Fine-Grain Dynamic Leakage Reduction

Seongmoo Heo, Kenneth Barr, Mark Hampton, and Krste Asanović

Introduction: Recently, leakage power dissipation in microprocessors is increasing so rapidly due to threshold voltage reduction and number of transistors increase that it is predicted leakage power could be comparable to dynamic power within a few process generations.

Approach: Previous work in leakage current reduction for digital circuits can be divided into two main categories: Static design-time Selection of Slow, low-leakage Transistors for non-critical paths (SSST) and Dynamic Deactivation of Fast, leaky Transistors on critical paths (DDFT). Leakage power is dominated by critical paths, and hence dynamic deactivation of fast transistors could potentially yield large savings. Existing dynamic leakage reduction techniques, although they have low steady-state leakage, have large deactivation energies and significant startup latencies. The large deactivation energies require long idle times to amortize their overhead, and large startup latencies impact performance, limiting the applicability of these techniques within an active microprocessor. DDFT schemes must be designed with all of these parameters in mind.

Progress: We have introduced new circuit techniques that have a low deactivation energy when transitioning a circuit block into a low leakage state from which it can be woken quickly. We have shown how these techniques can be applied at a fine grain within an active superscalar microprocessor and how microarchitectural scheduling policies can improve their performance [1, 2].

First, we focused on array structures: L1 SRAM caches and multiported register files. We have developed a simple circuit technique, *Leakage-Biased Bitlines* (LBB), that reduces bitline leakage current due to the access transistors of SRAM cells. LBB requires minimal transition energy and minimal wakeup time. Instead of forcing zero sleep values onto the read bit lines of inactive subbanks, this technique just lets the bitlines float by turning off the high-V_T NMOS precharging transistors. The leakage currents from the bit cells *automatically* bias the bitline to a mid-rail voltage that minimizes the bitline leakage current. Although the bitline floats to mid-rail, it is disconnected from the senseamp by the local-global bitline switch, so there is no static current draw. This technique has little additional transition energy because the precharge transistor switches exactly the same number of times as in a conventional SRAM, we only delay the precharge until the subbank needs to be accessed. The wakeup latency is just that of the precharge phase.

Figure 1 compares the steady-state leakage power of the leakage-biased bitline and the forced-zero/forced-one sleep vector techniques with the original leakage power for a 32-row$\times$16B SRAM subarray with varying numbers of stored ones and zeros. It is clear that the leakage-biased bitline technique has the lowest leakage power independent of stored bit values. Figure 2 compares the cumulative idle energy and the LBB DDFT energy consumption for different processes. The LBB DDFT technique must replace the lost charge on the bitline before the attached memory cells can be used. This one-time energy requirement (which is no more than the full-swing energy of the bitline) manifests itself as a penalty unless the cell is inactive long enough for the energy saved by LBB to exceed the energy spent due to transitioning. The break-even point is around 200 cycles in a 180 nm process. However, since active energy scales down faster than leakage energy, the break-even time decreases with feature size. In a 70 nm process, the break-even point is less than one cycle. We have shown that SRAM read path deactivation saves over 40% of idle circuit leakage energy and over 20% of total I-cache energy when using a 70 nm process.

The LBB scheme can be applied, in similar fashion, to multiported register files with no loss in performance in order to deactivate idle read ports or dead registers by subbank granularity. Dynamically deactivating dead registers reduces register file idle-circuit leakage energy by 41.1% and total register file energy by 12.4%. Deactivating read ports within a multiported register file saves 87.8-98.5% of leakage energy and 9.3-47.8% of total energy.

Second, we developed a Leakage-Biased Domino circuit family that maintains high speed in active mode but can be rapidly placed into a low-leakage inactive state by using leakage currents themselves to bias internal nodes [2]. Figure 3 shows a leakage-biased domino buffer which goes to sleep state regardless of input states by turning off two sleep transistors (one on top of the keeper and the other below the static logic). A 32-bit Han-Carlson domino adder circuit was used to compare LB Domino with conventional single and dual-V_T domino circuits. We found that for equal delay and noise margin, the LB Domino technique gives two decades reduction in steady-state leakage energy compared to a dual-V_T technique regardless of input states. Cumulative sleep energy graphs show that LB-Domino

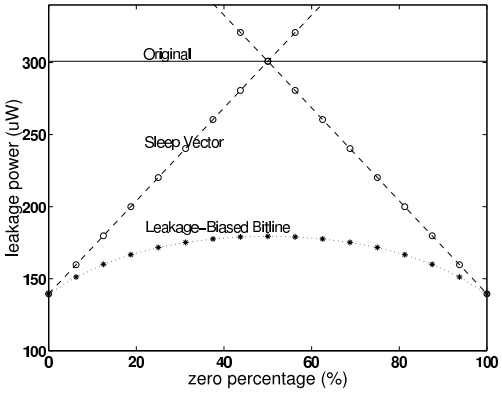

Figure 1: The leakage power of an SRAM subbank.

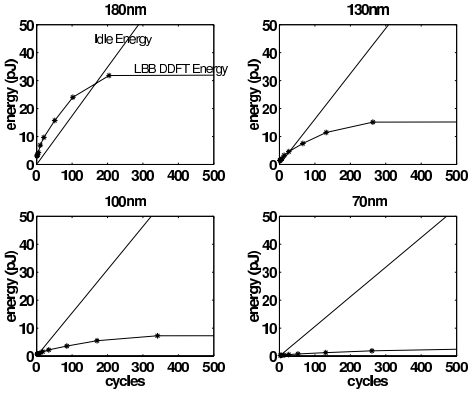

Figure 2: Idle energy and LBB energy of an SRAM subbank.

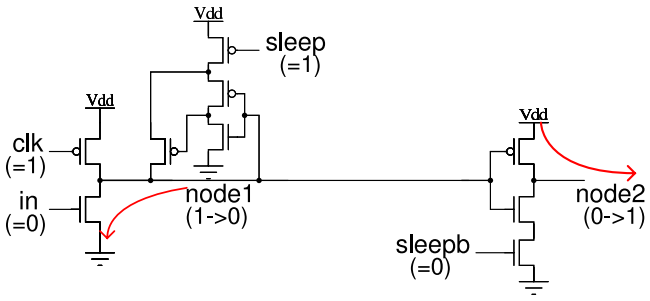

Figure 3: A leakage-biased domino buffer.

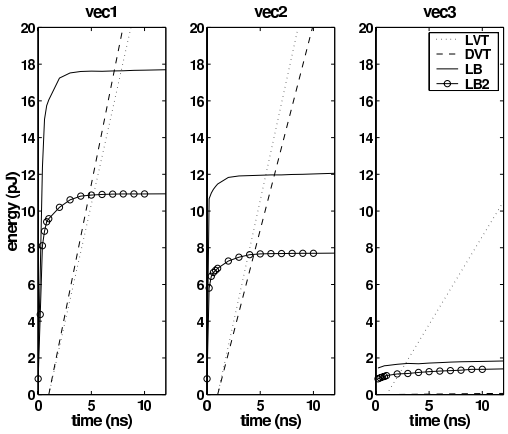

Figure 4: Cumulative sleep energy comparison.

techniques have less than 10 ns break-even point, which is the cross-over point between LB (or LB2: modified LB-domino) and LVT, the baseline adder (Figure 4).

Future: We are developing new fine-grain dynamic leakage reduction circuit techniques together with new microarchitectural algorithms that increase component sleep times. We are designing test chips to validate our circuit ideas.

Research Support: This work is supported by DARPA PAC/C award F30602-00-2-0562, NSF CAREER award CCR-0093354, NSF ITR Grant CCR-021-9545, a donation from Infineon Technologies, and an equipment grant from Intel.

References:

[1] S. Heo, K. Barr, M. Hampton, and K. Asanović, "Fine-grain dynamic leakage reduction," in *International Symposium on Computer Architecture*, May 2002.

[2] S. Heo and K. Asanović, "Leakage-biased domino circuits for dynamic leakage reduction," in *Symposium on VLSI Circuits*, June 2002.