Performance and Variability Optimization Strategies in a Sub-200mV, 3.5pJ/inst, 11nW Subthreshold Processor

Scott Hanson, Bo Zhai, Mingoo Seok, Brian Cline, Kevin Zhou, Meghna Singhal, Michael Minuth, Javin Olson, Leyla Nazhandali, Todd Austin, Dennis Sylvester, David Blaauw, *University of Michigan, Ann Arbor, MI*

Abstract

A robust, energy efficient subthreshold (sub-V_{th}) processor has been designed and tested in a 0.13µm technology. The processor consumes 11nW at V_{dd}=160mV and 3.5pJ/inst at V_{dd}=350mV. Variability and performance optimization techniques are investigated for sub-V_{th} circuits.

Keywords: subthreshold, variability, body-bias, low power

Introduction

Recent progress in low voltage circuit design has created opportunities for the development of inexpensive pervasive computing systems. Previous work has shown that minimum energy operation typically occurs in the sub-V_{th} regime ($V_{dd}<V_{th}$) [1,2]. In this work, we describe a sub-V_{th} processor for sensor network applications. The processor consumes 3.5pJ/inst at V_{dd}=350mV under zero body bias and 11nW at V_{dd}=160mV under a reverse body bias. There are two primary problems confronting sub-V_{th} designers: increased variability and reduced performance. We use extensive measurements to investigate both of these issues and demonstrate techniques for mitigating these problems. To control variability, we propose a body biasing strategy that leverages the unique sensitivities observed in sub-V_{th} circuits. To address performance, we investigate both chip-level and block-level performance enhancement techniques.

Design for Energy Efficiency

Sub-V_{th} operation must be accompanied by appropriate architectural design to ensure maximum energy efficiency. We use a simple architecture with 8-bit data and 12-bit instruction word sizes [3]. These word sizes have been chosen to maximize energy efficiency while still meeting the demands of sensor network processing. The core uses a 3-stage pipeline with instruction pre-fetching and branch speculation to reduce the number of clocks per instruction (CPI) and improve energy efficiency. A 1.5kb instruction memory and a 1kb data memory are both implemented using a latch-based memory with mux-based read/write schemes [4]. The processor has been fabricated in a 0.13µm twin-well technology with V_{th}~400mV at V_{ds}=50mV. Body biases have been routed in addition to supply rails, as shown in Fig. 2. V_{dd} and V_{ss} rails are routed in minimum pitch wires since interconnect resistance is negligible compared to device resistance in the sub-V_{th} regime. Three processor variants, each designed with a different gate sizing strategy, are highlighted in Fig. 3. Proc A uses minimum gate sizes, while Procs B and C (described later) use increased gate widths and lengths along critical paths.

Frequency and energy measurements of Proc A are shown for a typical die in Fig. 4. As predicted by [1,2], energy reaches a minimum due to increased leakage energy at low V_{dd}. The processor achieves a minimum of 3.5pJ/inst at V_{dd}=350mV with a frequency of 354kHz. The core (without memories, register file or pre-fetch buffer) reaches a minimum of 515fJ/inst at V_{dd}=290mV. In power-limited applications, a reverse body bias may be applied. Under a reverse bias of 300mV, the processor consumes 11nW at V_{dd}=160mV with the core consuming only 735pW. We focus on energy minimization for the remainder of this paper since it is more relevant for battery-powered applications.

Variability Control

Due to the exponential dependence of sub-V_{th} current on V_{th}, body biasing is useful for addressing variation in sub-V_{th} circuits. Previous work [5,6] has investigated body biasing, but we focus on its relevance to sub-V_{th} operation. For this discussion, we define two body bias quantities: *differential* ($V_{diff}=[V_{dd}-V_{B,PFET}]-V_{B,NFET}$) and *offset* ($V_{offset}=V_{B,NFET}$).

Matching between PFET and NFET devices, achieved by adjusting the body bias differential, maximizes noise margins in the sub-V_{th} regime. Fig. 5 shows how the minimum functional V_{dd} ($V_{dd,limit}$), a strong indicator of noise margins [6], is reduced from 180mV to 150mV when applying optimal differential. Fig. 5 also shows that the energy consumption of the core reaches a minimum at nearly the same differential. Though dynamic energy is largely insensitive to differential, leakage energy is minimized at the point where PFET and NFET strengths are matched. Fig. 5 stresses an important point: the differential that maximizes noise margins is almost identical to the differential that minimizes energy. Thus, body bias generation is simplified since the differential may be selected by matching the currents on NFET and PFET leakage monitors.

The body bias *offset* has exponential impact on performance in the sub-V_{th} regime. However, Fig. 6 shows that energy consumption is nearly independent of V_{th} as long as the system remains in the sub-V_{th} regime, supporting the derivation in [1]. Hence, the offset can be used to tune performance with minimal impact on energy.

With proper selection of differential and offset, we can address energy and performance variability due to systematic process variations. Fig. 7 shows that the mean $V_{dd,limit}$ for 20 dies reduces from 221mV to 168mV (a 24% improvement) when a unique energy optimal differential is applied to each die as in Fig. 5, suggesting a dramatic improvement in noise margins. Fig. 8 shows energy and frequency measurements at V_{dd}=300 mV over the same 20 dies for 4 different cases. In the first case, body biases are tied to the appropriate V_{dd} and V_{ss} rails (zero body bias). In the remaining 3 cases, the energy optimal body bias differential is applied, and body bias offset is chosen with 5mV resolution to meet frequency constraints of 66kHz (worst case frequency in Case 1), 100kHz, and 160kHz. The table in Fig. 8 summarizes the data from Cases 1 through 4 when all dies run exactly at the target frequency (66, 100 or 160kHz). Applying a body bias at 66kHz virtually eliminates delay variability and reduces the standard deviation of energy from 22fJ to 14fJ. Furthermore, a wide range of performance targets can be met with only minimal energy implications by tuning the body bias offset, as demonstrated by Cases 3 and 4. With a target frequency of 160kHz (Case 4), a 2.4X worst-case performance improvement and a 5% average energy improvement are achieved as compared to Case 1.

Similarly, proper selection of differential and offset can be used to compensate for variations in temperature. Fig. 9 shows that performance varies by 9.4X and energy increases by 74% between T=0 and 80ºC. By setting a constant differential as in Fig. 5 and tuning the offset at each temperature, frequency can be held nearly constant from T=0 to 80ºC and the energy increase can be reduced from 74% to 50%, as shown in Fig. 9.

Performance Control

We focus on four primary performance enhancement techniques: V_{dd} scaling, body biasing, gate length (L) sizing, and gate width (W) sizing. Fig. 10 compares V_{dd} scaling and body biasing as chip-level performance control techniques. Body biasing is clearly an energy-efficient alternative to V_{dd} tuning over the performance range shown since it allows PFET/NFET matching (Fig. 5) and leverages the insensitivity of energy to V_{th} (Fig. 6). However, since neither of these techniques can be easily applied at a block level, we also explore two gate sizing techniques. Improved drive strength is usually achieved by increasing W, but it can also be gained by increasing L in the sub-V_{th} regime. Due to halo implants, reverse short channel effects lead to V_{th} reductions at increased L. In the technology studied, on-current is increased by 2.4X when L increases from 120 to 200nm at V_{dd}=300mV. This high sensitivity suggests that L increases may be used to gain drive strength more efficiently than W increases alone. We measure energy and frequency for 3 processor variants: 1) minimum gate sizes (Proc A), 2) gate sizes increased using a typical standard cell library (Proc B) and 3) gate sizes increased using a standard cell library incorporating increased L and W (Proc C). The cores in Procs B and C were sized with energy as the objective as in [7]. Selected gate sizes from a subsection of the ALU are noted in Fig. 11 as an example. Proc C is both faster and more energy efficient than Proc B over the V_{dd} range shown, suggesting that L sizing is superior to W sizing. The use of increased L improves performance by 85% at V_{dd}=300mV for a 14% energy penalty compared to the case with minimum gate sizes, as shown in Figures 11-12. However, the frequency of Proc A can alternatively be increased by 85% for a ~7% energy penalty by increasing V_{dd} by 20-30mV, suggesting that L sizing is desirable only for block-level performance tuning.

References

[1] B. Zhai, et al., *DAC*, 2004.

[2] B. Calhoun, et al, *ISLPED*, 2004.

[3] L. Nazhandali, et al., *CASES*, 2005.

[4] A. Wang, A. Chandrakasan, *ISSCC.*, 2004.

[5] M. Miyazaki, et al., *ISSCC*, 2002.

[6] G. Ono, et al., *J. Solid-State Circuits*, 2003.

[7] S. Hanson, et al., *ISLPED*, 2006.

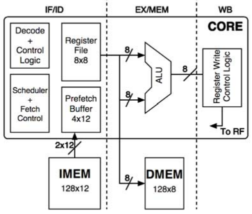

Figure 1: An 8-bit processor was implemented [3].

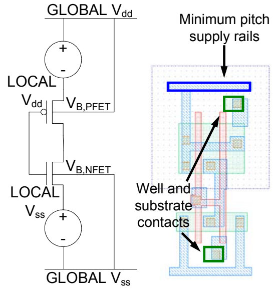

Figure 2: Body bias and power distribution schemes shown schematically

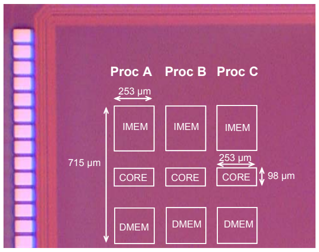

Figure 3: Die photograph highlighting core and memory for 3 processor variants

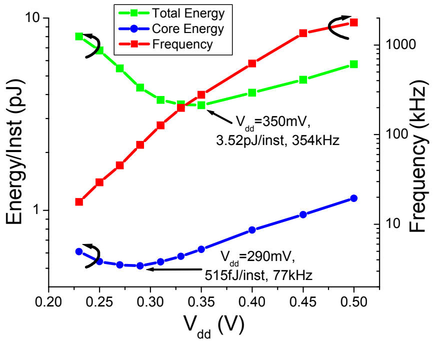

Figure 4: Energy and frequency measurements for an application with CPI~1.4

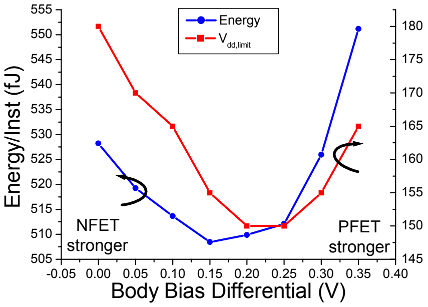

Figure 5: Energy and $V_{dd,limit}$ as functions of body bias differential for a single die

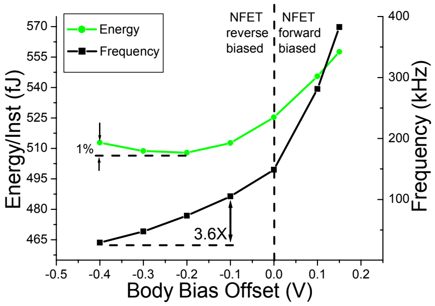

Figure 6: Energy and frequency as functions of body bias offset for a single die

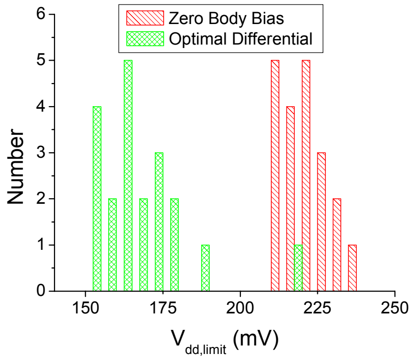

Figure 7: $V_{dd,limit}$ distribution over 20 dies with and without body biasing

Assuming freq. fixed at 66, 100, 160 kHz:

	Freq	μ_{energy}	σ_{energy}
Case 1: Zero body bias	66kHz	588fJ	22fJ
Case 2: Variable body bias	66kHz	528fJ	14fJ
Case 3: Variable body bias	100kHz	538fJ	17fJ
Case 4: Variable body bias	160kHz	559fJ	20fJ

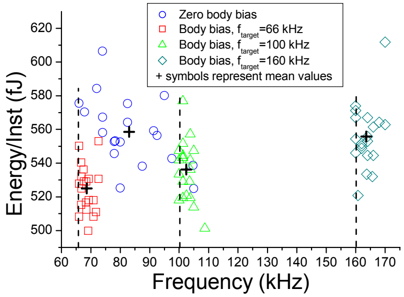

Figure 8: Energy and frequency distributions over 20 dies under various body bias schemes

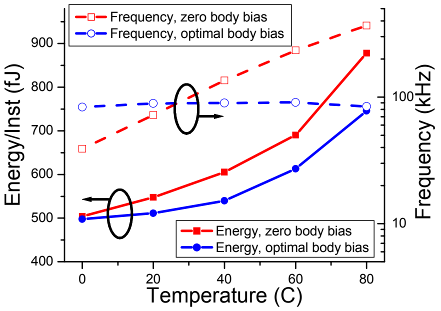

Figure 9: Energy and frequency temperature sensitivity for a single die with and without body biasing

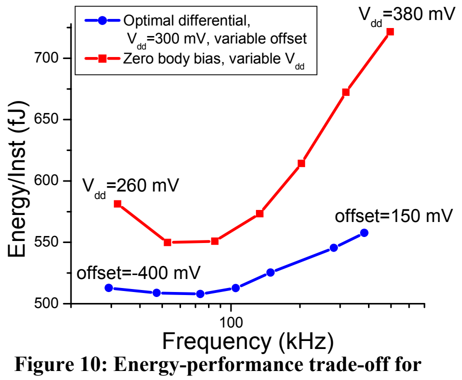

Figure 10: Energy-performance trade-off for variable body bias and variable V_{dd} systems

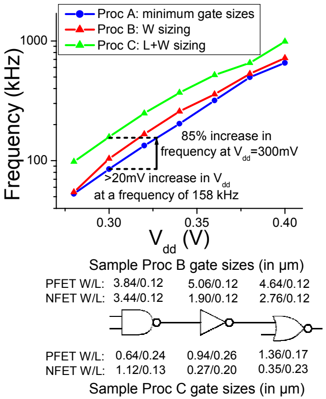

Figure 11: Frequency measurements for 3 core variants. Gate sizes along an ALU path shown.

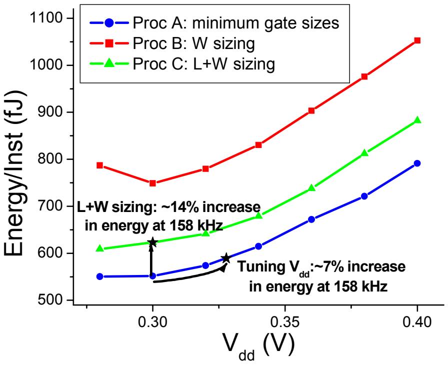

Figure 12: Energy measurements for 3 cores with different sizing schemes