A 175-mV Multiply-Accumulate Unit Using an Adaptive Supply Voltage and Body Bias Architecture

James T. Kao, Masayuki Miyazaki, *Member, IEEE*, and Anantha P. Chandrakasan, *Senior Member, IEEE*

***Abstract*—In order to minimize total active power consumption in digital circuits, one must take into account subthreshold leakage currents that grow exponentially as technology scales. This research develops a theoretical model to predict how dynamic power and subthreshold power must be balanced to give an optimal $V_{DD} - V_t$ operating point that minimizes total active power consumption for different workload and operating conditions. A 175-mV multiply-accumulate test chip using a triple-well technology with tunable supply and body bias values is measured to experimentally verify the tradeoffs between the various sources of power. The test chip shows that there is an optimum $V_{DD} - V_t$ operating point, although it differs from the theoretical limit because of excessive forward bias currents. Finally, we propose a preliminary automatic supply and body biasing architecture (ASB) that automatically configures a circuit to operate with the lowest possible active power consumption.**

***Index Terms*—Active power, adaptive body biasing, body biasing, dynamic power, dynamic voltage scaling, low power, subthreshold leakage, triple well.**

I. Background

LOW power considerations are becoming increasingly important in modern integrated circuits. As portable battery-powered devices such as cell phones, pagers, PDAs, and portable computers become more complex and prevalent, the demand for increased battery life will require designers to seek out new technologies and circuit techniques to maintain high performance and long operational lifetimes [1]. Power in modern digital CMOS integrated circuits has traditionally been dominated by dynamic switching power, given by

$$P_{\text{dynamic}} = C_{\text{switched}} {}^{*} V_{DD}^{2} {}^{*} f_{clk} \tag{1}$$

where C_{switched} is the total effective switched capacitance, V_{DD} is the supply voltage, and f_{clk} is the switching frequency. However, as technology scales leakage currents become increasingly large and must be taken into account to minimize total power consumption. Leakage currents can be broken into various components such as PN junction reverse bias current, gate-induced drain leakage, oxide tunneling, and hot carrier injection, but the dominant component is subthreshold leakage current given by [2]

$$I_{\text{leakage}} = I_0 \exp\left(\frac{V_G - V_S - V_{T0} - \gamma V_S + \eta V_{DS}}{n V_{\text{th}}}\right) \cdot \left(1 - \exp\frac{-V_{DS}}{V_{th}}\right) \tag{2}$$

where V_{th} is the thermal voltage, n is the subthreshold swing coefficient constant, γ is the linearized body effect coefficient, η is the DIBL coefficient, and I_0 a constant. Assuming that $V_{ds} \gg V_{\text{th}}$ and that the threshold voltage term includes barrier lowering and body effect terms, (2) can be further simplified to the well-known expression

$$I_{\text{leakage}} = I_0 10^{(V_{\text{gs}} - V_t)/S} \tag{3}$$

where $S = nV_{\text{th}} \ln 10$ is the subthreshold slope. For a typical technology with a subthreshold slope of 100 mV/decade, each 100-mV decrease in V_t will cause an order of magnitude increase in leakage currents. For extremely low $V_{DD} - V_t$ operating points, leakage power can actually dominate dynamic switching power. Fig. 1 shows a power trend for microprocessors and illustrates how leakage currents are projected to increase exponentially as technology scales.

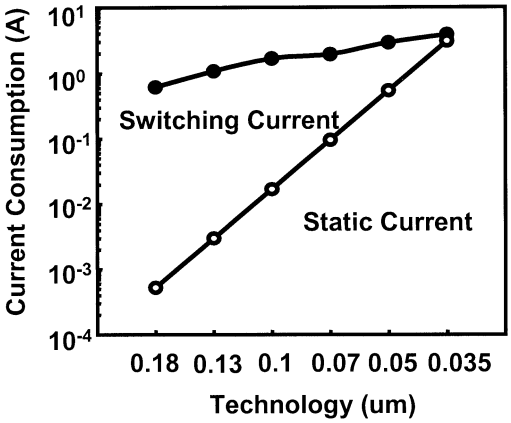

Fig. 1. Microprocessor trends from 1999 ITRS roadmap.

Reducing subthreshold leakage currents has been an important area of research over the past few years. Of particular value are techniques to reduce *standby* leakage currents, where circuits are placed in an ultralow leakage state during idle periods in order to lower power. Techniques such as MTCMOS, VTCMOS, and stack effect can all be used to effectively raise the V_t of the circuit during standby modes to reduce idle power, yet provide low V_t operation during the active modes for high performance [3]–[7].

Manuscript received April 8, 2002; revised June 10, 2002. This work was supported in part by the DARPA Power Aware Computing/Communication Program.

J. T. Kao and A. P. Chandrakasan are with the Massachusetts Institute of Technology, Microsystems Technology Laboratory, Cambridge, MA 02139 USA (e-mail: anantha@mtl.mit.edu).

M. Miyazaki is with the Massachusetts Institute of Technology, Microsystems Technology Laboratory, Cambridge, MA 02139 USA. He is also with the Hitachi Ltd., Central Research Laboratory, Tokyo 185-8601, Japan.

Digital Object Identifier 10.1109/JSSC.2002.803957

0018-9200/02$17.00 © 2002 IEEE

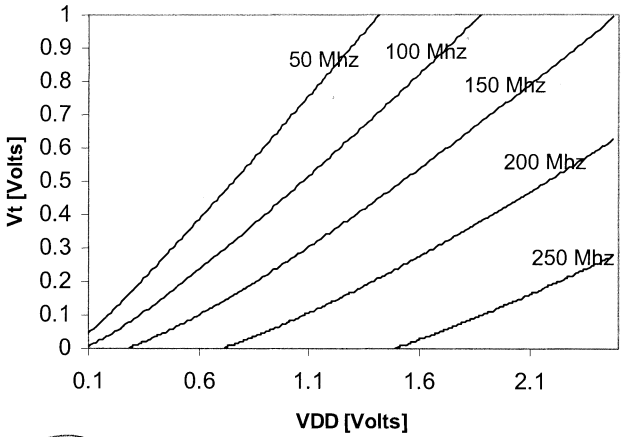

Fig. 2. V_{DD}/V_t combinations to give constant performance.

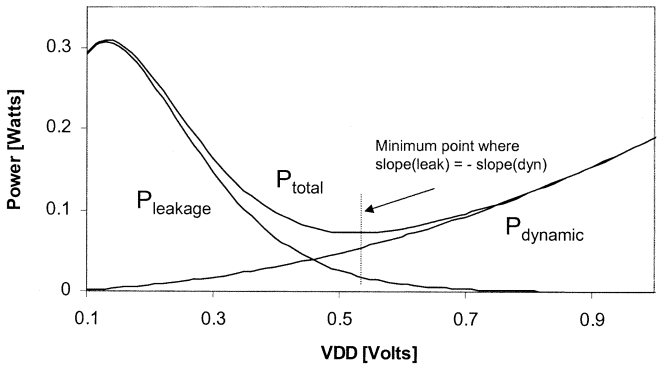

Fig. 3. Dynamic and subthreshold leakage power components for a fixed operating frequency (V_t implicitly set).

However, very little research has been devoted to managing leakage currents during the *active* state, which becomes more important as technology scales. This paper develops a novel technique to minimize total active power in digital circuits by dynamically adjusting both supply voltage and threshold voltages based on circuit operating conditions such as temperature, workload, or circuit architecture. It has been understood that active power can be balanced between dynamic and leakage power components, but this research is the first to our knowledge that explores how this optimum operating point changes with circuit operating conditions. This paper first develops a theoretical model to predict how the minimum active power point depends on circuit parameters and then describes experimental results for a test chip to verify the tradeoff between dynamic and leakage currents as workload changes. Finally, an automatic supply and body biasing architecture is proposed that will automatically bias a digital circuit to operate at its minimum active power level [8], [9].

II. Optimization of Supply and Threshold Voltage

For a digital circuit, it is possible to tradeoff dynamic and subthreshold leakage power by balancing between V_{DD} and V_t to maintain performance [10], [11]. Fig. 2 shows theoretical curves of V_{DD}/V_t combinations that give constant performance based on circuit parameters extrapolated from an SH4 microprocessor and models derived in the next section. As supply voltages increase, dynamic power increases quadratically, but, as threshold voltages decrease, leakage currents increase exponentially. Fig. 3 shows theoretical curves for power versus supply voltage, where the V_t is implicitly adjusted to maintain a fixed frequency. The total power is the sum of the dynamic power curve and the leakage power curve and the minimum operating power corresponds to the V_{DD} choice where the slope of the two curves are opposite in sign but equal in magnitude.

A. Model to Minimize Total Active Power Consumption

To better characterize how the minimum operating point varies with circuit parameters, it is useful to develop a simple model for total active power consumption taking into account both dynamic and leakage power. Some theoretical work was done in [12] to mathematically compute minimum energy–delay products for digital circuits, but for our purposes it is more appropriate to compute the actual average power as the metric of interest. The goal is to minimize the power consumed during the active operational period of a circuit, where the delay requirement is already fixed. The propagation delay for a CMOS gate can be modeled as

$$T_{pd} = \frac{KV_{DD}}{(V_{DD} - Vt)^{\alpha}} \tag{4}$$

where K is a constant of proportionality and α is the velocity saturation term that models short channel effects. This equation, for a given performance requirement, defines a locus of V_{DD} and V_t combinations that will exactly satisfy the overall performance. Assuming that the critical path delay (between registers for example) has a logic depth of n gates, then the operating frequency can be represented as

$$f \propto \frac{1}{nT_{pd}}. \tag{5}$$

The total power dissipation, subject to the above frequency requirement, is given by the sum of switching dynamic power and subthreshold leakage power given by

$$P_{\text{total}} = \underbrace{C_{\text{eff}} V_{DD}^2 f}_{P_{\text{DYNAMIC}}} + \underbrace{V_{DD} I_0 \exp\left(\frac{-V_t}{S}\ln(10)\right)}_{P_{\text{LEAKAGE}}}. \tag{6}$$

The minimum V_{DD}/V_t operating point can be mathematically derived from the above equations. One straightforward way to do this is with the Lagrange multiplier technique where the power in (6) is minimized subject to the constraint from (5). A more direct approach is to simply use direct substitution of (5) into (6) to give

$$P_{\text{total}} = C_{\text{eff}} V_{DD}^2 f + V_{DD} I_0 \exp\left(\frac{\sqrt[\alpha]{nfkV_{DD}} - V_{DD}}{S}\ln(10)\right) \tag{7}$$

which can be differentiated with respect to V_{DD} and set to 0 to compute the extrema. The resultant expression can then be solved numerically for the optimal V_{DD} and V_t combination that minimizes overall power.

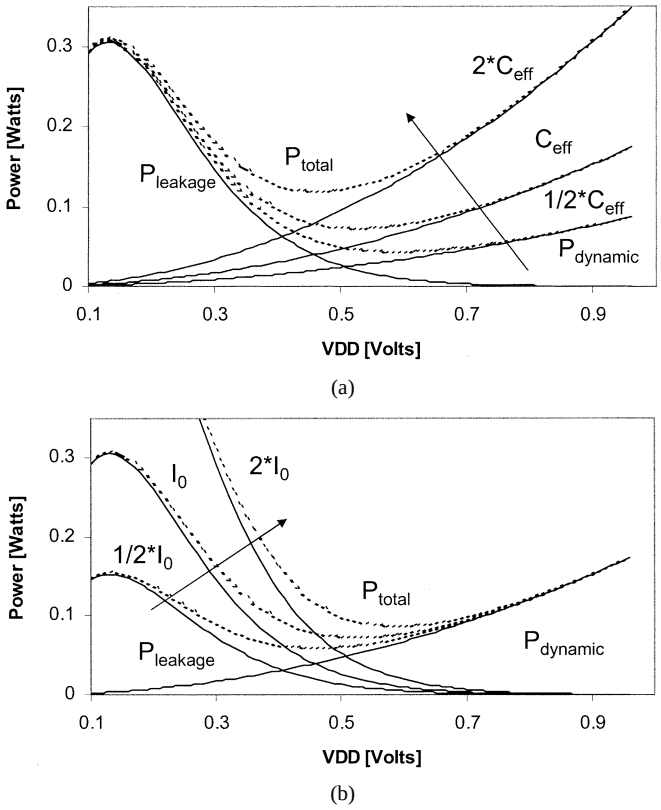

Fig. 4. Impact of changing (a) C_{eff} or (b) I_0 on optimal operating point.

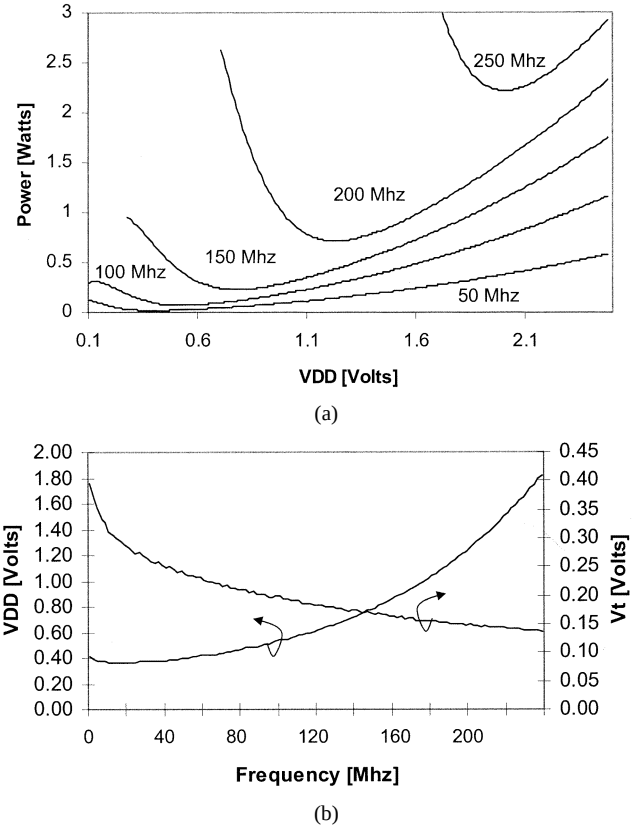

Fig. 5. (a) Impact of operating frequency on power versus V_{DD} curve (with V_t implicitly set). (b) V_{DD}/V_t optimal combinations for different operating frequencies.

B. Circuit Parameter Impact on Minimum Power

The simple theoretical models for total active power consumption shown in (6) and (7) are useful for understanding the tradeoffs between dynamic and leakage power as parameters vary. Qualitatively, architectural choices that determine C_{eff} and logic depth, technology and temperature choices that determine V_t and I_0, and workload requirements that determine operating frequency all affect the optimal $V_{DD} - V_t$ balance point.

Fig. 4(a) shows how the optimal $V_{DD} - V_t$ point shifts toward lower supply voltages when the average switched capacitance C_{eff} increases, assuming that the leakage component stays constant (a scenario for example where the activity factor increases but the leakage width remains fixed). This shows that, for circuits that are heavily skewed to having larger dynamic currents, the equations favor operating at lower supply voltages, at the expense of increased leakage power, to reduce total active power consumption. Fig. 4(b) shows a hypothetical case where the leakage constant I_0 increases independent of the C_{eff} term (in a real circuit both would likely vary) and the reverse effect occurs where optimum points shift toward higher V_{DD} and higher V_t combinations. This more effectively reduces subthreshold leakage currents at the expense of higher dynamic power. The effect of temperature variations on optimal operating points can be predicted from (6) and (7) as well. For a given operating frequency, the dynamic power term, as a function of different V_{DD} operating conditions, will be independent of temperature. However, I_0 and S will both increase with temperature and result in larger leakage currents for a given V_{DD}/V_t target. As temperatures increase, the optimal V_{DD} operating point shifts again toward higher supply voltages to minimize leakage power at the expense of increased dynamic power.

One particularly important parameter that greatly affects the optimal V_{DD}/V_t point is the workload requirement because it directly affects the balance between fast computation and slow computation. For a large class of circuits, it may be very useful to vary the operating frequency during runtime based on changing workload requirements. For example, a DSP that is processing image data will have different workloads depending on the dynamics of the image data [13], an encryption processor will require differing amounts of processing power depending on the level of security desired [14], or a microprocessor will operate at differing speeds depending on the application or the need to conserve battery power [15].

Fig. 5(a) shows theoretical power versus V_{DD} curves (with the V_t implicitly set) based on the SH4 extracted parameters and illustrates how the optimal power point changes as workloads change. If the operating point is not dynamically adjusted when different workloads are presented to the circuit, then the total power consumption can be far from the minimum. Fig. 5(b) shows in more detail the optimum V_{DD} and V_t values needed for different frequency targets. The trends indicate that, as performance increases, V_{DD} must increase, but threshold voltages should decrease. This illustrates that at higher operating frequencies it is more optimal to use low V_t devices because it is better to tolerate increased leakage currents using fast devices in order to minimize V_{DD} and dynamic power wherever possible. The opposite case holds for lower operating frequencies, where

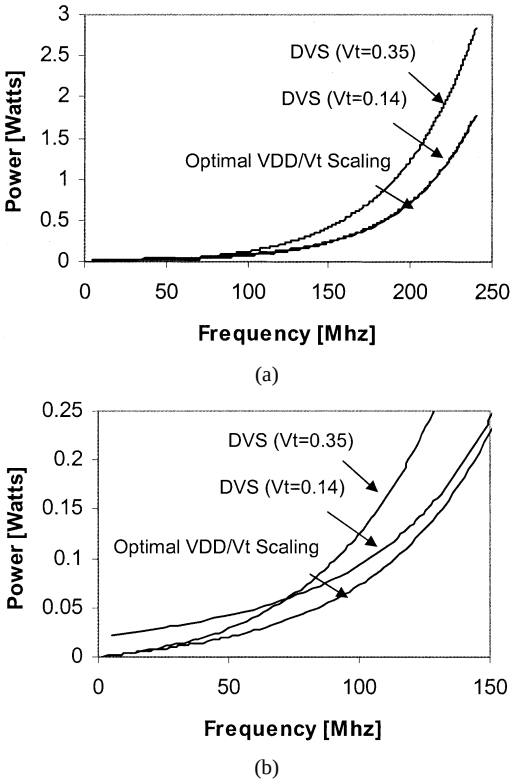

Fig. 6. (a) Adaptive V_{DD}/V_t scaling versus DVS as a function of frequency. (b) Magnified view.

the optimum shifts toward lower V_{DD} and higher V_t. In this case, V_{DD} is not scaled as aggressively because V_t increases, which helps lower subthreshold leakage currents that are more dominant at lower operating frequencies. At extremely low frequencies, leakage currents are so dominant that, in fact, it becomes more optimal to increase both V_{DD} and V_t in some cases. The behavior shown in these figures are dependent on the technology parameters used in the theoretical models. For more aggressive scaling, leakage currents will become more dominant, making the choice of V_t even more important. By adaptively tuning the different power components based on circuit operating conditions, significant power savings can be achieved.

C. Comparison of Dynamic Voltage Scaling to Dynamic Supply and Threshold Voltage Scaling

In the past, dynamic voltage scaling (DVS) has been proposed to reduce power consumption in scenarios where the chip workload varies in time [13]–[15]. By dynamically adjusting the supply voltage so that circuits operate only as fast as necessary, significant energy savings have been achieved. DVS is a significant improvement over simply dropping the operating frequency because it provides linear power savings from the frequency reduction as well as a quadratic savings in dynamic power due to supply voltage scaling. However, DVS completely neglects the effects of subthreshold leakage currents and thus will yield a nonoptimal solution in aggressive technologies where leakage currents are large. Fig. 6 shows theoretical comparisons between dynamic voltage scaling versus dynamic scaling of both V_{DD} and V_t. In the first case, the nominal process threshold voltage of 0.35 V was chosen to be constant. In the other case, it was assumed that the process threshold voltage was engineered to be a static 0.14 V, which corresponds to the optimal V_t value (with a corresponding V_{DD} of 1.7 V) for an operating frequency of 240 MHz using our models. For high frequencies, a threshold voltage of 0.35 V is significantly larger than ideal and, as a result, the supply voltage must be larger than necessary and the total power dissipation deviates greatly from the optimum case. In the case where the target threshold voltage is fixed at 0.14 V, this will give rise to optimal DVS operation at 240 MHz, but, at lower frequencies, the threshold voltage will be lower than optimum and again there will be significant deviation from the minimum power case. In general, with a fixed threshold voltage, the minimum power point cannot be achieved with DVS because there is no way to trade off dynamic and leakage power over a range of frequencies.

III. Test Chip to Explore Supply and Threshold Voltage Optimization

The theoretical analysis in the previous section shows that it is beneficial to dynamically tune V_{DD} and V_t separately in order to minimize total power dissipation. Practically, an easy way to dynamically tune threshold voltages is to utilize body biasing in a triple-well technology, as illustrated in Fig. 7. By applying both forward and reverse body bias, the threshold voltages can be shifted lower or higher to speed up or slow down devices dynamically. Forward bias is especially useful because modern technologies typically utilize threshold voltages skewed much higher than optimal from a power perspective. Furthermore, research has shown that forward body bias can also reduce short channel effects [16], thereby improving transistor performance. The following equation:

$$V_t = V_{t0} + \gamma\left(\sqrt{2\phi_B - V_{BB}} - \sqrt{2\phi_B}\right) \tag{8}$$

shows the well-known relationship relating how the body effect modulates the threshold voltage, where γ is the body factor and ϕ_B is the flatband voltage. Using this triple-well technology, an architecture to dynamically tune V_{DD} and V_t called automatic supply and body biasing (ASB) is developed. This methodology is similar to the theoretical approach of the previous section except that the variables to tune are V_{DD} and ΔV_{BB} (V_{BP} and V_{BN} offsets), which implicitly sets V_{tn} and V_{tp}.

A. Test Chip Implementation

A test chip consisting of a DSP core was fabricated and tested to explore the benefits of V_{DD}/V_t tuning as workload and operating conditions change in such a triple-well technology. The test chip was fabricated in a 0.14-μm technology with five metals, $T_{\text{ox}} = 3.2$ nm, $V_t = 0.05$ V defined as the V_{gs} corresponding to 1-nA drain current for a 1-μm device. This roughly translates to a standard linear extrapolated V_t of about 0.35 V. The chip consists of 16 parallel multiply-accumulate (MAC) units that can be used to model the core operation of a DSP and model the interaction between leakage power and dynamic power in a real circuit. The supply voltages and PMOS body and NMOS body voltages are all externally controllable

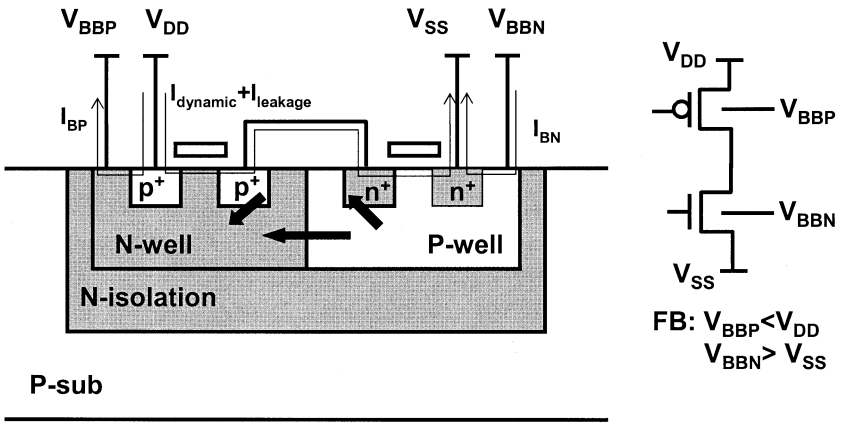

Fig. 7. Triple-well technology with switching, leakage, and forward bias drain-to-bulk currents shown. Heavy arrows show excess forward bias currents (P-well to N-well, P+ source to N-well, P-well to N+ source) that arise at low V_{DD} operation.

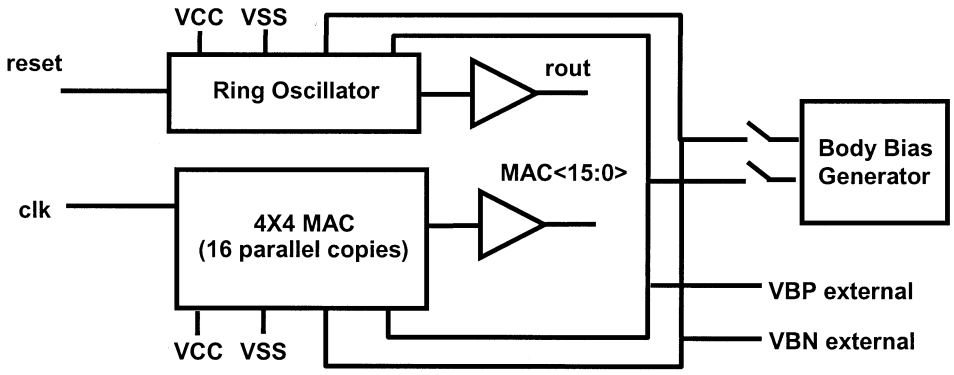

Fig. 8. ASB test chip block diagram.

so that different power combinations can be directly measured. A block diagram of the circuit layout is shown in Fig. 8.

The test chip is configured to operate at extremely low voltages below 1.0 V and very low frequencies below 100 MHz in order to explore circuit behavior where leakage power and dynamic power become comparable. Only at these very low voltages and low operating frequencies can the forward-biased devices have a low enough V_t such that subthreshold leakage currents are a reasonable fraction of the dynamic power dissipation, giving rise to a minimum power operating point. To provide ultralow-voltage operation, the MAC core is custom designed with pure CMOS gates, achieving operation down to a supply voltage of 175 mV.

B. Chip Architecture Details

The test chip consists of a 4×4 array of MAC units that operate in parallel. Each multiply-accumulate unit (MAC) consists of an 8-b array multiplier followed by a standard 24-b ripple carry adder and accumulator (Fig. 9). To model the DSP core operating in the active states, a random sequence of data vectors are input into the MAC unit. The data vectors are generated directly with a linear feedback shift register with a known starting seed pattern. The DSP thus cycles through a known test sequence and the chip functionality can be verified by simply comparing the accumulated output state to the predicted value after a fixed number of clock cycles. The MAC core has a separate power supply from the peripheral buffers to directly measure the DSP core power. The DSP test chip can be configured in a manual mode using external power supplies to control V_{DD} and the NMOS and PMOS body biases (V_{BN} and V_{BP}). This manual approach is to quantify how supply and body bias values vary for different operating frequencies.

Another available mode (which is useful for the ASB architecture proposed in Section V) is to generate the PMOS and NMOS body biases on chip using an adaptive body bias (ABB) generator [16]–[18]. A more complete description of the ABB implementation used in this chip, with experimental results, is given in [16]. The adaptive body bias generator uses a delay locked loop to tune the body bias values until the speed of a critical path replica exactly matches the target frequency (Fig. 10). Thus the body bias values are tuned so that this critical path delay is stretched or shortened until it exactly matches the period of the system clock. The DLL is configured to adjust both the PMOS and NMOS body biases by the same amount for each update period and will automatically adjust the threshold voltages to maintain a fixed chip performance as supply voltage or operating conditions change.

The critical path replica must properly track the performance of the DSP core critical path as supply and body bias values vary. Because the critical path replica and DSP core are located on the same die, they will track each other with process or temperature variations as long as the within-die matching is good. The matched delay line for the DLL is implemented as

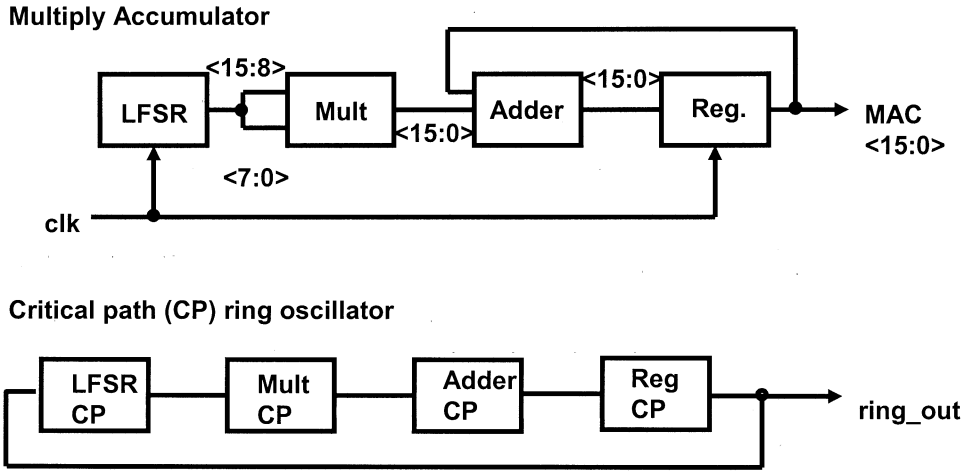

Fig. 9. MAC subblock and critical path ring oscillator.

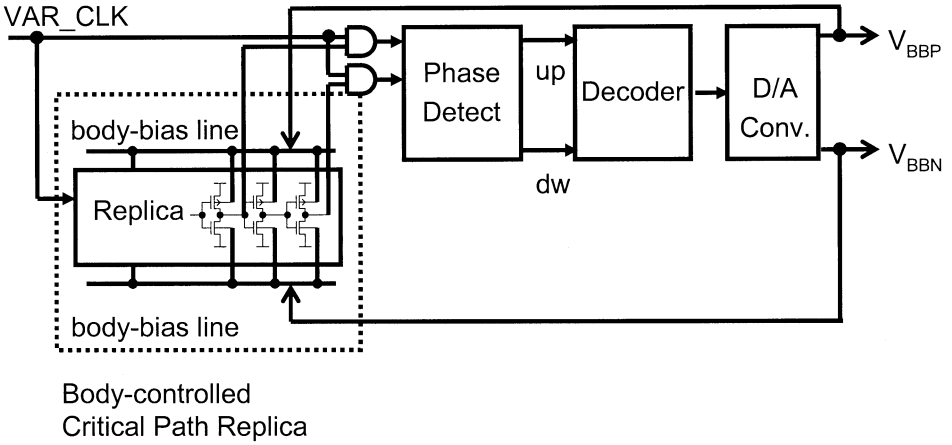

Fig. 10. Auto body bias generator based on DLL and critical path replica.

an inverter chain that uses the same gates with identical loading and identical layout orientation as the chip core circuitry. Each gate is configured to behave like a complex "inverter" with the worst case switching characteristics of the gate. This was accomplished by activating the worse case series paths with the minimum number of parallel chains for each gate in the matched delay line, which would then represent the worst possible delay for the MAC circuitry. The DLL matched delay line is also reproduced as a critical path ring oscillator that can be externally measured from the chip. This ring oscillator is used to measure the maximum operating frequency of the MAC circuitry during testing as supply and body bias knobs are tuned.

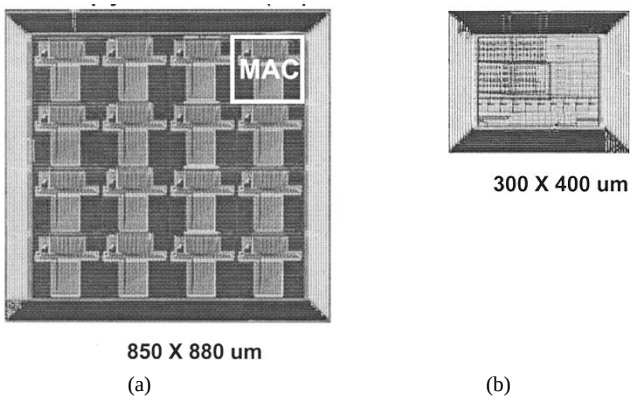

Fig. 11. Test chip die photograph. (a) An 8-b MAC (16 parallel units). (b) Body bias generator (ABB).

IV. Experimental Results

The MAC test chip was fabricated, measured, and shown to be functional down to 0.175 V, with the ring oscillator functioning as low as 0.1 V with zero body bias applied. A die photograph is shown in Fig. 11. The ring oscillator is constructed using exact replicas of the MAC critical path circuitry, so theoretically both should track exactly. For voltages greater than 0.175 V, the ring oscillator accurately tracks the speed of the MAC (i.e., the MAC is fully functional at the critical path clock speed). However, the MAC could not operate at as low a supply voltage as the ring oscillator, most likely due to the fact that the chip flip-flop circuits fail below 175 mV. At these very low supply voltages, the circuits operate in the subthreshold regimes since V_{DD} is less than the device threshold voltages. Fig. 12(a) shows a scope waveform corresponding to the complex ring oscillator operating at 0.1 V with the power supply shown. Fig. 12(b) shows the functioning DSP MAC with the input clock and a representative output pin.

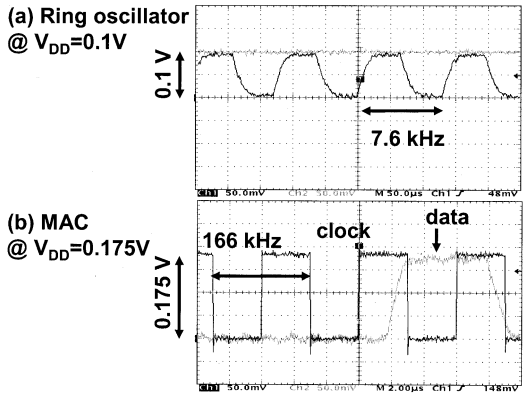

Fig. 12. Scope waveforms showing a ring oscillator functional at 0.1 V and a MAC unit functional at 0.175 V.

A. Threshold Voltage Tuning Limitations

The base technology used for the triple-well process has a nominal V_t of 0.05 V, defined as the V_{gs} corresponding to a 1-nA drain current for a 1-μm device. This roughly translates to a standard linearly extrapolated V_t of about 0.35 V. With body biasing, this V_t can be shifted approximately down to 0.3 V using 500 mV of forward bias and up to 0.45 V with 500 mV of reverse bias according to spice level simulations. Unfortunately, these threshold voltages are still high compared to the optimal threshold voltages that are needed for a circuit operating at a reasonably fast frequency. Consequently, even at very low operating frequencies, the threshold voltages could not be reduced enough to reach the theoretical balance point between subthreshold leakage currents and dynamic switching currents. However, the circuit can still achieve a minimum power point dictated by forward bias current limits that results in significant energy savings. The chip measurements show a more general minimum that takes into account subthreshold, dynamic and forward bias junction currents to find the minimum operating point. For example, the total power consumed by the circuit when forward bias currents exists can be written as

$$P_{\text{total}} = I_{\text{leakage}} V_{DD} + I_{\text{dynamic}} V_{DD} + I_{BN} V_{BN} + I_{BP} (V_{DD} - V_{BP}). \quad (9)$$

This increase in forward bias currents is an artifact of the technology threshold voltage being nominally too high and thus does not pose a fundamental limit to V_{DD}/V_t optimization. As technologies continue to scale, the more aggressive nominal threshold voltage will become closer to the theoretical optimum and only small amounts of forward bias (or even reverse bias) would be needed to dynamically adjust V_t's for all operating conditions

B. Forward Bias Currents

The test chip exhibits two different kinds of forward bias currents. The first type is the standard forward bias diode current that exists between the source and body junctions as shown in Fig. 7. These currents increase exponentially as expected and must be taken into account in the total active power dissipation. However, another unexpected parasitic forward bias current exists in our triple-well technology when operating at very low supply voltages. This "excess" forward bias current can flow between the p-well to n-well diode and also between the drain-to-body diodes. These parasitic forward bias currents only arise at low supply voltages. For example, when the supply voltage is at 0.8 V and the forward bias is 0.5 V, the p-well to n-well junction becomes forward biased by 0.2 V. In addition, when the output node is transitioning from low to high, the pMOS drain-to-body diode can become forward biased during the inverter high gain regime and, similarly, the nMOS drain-to-body diode can become forward biased when transitioning from high to low. These drain-to-body diode currents fight the dynamic switching current and can cause circuits to unexpectedly slow down. This mode of operation is detrimental because the increased forward bias not only increases static currents, but slows down devices as well. For the case of a larger supply voltage (e.g., 2.5 V, for example), there are no parasitic forward bias currents. The p-well to n-well junctions remain reverse biased and the pMOS and nMOS drain-to-body diodes also remain reverse biased during the critical transition periods and only exhibit forward biased junctions after the gate has already switched. An important conclusion is that forward bias benefits become limited at very low supply voltages and cannot be used to compensate for a technology with an intrinsically high threshold voltage.

C. MAC Performance Versus Body Bias

Fig. 13(a) shows measurements of the chip operating frequency as a function of the applied body bias for different V_{DD} values. The body bias amounts were applied equally to the pMOS and nMOS devices, where a bias of ΔV_{BB} corresponds to an nMOS well bias of $0\text{ V} - \Delta V_{BB}$ and a pMOS well bias of $V_{DD} + \Delta V_{BB}$. As such, a positive ΔV_{BB} corresponds to reverse body biasing and a larger V_t, while a negative ΔV_{BB} corresponds to forward body biasing and a smaller V_t. As more forward bias is applied, performance increases but, beyond a certain threshold, performance begins to roll off. This frequency degradation results from the excess forward bias described in the previous section where the p-well to n-well diode and drain-to-well diodes become forward biased. As expected, these diodes become more easily forward biased at lower supply voltages, which is confirmed in the figure where the roll off point shifts to the right for lower V_{DD}.

If the forward biased source and drain junction currents for the pMOS and nMOS are comparable, then the body currents I_{BN} and I_{BP} are approximately the same in magnitude. Fig. 13(b) shows measurements of I_{DD} and I_{BB} currents as a function of the applied body bias. The graph shows how the measured body currents are a strong function of both the applied forward bias and the supply voltage, which lends credence to the assumption that, at low voltages, the "excess" forward body bias causes large diode leakage currents to develop between the actual pMOS and nMOS wells. For example, for a fixed amount of applied body bias ΔV_{BB}, the amount of substrate current associated with a low V_{DD} value is much higher than that associated with a high V_{DD} value. If it were true that the substrate currents are dominated by the highly doped source junction currents for forward bias conditions, then there would be very little dependency on supply voltage because

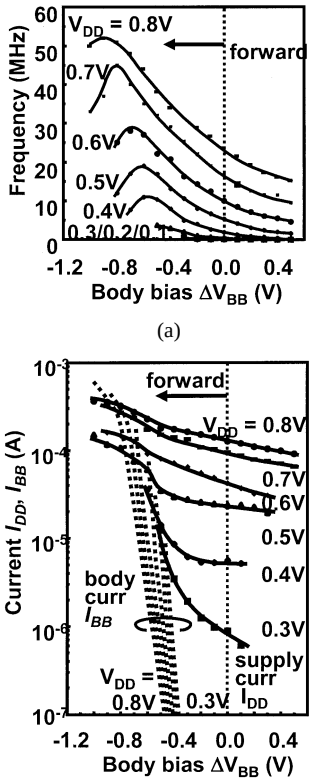

Fig. 13. (a) Frequency and (b) body currents as a function of body bias. Excess forward body bias shown performance degradation and large diode currents.

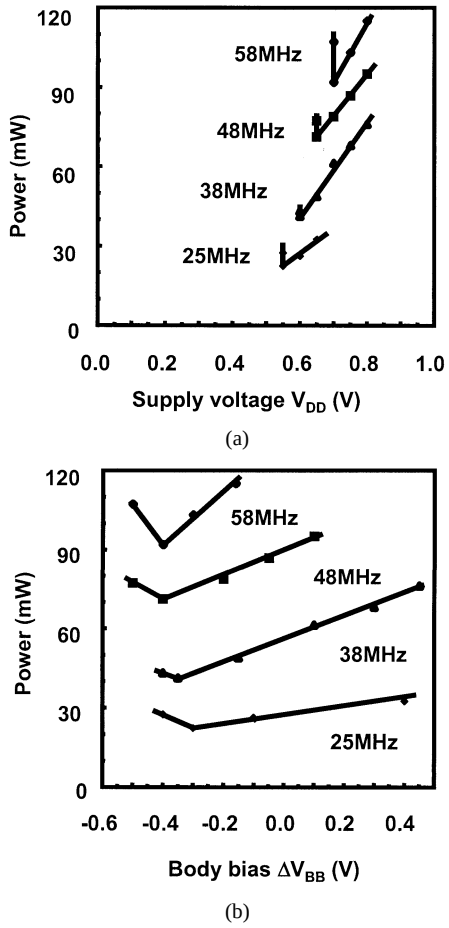

Fig. 14. Power versus (a) supply and (b) body bias showing a minimum point (limited by forward bias currents).

the voltage across the junction p–n diodes would be fixed. On the other hand, the current through the p-well to n-well diode is a strong function of the supply voltage as illustrated in the measurements. Once the substrate current suddenly increases with added forward body bias, the circuit reaches a physical limit where the devices can no longer improve with increased forward bias.

D. Optimal $V_{DD}/\Delta V_{BB}$ Operating Points

Fig. 14(a) shows power versus V_{DD} curves for several different target frequencies, where the ΔV_{BB} amount is implicitly chosen to satisfy the target frequency. As can be seen, there is a definite operating point that minimizes the total power dissipation, but it is limited by a sharp increase in forward bias currents rather than the theoretical balance between subthreshold leakage and dynamic currents. Fig. 14(b) shows the same data of power curves plotted against the applied body bias ΔV_{BB} instead of the supply voltage. In this case, the supply voltage is implicitly set by the choice of ΔV_{BB} needed to satisfy the target frequency. As frequencies drop, the local minimum in power for lower operating frequencies shifts to the right in the power versus ΔV_{BB} curves. This occurs because the excess forward bias currents are amplified at lower supply voltages and also because at low frequencies the dynamic power components are reduced and the optimal point is shifted toward reducing leakage currents.

Measurements show that significant power saving can be achieved by choosing V_{DD} and ΔV_{BB} appropriately depending on operating conditions. As a result, even if the theoretical minimum cannot be achieved in a given technology, aggressive use of forward biasing to speed up devices as much as possible can yield large power savings. Fig. 15 illustrates the benefits of the ASB approach versus standard dynamic voltage scaling for the DSP test chip. From these sample points, one can see that significant energy savings are achievable by tuning both V_{DD} and V_t rather than employing simple dynamic voltage scaling.

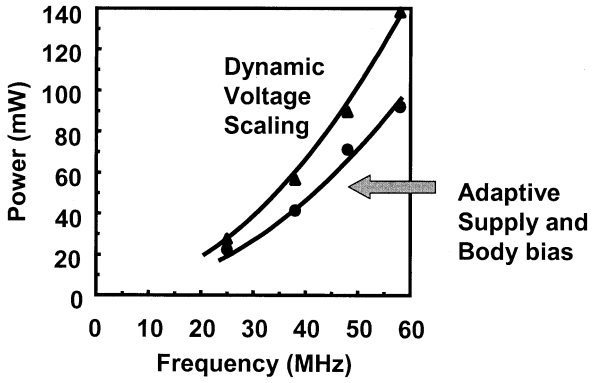

Fig. 15. ASB approach improvement over DVS.

V. ASB

To characterize the test chip, the supply voltage V_{DD} and body bias value ΔV_{BB} are manually adjusted to determine the

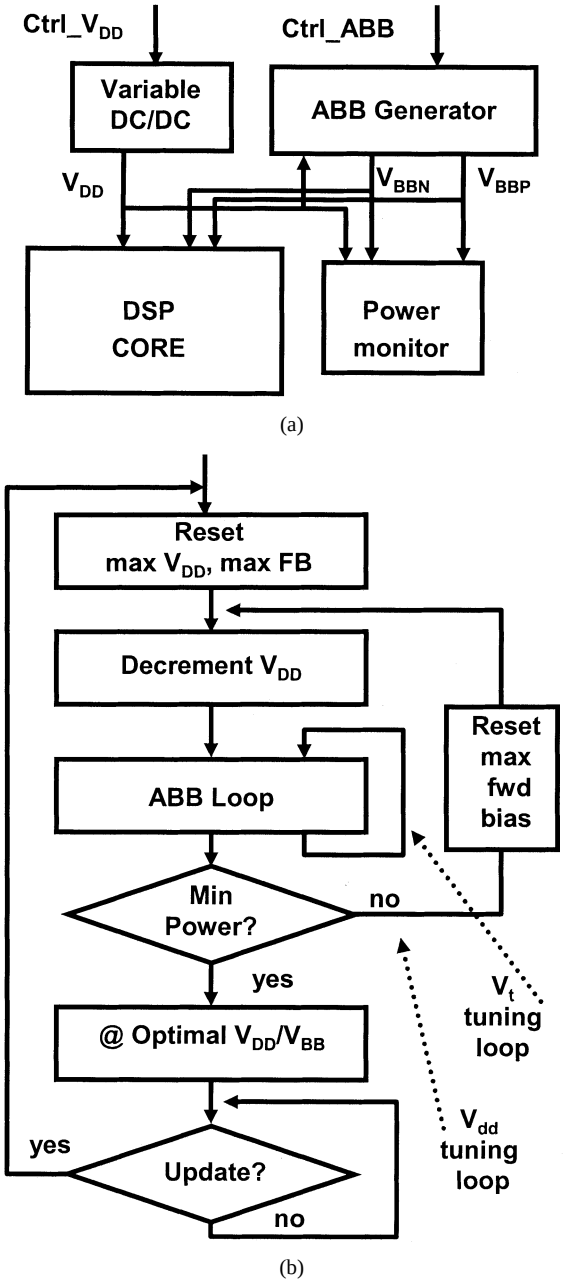

Fig. 16. (a) Dual-loop ASB architecture and (b) flowchart for automatic V_{DD}/V_T optimization.

minimum power point. Because the main goal of this research is to show how frequency and operating conditions affect this optimal point, a manual testing approach is appropriate. In a real application though, it would be necessary to automatically adjust supply voltages and threshold voltages based on operating conditions. A preliminary architecture for ASB is proposed below. One complication is that this is a three-dimensional (3-D) control problem because the supply voltage, the PMOS bias, and the NMOS bias can all be independently tuned. By always tuning PMOS and NMOS devices by the same amounts, the problem can be reduced to a two-dimensional (2-D) control loop.

Fig. 16(a) shows an approach for a dual-loop ASB architecture where the supply loop and body bias loops are decoupled and independently tuned to achieve minimum power. This approach relies on sweeping the supply voltage (the V_{DD} loop) and letting an auto body bias generator (V_{BB} loop) automatically lock to the appropriate V_{BB} value needed to maintain performance. This effectively decouples the 2-D control loop into two one-dimensional (1-D) control loops. The auto body bias generator loop is especially important because it automatically tunes the body voltages to force the chip to remain on an iso-performance curve independent of the operating conditions. If the controller sweeps the supply voltage, then the minimum active power point can be found by monitoring power through the chip or a representative cell. This idea is illustrated in the flow chart of Fig. 16(b). Initially maximum supply voltages and maximum forward bias are applied and the supply voltages are slowly lowered. At each step, the body bias is reset to maximum forward bias and locks to the target frequency to ensure that the chip stays functional. Once the minimum power point is detected, the bias values can be held in a register and the controller turned off. The loop can be reactivated whenever the workload changes, or periodically updated to reflect changes in temperature or operating conditions. Because the ASB loop does not have to be constantly running, the overhead power consumption, which is already amortized across the whole chip, can be reduced even further.

This ASB control loop is completely self-contained and should lock to the true minimum power configuration taking into account all possible current paths. This minimum will result in the physically lowest power consumption achievable by tweaking both V_{DD} and V_{BB} subject to the constraint that the chip satisfy a target frequency. Even for technologies where the theoretical limit is not yet achievable using body biasing techniques or in cases where excessive forward bias at low V_{DD} degrades performance, it is still possible to use this architecture to find the minimum physical power condition.

VI. Conclusion

As technology scales, new methodologies for low-power circuit operation need to be developed. This work shows how the balance between subthreshold leakage current and dynamic switching currents can greatly affect total active power consumption in future technologies. Significant energy savings can be achieved by optimally choosing V_{DD} and V_t for different circuit architectures and operating conditions and dynamically tuning these values based on workload variations during runtime. A test chip using a triple-well technology was designed to tune supply and body biases to implicitly adjust threshold voltages. Measurements show that forward bias currents can play a large role in the minimum achievable power, but this effect will decrease as technology is scaled more aggressively. A preliminary ASB architecture is also proposed to automatically adjust supply and body bias values by using a controller that decouples the V_{DD} and V_t loops by using a separate auto body bias generator circuit. It will be important to make the overhead in the ASB approach as small as possible in order to reduce total power. This can be done through careful design, amortizing the circuitry over the whole chip, and selectively updating the ASB controller only when updates are needed. Other architectures to tune threshold

voltages (for example, dual-gated SOI) can also be explored in the future since body biasing ranges could be less effective in future technologies [19], [20]. Because subthreshold leakage currents will contribute a large portion of the active power in future digital circuits, it is important to develop techniques such as ASB that will intelligently balance dynamic and leakage power as technology scales.

ACKNOWLEDGMENT

The authors are grateful to Hitachi, Ltd., for fabricating the test chip.

REFERENCES

[1] A. Chandrakasan, S. Sheng, and R. Brodersen, "Low-power CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 27, pp. 473–484, Apr. 1992.

[2] A. Keshavarzi, K. Roy, and C. Hawkins, "Instrinsic leakage in low power deep submicron IC's," in *Proc. Int. Test Conf.*, Nov. 1997, pp. 146–155.

[3] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, pp. 847–854, Aug. 1995.

[4] J. Kao and A. Chandrakasan, "Dual-threshold voltage techniques for low power digital circuits," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1009–1018, July 2000.

[5] M. Mizuno, K. Furuta, S. Narita, H. Abiko, I. Sasaki, and M. Yamashina, "Elastic-V_t CMOS circuits for multiple on-chip power control," in *ISSCC Dig. Tech. Papers*, Feb. 1996, pp. 300–301.

[6] T. Kuroda *et al.*, "A 0.9 V, 150 MHz, 10 mW, 4 mm2, 2-DCT core processor with variable V_t scheme," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1770–1778, Nov. 1996.

[7] Y. Ye, S. Borkar, and V. De, "A new technique for standby leakage reduction in high-performance circuits," in *Proc. 1998 Symp. VLSI Circuits*, June 1998, pp. 40–41.

[8] M. Miyazaki, J. Kao, and A. Chandraksan, "A 175 mV multiply-accumulate unit using an Adaptive Supply Voltage and Body Bias (ASB) architecture," in *ISSCC Dig. Tech. Papers*, Feb. 2002, pp. 58–59.

[9] J. Kao, "Subthreshold Leakage Control Techniques for Digital Circuits," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, 2001.

[10] V. Kaenel, M. Pardoen, E. Dijkstra, and E. Vittoz, "Automatic adjustment of threshold and supply voltages for minimum power consumption in CMOS digital circuits," *Proc. Int. Symp. Low-power Electronics and Design (ISLPED)*, pp. 78–79, 1994.

[11] D. Liu and C. Svensson, "Trading speed for low power by choice of supply and threshold voltages," *IEEE J. Solid-State Circuits*, vol. 28, pp. 10–17, Jan. 1993.

[12] R. Gonzalez, B. Gordon, and M. Horowitz, "Supply and threshold voltage scaling for low power CMOS," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1210–1216, Aug. 1997.

[13] V. Gutnik and A. Chandraksan, "An efficient controller for variable supply-voltage low power processing," in *Proc. Symp. VLSI Circuits*, June 1996, pp. 158–159.

[14] J. Goodman, A. Dancy, and A. Chandrakasan, "An energy/security scalable encryption processor using an embedded variable voltage DC/DC converter," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1799–1809, Nov. 1998.

[15] T. Burd, T. Pering, A. Stratakos, and R. Brodersen, "A dynamic voltage scaled microprocessor system," in *Proc. ISSCC*, Feb. 2000, pp. 294–295.

[16] M. Miyazaki *et al.*, "A 1.2-GIPS/W microprocessor using speed-adaptive theshold-voltage CMOS with forward bias," *IEEE J. Solid-State Circuits*, vol. 37, pp. 210–217, Feb. 2002.

[17] ——, "A delay distribution squeezing scheme with speed-adaptive threshold-voltage CMOS for low voltage LSI's," in *Proc. Int. Symp. Low-Power Electronics and Design (ISLPED)*, 1998, pp. 49–53.

[18] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. Chandrakasan, and V. De, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," in *ISSCC Dig. Tech. Papers*, Feb. 2002, pp. 422–423.

[19] S. Narendra, D. Antoniadis, and V. De, "Impact of using adaptive body bias to compensate die-to-die V_t variation on within-die V_t variation," in *Proc. Int. Symp. Low-Power Electronics and Design (ISLPED)*, Aug. 1999, pp. 229–232.

[20] I. Yang, C. Vieri, A. Chandraksan, and D. Antoniadis, "Back gated CMOS on SOIAS for dynamic threshold control," in *Proc. IEDM*, Dec. 1995, pp. 877–880.

James T. Kao received the B.S. degree from the University of California at Berkeley, in 1993 and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1995 and 2001, respectively, all in electrical engineering and computer science. His Ph.D. dissertation was entitled "Subthreshold Leakage Control Techniques for Low Power Digital Circuits."

His interest include low-power and mixed signal circuits, and he is currently working at Silicon Laboratories, Austin, TX, designing phase-locked loops.

Masayuki Miyazaki (M'01) was born in Tokyo, Japan, in 1966. He received the B.E. degree in electrical engineering from the University of Tokyo, Japan, in 1990 and the M.E. and Ph.D. degrees in superconductivity engineering from the University of Tokyo, Japan, in 1992 and 1995, respectively.

In 1995, he joined Hitachi, Ltd., Central Research Laboratory, Tokyo, Japan, where he has been engaged in the research and development of high-speed and low-power CMOS circuits techniques for microprocessors. From 2000 through 2001, he was a Visiting Scientist at Massachusetts Institute of Technology, Cambridge, doing research in the field of low-power systems.

Dr. Miyazaki is a Member of the IEEE Solid-State Circuits Society and the Institute of Electronics, Information and Communication Engineers of Japan.

Anantha P. Chandrakasan (S'87–M'95–SM'01) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer sciences from the University of California at Berkeley, in 1989, 1990, and 1994, respectively.

Since September 1994, he has been at the Massachusetts Institute of Technology, Cambridge, and is currently an Associate Professor of Electrical Engineering and Computer Science. He held the Analog Devices Career Development Chair from 1994 to 1997. His research interests include the ultralow-power implementation of custom and programmable digital signal processors, distributed wireless sensors, multimedia devices, emerging technologies, and CAD tools for VLSI. He is a coauthor of the book titled *Low Power Digital CMOS Design* (Norwell, MA: Kluwer) and a co-editor of *Low Power CMOS Design* and *Design of High-Performance Microprocessor Circuits* (Piscataway, NJ: IEEE Press).

Dr. Chandrakasan received the National Science Foundation Career Development Award in 1995, the IBM Faculty Development Award in 1995, and the National Semiconductor Faculty Development Award in 1996 and 1997. He has received several best paper awards, including the 1993 IEEE Communications Society's Best Tutorial Paper Award, the IEEE Electron Devices Society's 1997 Paul Rappaport Award for the Best Paper in an EDS publication during 1997, and the 1999 Design Automation Conference Design Contest Award. He has served on the technical program committee of various conferences including ISSCC, VLSI Circuits Symposium, DAC and ISLPED. He has served as a technical program co-chair for the 1997 International Symposium on Low-power Electronics and Design (ISLPED), VLSI Design '98, and the 1998 IEEE Workshop on Signal Processing Systems and as a general co-chair of the 1998 ISLPED. He was an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS from 1998 to 2001. He served as an elected member of the Design and Implementation of Signal Processing Systems (DISPS) Technical Committee of the Signal Processing Society. He was the Signal Processing Sub-committee chair for ISSCC 1999 through 2001 and the program vice-chair for ISSCC 2002. He is the technical program chair for ISSCC 2003.