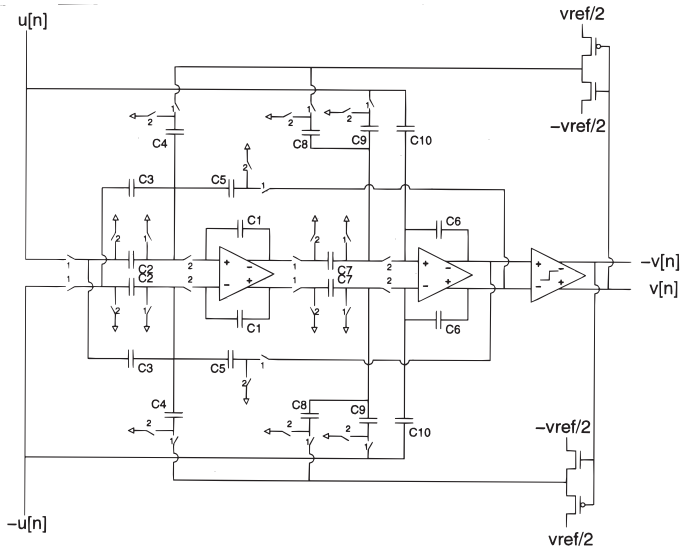

Fig. 22: Second Order Delta Sigma Modulator - Cascade of Integrators.

The focus of this research is the design and implementation of a general, low-power, oversampling, bandpass, one-bit delta-sigma analog to digital converter for small voltage measurements, particularly MEMs applications. This tool allows a user to enter parameters for the desired design (order of modulator, oversampling ratio, etc.) and a full netlist of capacitors, switches, operational amplifiers, and comparators will be generated. The generated design has dynamic range scaling and power optimized operational amplifiers, important characteristics of modern designs.

The current architecture that this tool utilizes is a cascade of integrators/resonators for either low-pass or bandpass delta-sigma architecture. The architecture is implemented as a discrete-time, fully differential design as shown in Figure 22. Switch capacitors are used which makes implementation of the integrators/resonators straightforward.

The behavioral circuit synthesis part of this project is based on the Delta Sigma Converter MATLAB toolbox developed by Dr. Richard Schreier at Oregon State University. ❑

Assessing Circuit-Level Hot-Carrier Reliability

Personnel
W. Jiang
(J. E. Chung)

Sponsorship
DARPA/ONR/ARO

Existing Hot-Carrier (HC) reliability criteria, based on device-performance metrics alone, are too ambiguous, resulting in excessive guard-banding, sacrificed device performance, and unnecessary process complexity. Although significant work has been carried out on hot-carrier degradation modeling and reliability circuit-simulation tool development, no consistent method yet exists as to how to assess hot-carrier lifetime in terms of more realistic circuit, rather than device, performance. What is lacking is a comprehensive understanding of the major factors that contribute to circuit-level hot-carrier reliability.

This project attempts to define and characterize the major factors that cumulatively determine circuit-level hot-carrier reliability and to determine each particular factor's relative importance. The inherent inverse relationship between lifetime-underestimation/criteria-overspecification and the amount of known device/circuit information will be explored. Lifetime-underestimation/criteria-overspecification is shown to depend quite strongly on the particular "worst-case" approximations used.

Major Factors Determining Hot-Carrier Reliability:
Digital circuit-level hot-carrier reliability can be expressed as a cumulative function of the following five major Factors:

I. Hot-Carrier Degradation Model Precision and Accuracy
II. The Specific MOSFET Terminal Voltage Waveforms
    A. Circuit Topology and Operating Conditions
    B. Capacitive-Coupling-Induced Voltage Overshoot
III. MOSFET Switching Activity
IV. Circuit Performance Sensitivity to Device Degradation
V. Relative Importance of the Degraded Circuit Path

Factors I - III determine both the magnitude and distribution of hot-carrier-induced oxide damage throughout the circuit. Factors IV and V determine how this created damage impacts overall circuit performance. Depending upon the particular technology or circuit design, each of the above reliability Factors can vary significantly in magnitude, having a direct impact on the estimated lifetime.

Assessing Hot-Carrier Lifetime Based on "Worst-Case" Approximations: The degree of lifetime-underestimation/criteria-overspecification strongly depends on the realism of each Factor's "worst-case" approximations. If insufficient data about a Factor exists, then more conservative "worst-case" approximations must be adopted, resulting in greater lifetime-underestimation. However, with sufficient data, more realistic "worst-case" approximations can be used, reducing lifetime-underestimation. Note, that acquiring detailed device/circuit information may not always be feasible (e.g. there is a circuit-size limit to practical SPICE-level simulation; higher-level simulation tools have reduced model accuracy). Thus, each reliability Factor possesses an inherent trade-off between its degree of realism and its ease of evaluation.

continued

continued

For Factors II-IV, Table 7 lists different "worst-case" approximations along with the corresponding required level of device/circuit information. Approximation (A) is the most conservative; (B) is more realistic, etc. Depending upon the level of approximation chosen for each Factor, a wide range of different lifetime values can be calculated. ❑

Reliability Factor	Different Levels of "Worst-Case" Approximation	Necessary Device/Circuit Data
II. The Specific MOSFET Terminal Voltage Waveforms	(A1) Determine device-level "worst-case" DC stress condition (A2) Apply this DC stress condition identically to all devices	(A) Device-level stress data
	(B1) Determine "worst-case" AC stress per switching-transition over circuit's design/operating space (B2) Apply this AC stress condition identically to all devices	(B) All of (A) plus AC hot-carrier reliability simulation
	(C1) Evaluate AC stress/transition over circuit's design/operating space (C2) Back-annotate each device with its specific AC stress condition	(C) All of (B) plus circuit back-annotation of AC stress information
III. MOSFET Switching Activity	(A1) DC stress condition implies 100% duty cycle	(A) None
	(B1) Use operating clock frequency	(B) Basic circuit operating condition
	(C1) Determine "worst-case" circuit switching activity value (C2) Apply this switching activity value identically to all devices	(C) All of (B) plus switching activity evaluation
	(D1) Evaluate circuit switching activity distribution (D2) Back-annotate each device with its specific switching activity	(D) (C) plus circuit back-annotation of specific switching activity
IV. Circuit Performance Sensitivity to Device Degradation	(A1) Evaluate device degradation over range of possible stress conditions determined from Factors I-III (A2) Simulate performance sensitivity using "worst-case" age for all devices	(A) Parameter extraction for stressed devices and circuit performance simulation
	(B1) Evaluate device degradation over range of possible stress conditions determined from Factors I-III (B2) Back-annotate each device with its specific degraded I-V characteristic (B3) Simulate circuit performance sensitivity	(B) All of (A) plus circuit back-annotation of degraded device characteristics and circuit performance simulation

Table 7

Monolithic Optoelectronics and Electronics on Si

Personnel
S. Ting, M. Bulsara, and M. Currie
(E. A. Fitzgerald)

Sponsorship
Discovery Semiconductors and Lincoln Laboratories

We are fabricating optoelectronic circuits on Si using the MIT MTL CMOS baseline process in conjunction with our III-V OMVPE growth. Currently, process integration of optoelectronic chips is being investigated by employing direct III-V epitaxy on Si. Optoelectronic chips will be analyzed for reliability and performance. ❑

Circuit Techniques for Multi-Threshold CMOS Technology

Personnel
J. Kao
(A. Chandrakasan)

Sponsorship
DARPA

Low power circuit operation is becoming extremely important for many electronic systems. In portable equipment (PDAs), long battery life is requisite, while in high performance systems (desktop microprocessors), managing (reducing) heat dissipation is also extremely important. The best way to lower power consumption is to reduce dynamic dissipation by lowering the power supply; however, to maintain performance, the threshold voltage must also be scaled, which exponentially increases leakage power. One approach to reducing power in burst mode or event-driven applications is to use an emerging technology like Multi-Threshold CMOS (MTCMOS) to allow circuits to operate at low threshold voltages (high speed) during computation and at high threshold voltages (low leakage) during idle periods.

MTCMOS circuits involve placing a high Vt sleep device in series with low Vt logic so that when the sleep transistor is on, the internal logic switches quickly, but when the sleep transistor is turned off, leakage is reduced (Figure 23).

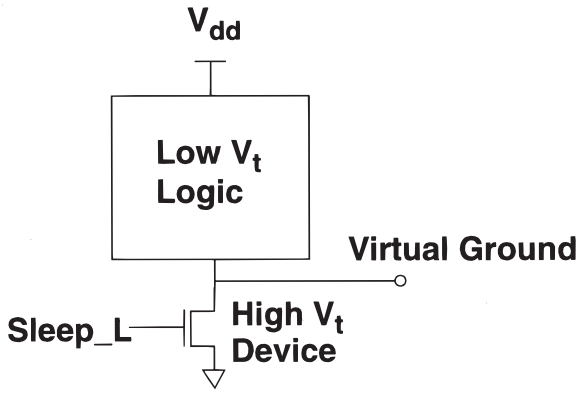

Fig. 23: Multi Threshold CMOS.

continued

continued

MTCMOS circuits are difficult to size though because the worst case delay through a logic block depends greatly on the discharge pattern of current that must flow through the sleep transistor. The high Vt device (when on) is very close to a linear resistor, and any discharge current passing through the device will cause the virtual ground node to rise, causing internal gates to slow down. As a result, the input pattern will strongly influence the delays through the logic block in a complicated fashion.

One approach to sizing the sleep transistor is to independently provide a sleep transistor for every single gate in the circuit. By making the transistors large enough, circuit performance will be maintained, independent of the input pattern applied, although the area penalty will be very large. A hierarchical sizing strategy has been proposed that will help reduce the number of these individual sleep transistors through merging of mutually exclusive gates. Gates that do not switch at the same time can share a common sleep transistor, and the total area for sleep transistors can be reduced dramatically. This algorithm for merging sleep transistors can be applied to all hierarchical levels including individual gates, cells within an array, and modules within a chip.

Another important issue in designing MTCMOS circuits is to ensure correct functionality for memory structures. For an ordinary MTCMOS implementation of a latch, where power and ground are disconnected from the circuit, it will not be possible to hold state during the sleep mode. We are currently developing novel circuits that will retain state with low leakage during sleep modes, yet will still provide very high performance during active modes. Work is being done to develop a complete methodology for designing systems in the MTCMOS circuit style. ❑

Electromigration in Single Crystal Interconnects

Personnel

V. T. Srikar
(C. V. Thompson)

Sponsorship

SRC

Electromigration-induced failures of interconnects with bamboo or near-bamboo microstructures is affected by transgranular diffusion and failure mechanisms. In order to isolate transgranular failure mechanisms for characterization through accelerated tests, we have developed techniques for producing bi-crystal and single crystal Al films on oxidized silicon wafers. We have used these techniques to experimentally investigate the rates and mechanisms of failure in pure Al single crystal interconnects with different textures and in-plane orientations. Kinetic results are being used in simulations of electromigration-induced failure in bamboo and near-bamboo lines. We are currently carrying out experiments on Cu-doped single crystal Al interconnects and single crystal Al with Al_3Ti overlayers to investigate the effects of these factors on the kinetics and mechanisms of transgranular electromigration and electromigration-induced failure. ❑

The Effects of Thermal History on the Stress and Reliability of IC Interconnects

Personnel
V. T. Srikar, A. Gouldstone, and S. Suresh
(C. V. Thompson)

Sponsorship
SRC

The metal films used for IC interconnects are deposited at elevated temperature, patterned at room temperature and then passivated by encapsulation with glasses deposited at elevated temperature (350 to 400° C), before they are used at temperatures around 100° C. This thermal history affects the stress state of the interconnects and ultimately affects their reliability, as limited by stress or electromigration-induced voiding. The stress state is not only a function of the details of the thermal history, but is also a function of the interconnect thickness, width, and spacing, as well as of the thickness and mechanical properties of the passivation layers. We have used a generalized plane strain formulation to predict the stress state as a function of these parameters, allowing for both elastic and plastic accommodation of the thermal strain.

We are testing these models through comparisons with experiments in which the radii of curvature of wafers coated with Al and Al-alloy interconnects with different widths, thicknesses and spacings are measured as the wafers are thermally cycled. Results from experimentally validated stress simulations will be incorporated into our electromigration simulation tools. ❑

Three-Dimensional Integration for Future ULSI

Personnel
A. Rahman and A. Fan
(J. Chung and R. Reif)

Sponsorship
SRC

It is well known that interconnects will play a significant role in future generation of high performance microprocessors. The performance and cost of future ULSI technology will be determined by the interconnect delay and process the complexity for fabricating sub-micron (~ .1 μm) devices with 7 - 8 layers of interconnects. Building high-speed (fc > 1 GHz) integrated circuits will require innovations in on-chip wiring networks and device designs. Cu and low k dielectric materials will enable us to scale the interconnects along with devices to maintain high transistor density and system performance as guidelined in the SIA roadmap. However, even with the addition of Cu and low k dielectrics, the interconnect bottleneck will not be completely alleviated. Novel structures, such as optical and/or three-dimensional (3-D) interconnects may be required to meet the performance requirements for future ULSI technology. The advantages of the 3-D architecture include reduced die size, high device density, increased functionality, and higher performance.

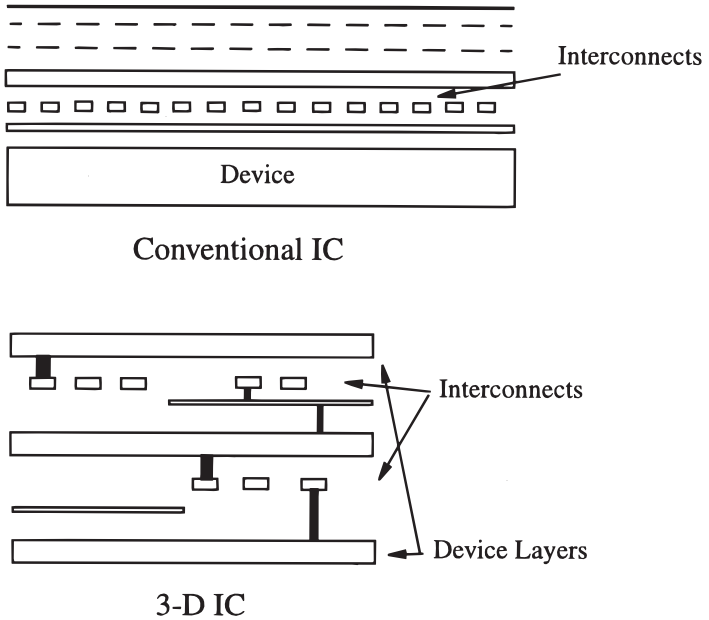

Fig. 24

continued

continued

In Figure 24 conventional and 3-D circuit implementations are illustrated. Metal-device layers are stacked to form the 3-D circuits. In our research project we will explore the feasibility of 3-D integration and evaluate the performance limits of 3-D architecture. Using wire-length distribution, optimal chip area, clock frequency and interconnect architecture in 3-D integration will be derived. Key technologies such as low temperature wafer bonding or epitaxial overgrowth that will enable us to fabricate 3-D circuits will be identified. As a technology demonstration, 3-D implementation of some building blocks of high performance circuits will be fabricated. ❑

Intelligent Train Control Systems

Personnel

K. Takeshi
(I. Masaki and J. Sussman)

Sponsorship

East Japan Railway

System architecture is restricted by available component technologies. Today's fundamental train control system architecture is based on the component technologies which were available over 100 years ago. Examples of those component technologies are communication systems which use rails and track circuits for locating trains. We are developing a novel train control scheme with current components such as machine vision chips, global positioning systems, and cellular phones.

The other motivation of this project is to reduce the cost of trains by using emerging components for automobiles. Examples include adaptive cruise control systems, collision warning systems, dynamic navigation systems, platooning systems, advanced behavior-control processor chips, and in-vehicle communication network chips. The cost can be typically reduced by one order of magnitude by replacing custom-made components dedicated to trains with mass-produced automotive components. In many cases, however, it is impossible to simply replace existing train components with automotive ones because the reliability specifications of train components are higher. A technical challenge is to develop train-control architecture which accepts less reliable components. Technologies for developing reliable systems with unreliable components are not limited to redundant architecture. Various fault tolerant architectures are being explored.

An example of possible architecture includes global positioning systems and odometers for locating trains, machine vision systems for monitoring, and hierarchical wireless communication network for communications among control centers, trains, on-ground facilities such as rail-crossing, and various components on trains. An emphasize is on concurrent development from the following two different viewpoints: components and system design. ❑

A Framework for Distributed Web-based Microsystem Design

Personnel

N. B. Gangadhar Konduri and D. Saha
(A. Chandrakasan and D. A. Antoniadis)

Sponsorship

DARPA

The design of future high-performance "system-on-a-chip" will require a distributed and collaborative design and verification methodology due to the diverse expertise required at various levels of abstraction. The objective of this project is to develop the infrastructure required to implement distributed and collaborative design tools. The simple user interface and global availability of the World Wide Web makes it an ideal platform for distributed design. Emergence of Java also enhances the power of Web-based distributed applications.

We have developed a hierarchical schematic editor, WebTop, which has a simple GUI. The tool has a number of primitive cells with which a designer can build other cells. WebTop also supports block editing, in which the user creates a template of the cell and provides a behavioral and/or structural view (e.g., Verilog); the new cell can then be used within other designs.

Providing access to remote tools is one objective of this project. For example, PowerPlay is a web-based tool developed at U.C. Berkeley which helps in the exploration of the system design space. From the WebTop editor, the designer can extract a PowerPlay compatible netlist and submit it to the PowerPlay tool to obtain power estimates of the system. WebTop is also capable of extracting a Verilog netlist and tools such as Pythia (a tool developed at MIT) can be used to estimate power dissipation at an RTL level.

In addition to estimation tools, it is desirable to invoke module generators that can generate low-level circuits/logic from high-level design descriptions. To illustrate this capability, a generator for an adaptive power supply was developed in WebTop. Based on user inputs of load current and voltages, several aspects of the design are generated (e.g., buffer sizes, power transistor sizes, filter inductor/capacitor values, etc.). We used Java RMI mechanisms to perform this function.

We are currently involved in a multi-university project, that includes U.C. Berkeley, Stanford and NCSU, to demonstrate concepts of distributed system design. As a driver application, we have chosen a single-chip multimedia processor that includes an embedded ARM core, a video compression module and power supply circuits. The design has been entered in WebTop and is currently being validated using a distributed simulation strategy. ❑

A Nyquist-rate Pipelined Oversampling Delta-Sigma A/D Converter

Personnel

S. Paul and J. Goodrich
(H.-S. Lee)

Sponsorship

NSF Fellowship/Lincoln Laboratory

A primary objective of the pipelined oversampling A/D concept is to circumvent the trade-off between accuracy and speed, inherent in conventional oversampling techniques, so that both high accuracy and high speed are simultaneously achievable. The concept is expected to provide a substantial improvement in bandwidth over conventional oversampling techniques. Estimated performance for 1.2 um CCD/CMOS technology is 14 bits of accuracy with Nyquist input sampling rates of 50 MHz.

The sampling rate of conventional oversampling architectures is limited by the fact that the modulator loop must operate over many clock cycles to produce a single digital output word. Their additional resolution in amplitude is achieved at the expense of resolution in time. Pipelining may be applied to eliminate this bottleneck. The modulator loop may be unraveled into a pipeline so that consecutive cycles of its operation occur in consecutive pipeline stages. The entire pipeline may then be operated at the input sampling rate and the device's throughput is one result per clock cycle, even though the entire result requires many stages to complete. In this way, a pipelined oversampling converter is able to achieve the accuracy benefits of oversampling without the corresponding detrimental impact on speed. The pipeline length determines the oversampling ratio and the pipeline clock rate determines the input bandwidth. There is no longer a trade-off between these specifications and each may be independently adjusted.

The decimator is pipelined in parallel with the analog signals and has an impulse response length equal to the number of pipeline stages. Because the input signal is sampled once and processed identically along the pipeline, the input signal, as viewed by the modulator's feedback loop, is constant. This simplifies design of the digital decimation filter because a sharp cutoff between the pass and stop bands is not needed.

The pipelined oversampling concept is not fundamentally tied to CCD technology. However, CCD devices are exceptionally well suited for implementing such a converter because high accuracy, low power pipelines with hundreds of stages are feasible. In contrast, CMOS pipeline lengths are limited by inaccuracies in, as well as hardware and power requirements of, the circuits at each stage.

Power requirements for a pipelined oversampling converter are expected to be significantly less than those of conventional converters operating at the same input sampling rate. A pipelined architecture is operated at an internal clock rate which is orders of magnitude slower than that of conventional converters and may use low speed, low power circuits. CCD charge-domain circuits require substantially less power than voltage-domain circuits. The only source of power consumption in a CCD device is dynamic switching current, which is needed during charge transfers. An additional advantage is achieved when the proposed converter is used to process charge quantity inputs. In this case, the charge packets may be converted directly to digital, in the charge domain. This approach reduces system power and complexity by eliminating the intermediate charge-to-voltage conversion, correlated double sampling, and resampling operations, which are otherwise needed. ❑

Oversampled Pipeline A/D Converters with Mismatch Shaping

Personnel
A. Shabra
(H.-S. Lee)

Sponsorship
IBM

In recent years, delta-sigma modulators and pipeline converters have become very popular as analog-to-digital converters. In comparing these converters for wide-band signals, we recognize a few important attributes. Due to the wide bandwidth of the input signal and limited circuit speed, delta-sigma converters afford only low oversampling ratios, which makes high-resolution conversion extremely difficult. The low oversampling ratio generally nullifies the primary advantage of delta-sigma converters; the tolerance to component mismatches. As a practical matter, at input frequencies over 50 MHz, it would be extremely difficult to achieve oversampling ratio greater than 8 with final accuracy over 12 bits with present technologies. At this low oversampling ratio, not only the benefits of delta-sigma modulation disappears, but more importantly many delta-sigma converters are incapable of providing good enough performance.

We believe that a more efficient approach would be to oversample a standard pipeline converter, and shape the mismatch out of band which will be removed by a subsequent digital filter. Since no attempt is made to shape the quantization noise, there is none of the concerns associated with delta- sigma converters with a low oversampling ratio.

This work applies mismatch shaping to a Commutative Feedback Capacitor Scheme (CFCS) pipeline converter proposed in. The SNDR improvement is achieved through a combination of oversampling and mismatch shaping, which modulates the distortion energy out-of-band. Simulation results show that at an oversampling ratio of 4 and 64 and component mismatch of 0.1% a 9 dB and 35 dB improvement in SNDR is achieved respectively, compared to a converter with no mismatch shaping. ❑

Oversampling A/D Converter Based on Superconducting Electronics

Personnel
J. F. Bulzacchelli
(H.-S. Lee and M. Ketchen - IBM)

Sponsorship
IBM Cooperative Fellowship and Consortium for Superconducting Electronics

Because of the high speed of Josephson junctions, superconducting electronics offers the possibility of producing A/D converters which achieve higher speed and resolution than do converters based on semiconductor technologies. In this program, we are investigating a Josephson delta-sigma converter which achieves high resolution by oversampling and digital filtering; the very high sampling rates available in superconducting technology make such a tradeoff between speed and resolution attractive. Based on published clock rates for Josephson Single-Flux-Quantum (SFQ) logic, a 20 GHz sampling frequency should be feasible.

While such a high clock rate enhances the performance of oversampled A/D converters, the challenges of high speed testing in a cryogenic environment are formidable. Even in the best cryogenic sample holders, the long cables used to connect the superconducting chip to room-temperature electronics have significant losses at frequencies above 10 GHz. In our previous reports, we described an optoelectronic clocking technique, which bypasses the bandwidth limitations of conventional electrical testing. In this approach, picosecond optical pulses from a mode-locked laser are delivered (via optical fiber) to an on-chip photodetector, which generates the clock pulses needed by the Josephson circuitry. We have already demonstrated our optoelectronic clocking system up to 20 GHz and will use it in testing our A/D converter later this year.

While the high speed clocking will be done optoelectronically, the digital output of the A/D converter will still be read out over standard coaxial cables. Consequently, direct transfer of the output of the delta-sigma modulator (at 20 Gbits/s) to the room-temperature test equipment is impractical in our setup. Instead, on-chip processing of the data will be used to reduce the bandwidth requirements for readout. In principle, an on-chip decimator could be used for this purpose, but the circuit complexity is high, and the data reduction that

continued

can be achieved is limited by the desired resolution at the Nyquist rate (e.g., 12 bits at 80 MHz, still close to 1 Gbit/second). Another approach would be to capture part of the modulator bit stream with an on-chip shift register, which could later be unloaded to room-temperature electronics at low speed. In order to evaluate the performance of the delta-sigma converter, the length of the shift register must be significantly larger than the oversampling ratio (e.g., 256). Unfortunately, even the largest (1024 bits) SFQ shift register demonstrated to date is too small for our purposes.

Our approach to the readout problem is based on the following observation. Since the power spectrum is the Fourier transform of a signal's autocorrelation function, estimating the power spectrum with fine frequency resolution is equivalent to estimating the autocorrelation function R[n] up to a large value of n. If, as mentioned above, a segment of the bit stream is captured with a single shift register, R[n] can be estimated only for values of n up to the shift register length. On the other hand, if two segments of the bit stream are captured with a pair of shift registers, as illustrated in Figure 25, R[n] can be estimated for values of n much larger than the shift register lengths. For the example shown in the figure, in which each shift register is 128 bits long, cross-correlation between the A and B segments generates estimates of R[n] for values of n between N + 1 and N + 255, where N is the number of bits skipped between acquiring the A and B segments and is set by an on-chip timing controller. By making N programmable from 0 to 8000, for instance, different sections of R[n] can be estimated through successive measurements, after which the entire function R[n] (up to n = 8255) can be assembled. Fourier transforming R[n] then yields a power spectrum with frequency resolution comparable to an 8K FFT of the original bit stream. An advantage of the technique over using an on-chip decimator is that one can observe the entire output spectrum, from dc to half the clock rate.

A simple and fast on-chip timing controller is a key element in making this technique practical. We have recently completed the design and layout of a timing controller employing 440 Josephson junctions. By setting 11 external control currents, the value of N is digitally programmable from 0 to over 8000 (in increments of 4). According to simulations, the circuit is much (4X) faster than necessary for testing the A/D converter at 20 GHz. We are currently completing the layout of the entire A/D converter test chip, which will be fabricated by HYPRES, Inc. ❑

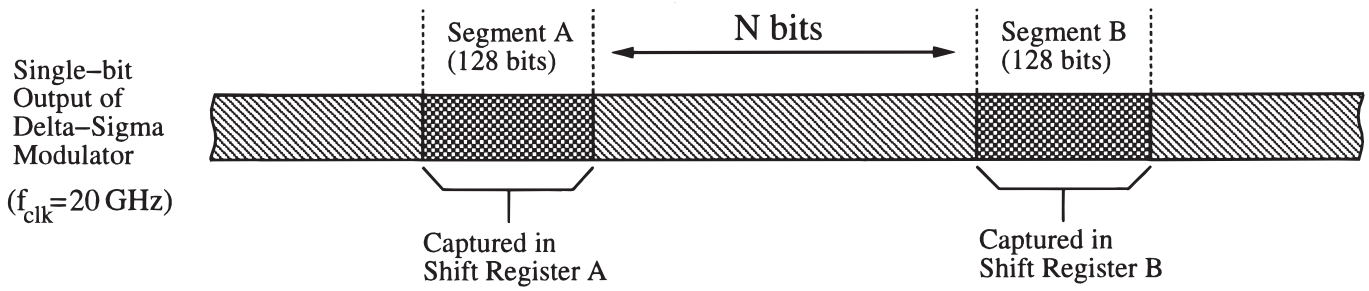

Fig. 25