

19.2 A 6.4pJ/Cycle Self-Tuning Cortex-M0 IoT Processor Based on Leakage-Ratio Measurement for Energy-Optimal Operation Across Wide-Range PVT Variation

Jeongsup Lee¹, Yiqun Zhang¹, Qing Dong¹, Wooteak Lim¹, Mehdi Saligane¹, Yejoon Kim¹, Seokhyeon Jeong¹, Jongyup Lim¹, Makoto Yasuda², Satoru Miyoshi³, Masaru Kawaminami^{2,3}, David Blaauw¹, Dennis Sylvester¹

¹University of Michigan, Ann Arbor, MI

²Mie Fujitsu Semiconductor Limited, Kuwana, Japan

³Fujitsu Electronics America, Inc., Sunnyvale, CA

Wireless sensors for IoT applications have become a prominent computing class and are typically severely power constrained. IoT devices are deployed in a wide range of environments and low power consumption must be guaranteed across a wide temperature range. The combination of dynamic voltage scaling (DVS) and adaptive body biasing (ABB) can achieve minimum total energy per cycle, where dynamic and leakage energy are at an optimal trade-off point [1]. However, due to the dependence of leakage on temperature and workload fluctuations over time, this optimal operating point requires runtime adjustment. Traditional approaches to track an optimal operating point involve repeated measurement of processor power inside a body-bias and supply voltage search loop. However, determining processor power consumption adds significant complexity and/or power overhead since it requires separate measurement of supply voltage and current followed by their multiplication.

In this paper, we make the key observation that at the optimal energy operating point, the ratio of leakage to dynamic power is a nearly constant value across temperature, process variation, and workload. We refer to this as the *optimal leakage ratio*. We derive the optimal leakage ratio value from basic MOSFET characteristics and show that it results in energy per cycle within 4.6% of optimal across PVT conditions. We then show how the leakage ratio can be efficiently and simply measured by counting the modulated frequency of a DC-DC converter and comparing the count obtained at two different clock frequencies. The proposed method was implemented in a Cortex-M0 processor in MIFS 55nm deeply depleted channel (DDC) CMOS with a custom DC-DC converter for supply voltage regulation and charge pumps for body bias generation. A control loop dynamically measures the leakage ratio and processor speed using a critical-path replica and automatically adjusts the body bias and supply voltage. Measurements demonstrate that the design can maintain the optimal ratio across a wide temperature range (-20°C to 125°C) and 5 process corners from split-wafer lots. The processor achieves 6.4pJ/cycle, making it the most energy efficient reported among recent low-power industry-standard processors and operates from 100kHz-6MHz for IoT applications.

Figure 19.2.1 provides a derivation of the optimal leakage ratio in a typical digital circuit. It is well known that total energy shows a minimum at a particular supply voltage (referred to as the minimum energy point, or MEP) due to the growing leakage energy associated with increased delay at reduced voltage. Theoretically, the MEP occurs when the slopes of dynamic and leakage energy become equal and opposite. Dynamic power can be modeled as a simple quadratic with supply voltage (Fig. 19.2.1, Eqn. 2), and we use a power model ($y = aV_{DD}^{-b}$, Eqn. 1) to fit leakage, similar to [1], to facilitate the derivation. Combining these two expressions and minimizing power (setting slopes equal) we find that the ratio of dynamic power to leakage power at MEP is half the parameter value of b , which is 9 in the targeted 55nm technology leading to the P_{leak}/P_{total} ratio of 18%. While this ratio varies with technology node (0.32 in 28nm and 0.10 in 40nm CMOS, simulation) it is nearly constant across temperature and process corners in a particular technology node, as confirmed in Fig. 19.2.1 (top right, simulation), where the 18% ratio yields power within 2% from the minimum.

The overall system architecture is shown in Fig. 19.2.2 and consists of a configurable DC-DC converter that supplies the processor and memory and two bias-voltage generators for compensating leakage across PVT. To tune the processor operation to the optimal leakage ratio (1/4.5) we need to measure both leakage and dynamic power of the processor which we accomplish by monitoring the frequency of the DC-DC converter. The DC-DC converter supplied current is determined by its flying cap, voltage drop (Δh), and frequency (Fig. 19.2.2). At runtime, the flying cap size is constant and Δh is regulated, implying that the DC-DC converter frequency is proportional to its load current (consisting of leakage and dynamic currents). Since dynamic current is proportional to clock frequency, two equations (1), (2) in Fig. 19.2.2 can be derived by modulating processor clock frequency occasionally to measure its leakage/dynamic current ratio. Since the desired optimal leakage ratio is 1/4.5 for this system, the ratio of CNT_{slow}/CNT_{norm} (R_{CNT}/FF_{16}) should be 0.59 (i.e., $96_{16}/FF_{16}$), as derived in Fig. 19.2.2. By tracking this optimal leakage ratio, the system determines if the system is at MEP with very low hardware and power overhead.

Figure 19.2.3 shows a detailed block diagram of the implemented system. The system has two power domains; 1.2V and the DC-DC output power domain. The 1.2V power domain supplies the DC-DC converter, charge pumps, clock generator, and controller. All blocks within this power domain use extremely low leakage (ELL) thick-gate-oxide devices to achieve low leakage power across all corners and temperatures. Due to the large V_{th} of ELL devices, we adopt a new body-controlled transmission gate (Fig. 19.2.3) for the configuration switches in the DC-DC converter to ensure small on-resistance and large off-resistance across all corners and wide temperature range, improving DC-DC converter efficiency. Charge pumps for the processor and SRAM were designed to avoid leakage paths through the power and/or body since the system supports both forward and reverse body bias to enable MEP operation across all process and temperature corners. The Nwell charge pump automatically adjusts its output voltage matching the V_{th} of PMOS and NMOS [2], hence the system is able to compensate for skewed corners as well. Charge pumps for ELL devices were designed for reverse body bias only. To prevent a current path from the body to power or ground when charge pumps are running, 3-transistor switches are used. By default, the charge pumps for the ELL devices are turned off and 3-transistor switches are turned on. At extremely high temperature, such as 125°C and at the FF corner, where the leakage of ELL devices become significant, this charge pump is enabled to limit leakage in ELL devices. A custom 8KB SRAM is implemented to support the Cortex-M0 processor. SRAM word line drivers are boosted using the 1.2V supply. Since the processor and SRAM share the same power domain, the SRAM is designed to match processor speed.

Figure 19.2.4 describes the control algorithm to find the MEP of the system. The target range of R_{CNT} is $90_{16} \sim 98_{16}$, within which the operating power is within 4.6% of optimal. For R_{CNT} smaller than 90_{16} the leakage power is too small, indicating that we should apply further forward body bias first before decreasing V_{DD} to prevent the processor from exiting its functional voltage window. A critical-path replica is used to detect how close to the boundary of the functional voltage window the processor is operating. Two thresholds for the number of errors detected by this replica circuit are used to tune the processor close to the maximum operating frequency, while maintaining functionality. Fig. 19.2.4 also includes oscilloscope traces demonstrating the fully automated control loop running on the test chip. The system finds the MEP at 60°C from a default non-optimal starting point. Then, as temperature gradually changes from 60°C to 10°C, the system automatically adjusts its V_{DD} and body bias in the direction of the MEP. As the temperature drops, leakage power reduces and forward body bias is automatically applied to maintain the optimal leakage ratio, while V_{DD} decreases to lower dynamic power. These changes are made while maintaining safe operation in the functional voltage window using the critical-path replica.

Figure 19.2.5 provides measurement results for the implemented system. The design is tested at TT/FF/SS/FS/SF corners, as well as across a wide temperature range from -20°C to 125°C. Various target frequencies are chosen, ranging from 100kHz to 6MHz. The proposed approach achieves power consumption within 4.6% of optimal at 1MHz across all mentioned process and temperatures (Fig. 19.2.5, top, labeled MEP). Fig. 19.2.5 also shows the optimal V_{DD} and body bias values across target frequencies and temperatures. At fixed temperature, optimal V_{DD} is nearly constant, while body bias adjusts as frequency varies (Fig. 19.2.5, bottom right). The Shmoo plot shows R_{CNT} gradually changes according to the V_{DD} and body bias and indicates the optimal value at the selected operating point. The pie chart gives the power breakdown of the system: the DC-DC converter supplies the core and 8KB SRAM for instructions and data, while support circuits such as interfaces, the ELL charge pump, clock, and the bus interface run at constant 1.2V and consume 10.3%. Figure 19.2.6 provides a comparison table to other small IoT processors. The proposed approach offers closed-loop MEP-tracking via clock frequency modulation embedded in the DC-DC converter, and lowest energy per cycle of the listed commercial processors. Figure 19.2.7 provides the die photo.

References:

- [1] S. M. Martin, et al., "Combined Dynamic Voltage Scaling and Adaptive Body Biasing for Lower Power Microprocessors under Dynamic Workloads," *IEEE/ACM ICCAD*, pp. 721-725, 2002.
- [2] G. Ono, et al., "Threshold-Voltage Balance for Minimum Supply Operation," *IEEE JSSC*, vol. 38, no. 5, pp. 830-833, 2003.
- [3] J. Myers, et al., "An 80nW Retention 11.7pJ/Cycle Active Subthreshold ARM Cortex-M0+ Subsystem in 65nm CMOS for WSN Applications," *ISSCC*, pp. 144-145, 2015.
- [4] D. Bol et al., "SleepWalker: A 25-MHz 0.4-V Sub-mm² 7-μW/MHz Microcontroller in 65-nm LP/GP CMOS for Low-Carbon Wireless Sensor Nodes," *IEEE JSSC*, vol. 48, no. 1, pp. 20-32, 2013.
- [5] G. Chen, et al., "Millimeter-Scale Nearly Perpetual Sensor System with Stacked Battery and Solar Cells," *ISSCC*, pp. 288-289, 2010.
- [6] B. Zhai, et al., "A 2.60pJ/Inst Subthreshold Sensor Processor for Optimal Energy Efficiency," *IEEE Symp. VLSI Circuits*, pp. 154-155, 2006.

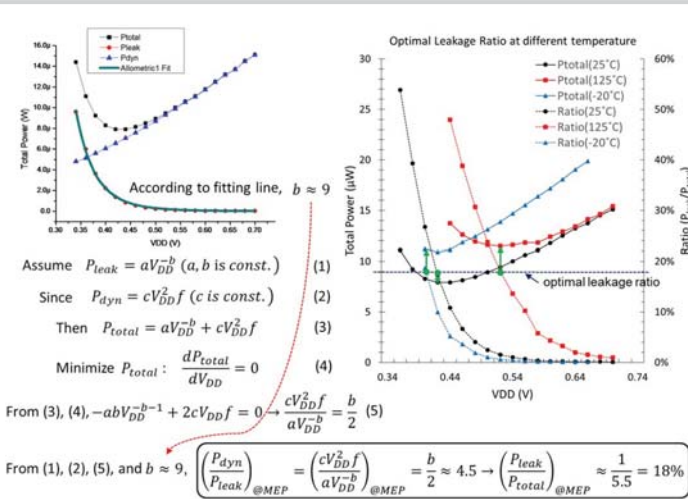


Figure 19.2.1: Derivation of optimal leakage ratio (left); Plot at right shows that total power is within 2% of optimal at derived 18% leakage ratio across temperatures from -20°C to 125°C.

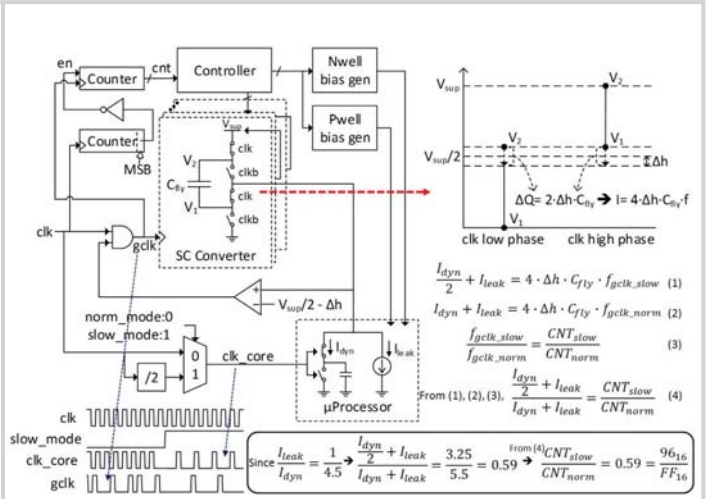


Figure 19.2.2: Proposed approach to measure the ratio of leakage to dynamic power and achieve the desired ratio of CNT_{slow} to CNT_{norm}.

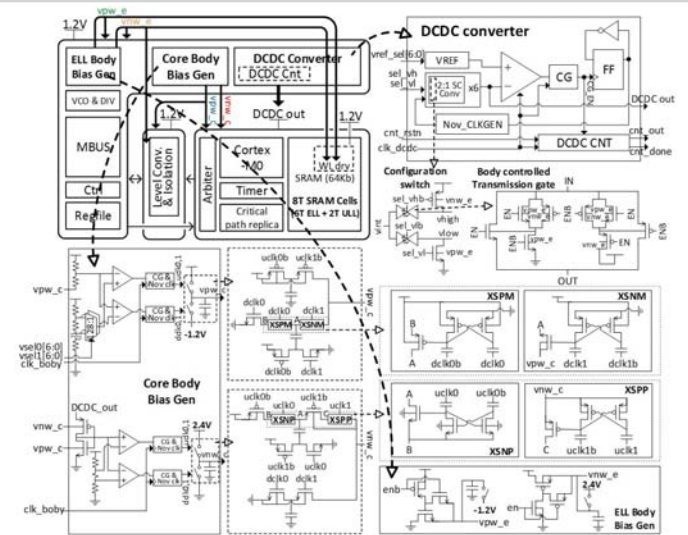


Figure 19.2.3: Top-level block diagram and detailed circuits of DC-DC converter and charge pumps.

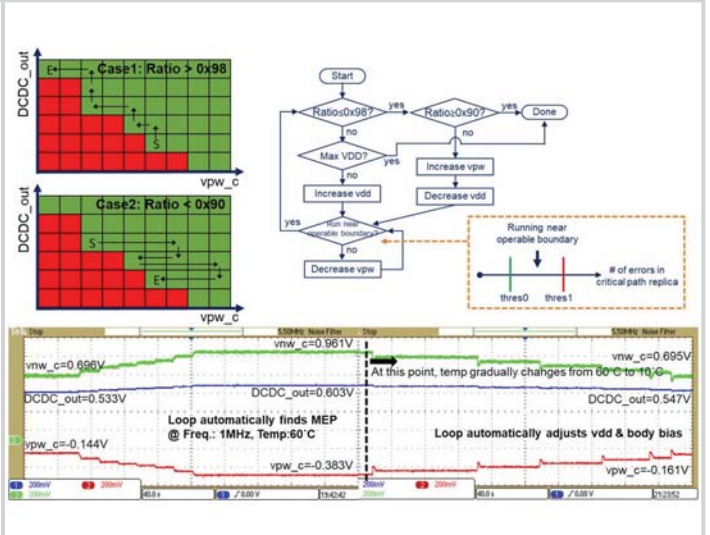


Figure 19.2.4: Algorithm describing system operating loop and measured automatic DC-DC output and body bias adjustment.

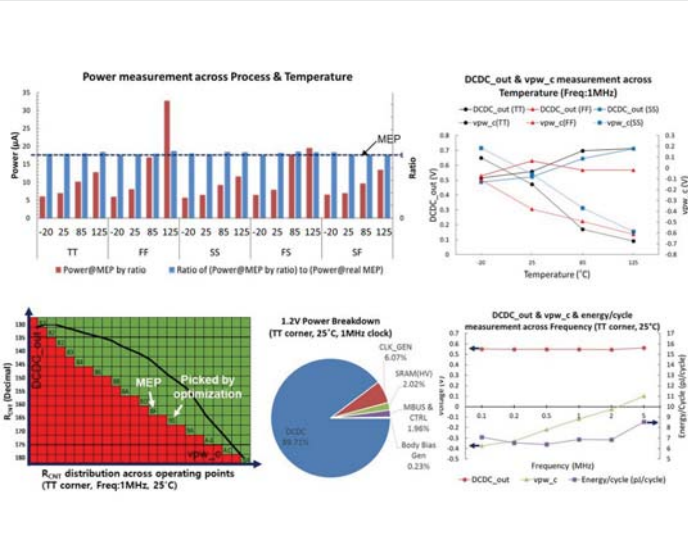


Figure 19.2.5: Measurement results.



Figure 19.2.6: Comparison table and photo of the proposed system.

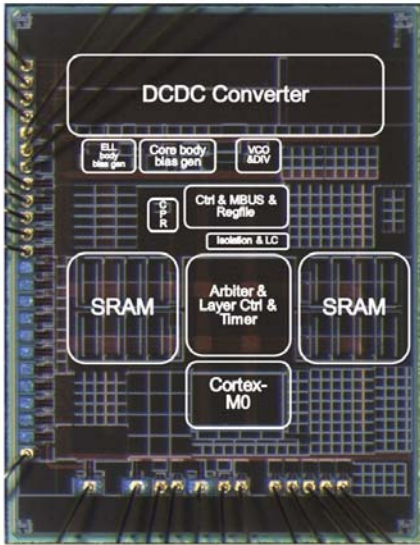


Figure 19.2.7: Die photo in 55nm CMOS.