

22.3 A Process-Variation-Tolerant Floating-Point Unit with Voltage Interpolation and Variable Latency

Xiaoyao Liang, David Brooks, Gu-Yeon Wei

Harvard University, Cambridge, MA

Process variation will greatly impact the power and performance of future microprocessors. Design approaches based on multiple supply or threshold voltage assignment provide techniques to statically tune critical path delays for energy savings [1]. However, under process variation, the delay of critical paths may vary, and a large number of critical paths in circuits reduces the maximum operating frequency of pipelined processors. One proposed post-fabrication solution is to adaptively tune the back-body bias to combat variations for logic structures [2]. Dynamic voltage switching between two power supplies, using level shifters to cross voltage domains, has also been proposed to primarily reduce power [3-5]. This paper explores two fine-grained, post-fabrication circuit-tuning techniques to combat process variation for pipelined logic components—voltage interpolation and variable latency. These techniques are applied to a single-precision floating-point unit (FPU) designed using a standard CAD synthesis flow in a 0.13 μ m CMOS logic process with 8 metal layers. Measured results from fabricated chips show that both techniques provide wide frequency tuning range to deal with frequency fluctuations arising from process variations with minimal power overhead, and in some configurations, power savings.

Figure 22.3.1 illustrates the circuit architecture. The FPU is pipelined into 6 stages with two power supplies (VddH, VddL) provided across the unit. Each pipeline stage independently selects one of the two voltages, resulting in 64 different voltage configurations. By maintaining a small difference between VddH and VddL, the design does not need level shifters. Latch-based clocking enables time borrowing across pipeline stages such that choosing different voltage configurations leads to different *effective voltages*, somewhere between VddH and VddL, across the FPU. This spatial voltage dithering provides broad frequency tunability. Each pipeline stage is divided into two clocking domains, controlled by complementary clocks ($\Phi 1$, $\Phi 2$). To increase borrowing, an additional stage can be introduced by adding one extra latch in the middle (Stage 3) and at the end (Stage 6) of the pipeline, as shown in Figure 22.3.2. For long pipeline units (without tight loops) in a microprocessor, the additional cycle of latency introduces very little system-level performance degradation. When the system is configured in 6-stage mode, the extra latches are set to let data flow through. In 7-stage mode, the two latches connected to the pipeline add two half stages. These extra stages are purely used for time borrowing to provide almost one cycle of timing slack into the overall pipeline. Clock selection circuits feed each latch with the proper clock phase, as shown in Figure 22.3.2.

With two supply voltages, one concern is the potential for increased static current at the voltage domain boundaries. If a VddL stage drives a VddH stage, the interface PMOS transistors connected to the VddH domain may not fully shut off, resulting in short-circuit current. Figure 22.3.3 plots the static power measured for the FPU when set to a worst-case voltage configuration to highlight this problem. The amount of short-circuit power consumption depends on ΔV , as well as VddH. For ΔV less than 200mV (less than V_{tp}), the increase in static power consumption is negligible and dominated by leakage. Measured results show that $\Delta V \approx 200$ mV is sufficient to enable $\sim 30\%$ frequency tuning and cover large delay variations. Hence, the design does *not* use level shifters and avoids associated overheads. At low voltages and large ΔV settings, circuit operation fails.

Measured results of the frequency tuning provided by voltage interpolation for the 6-stage mode are presented in Figure 22.3.4. The voltage-interpolated configurations use two power supplies: $VddH = VddNom + \Delta V/2$ and $VddL = VddNom - \Delta V/2$. The max fre-

quencies measured for all 64 voltage configurations at 8 different VddNom and ΔV settings are overlaid onto the traditional frequency tuning versus nominal voltage curve (dark line). Voltage interpolation provides a well-distributed frequency tuning range about the nominal frequencies. This tuning range depends on the selection of ΔV and the nominal voltage. By linearly scaling ΔV with respect to nominal voltage, $\sim 30\%$ frequency tuning range is achieved across all nominal voltage levels. Figure 22.3.5 presents a scatter plot of the measured power versus the delay for all of the configurations in Figure 22.3.4 and demonstrates good tracking with respect to the nominal power-delay curve. The zoomed-in region of the plot shows that some voltage configurations achieve equivalent frequency with lower power consumption.

Figure 22.3.4 (inset) shows the effectiveness of voltage interpolation to combat variability across 15 measured FPU chips. The maximum frequency and power consumption of each FPU with a single 1V supply is plotted, showing frequency and power variations around a 240MHz median frequency. With voltage interpolation, where all FPUs use the same VddH (1.085V) and VddL (0.915V), all FPUs can be binned to one median frequency with minimum power for each. The slowest FPU (#14) can be sped up at the expense of higher power. A faster FPU (#2) can trade frequency for reduction in power. These results show that voltage interpolation can be an effective post-fabrication performance-tuning knob to combat process variation.

Variable-latency operation further mitigates the effects of process variation and saves energy when combined with voltage interpolation. If a 6-stage FPU fails to meet timing due to large delay variations, the 7-stage mode provides 17% additional frequency headroom and the opportunity to reduce power. Figure 22.3.6 shows the measured power-delay space for a 7-stage FPU with voltage interpolation, and compares to the power-delay curves (dashed lines) for 6-stage and 7-stage modes generated by sweeping a single nominal voltage. To achieve the same delay, the 7-stage pipeline consumes less power, and the voltage-interpolated configurations again scatter close to the nominal 7-stage power-delay curve. Figure 22.3.6 (inset) plots the measured power in 7-stage mode operating at 233MHz across the 64 voltage configurations. Configuration #64 saves 12.5% of power when compared to the 6-stage FPU with a fixed 1V supply.

Leveraging time borrowing in latch-based designs, voltage interpolation offers fine-grain *effective voltage* tuning with only two supply voltages. Variable latency provides an additional knob to combat process variations. This tunability is important for variation-tolerant design since different units on the same chip may have localized worst-case operating frequencies that deviate from the nominal. Combining voltage interpolation with traditional voltage-frequency binning covers both fine- and coarse-grain variations and global adjustment of VddH and VddL balances the current loads on the two supplies. A die micrograph with floorplan overlay is shown in Figure 22.3.7.

Acknowledgements:

This work is funded by NSF CCF-0429782. We thank D. Kahn and M. Hempstead for help in testing and UMC for chip fabrication.

References:

- [1] K. Usami, M. Horowitz, "Clustered voltage scaling technique for low-power design," *Int. Workshop on Low Power Design*, Apr. 1995.
- [2] J. Tschanz, et al., "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *ISSCC Dig. Tech. Papers*, Feb. 2002.
- [3] H. Li, et al., "Combined circuit and architecture level variable supply-voltage scaling for low power," *Trans. VLSI*, May 2005.
- [4] C. Tran, et al., "95% Leakage-Reduced FPGA using Zigzag power-gating, dual-Vth/Vdd and micro-Vdd-hopping," *ASSCC Dig. Tech. Papers*, Nov. 2005.
- [5] K. Agarwal, K. Nowka, "Dynamic Power management by combination of dual static supply voltages," *Int. Symp. Quality Elec. Design*, Mar. 2007.

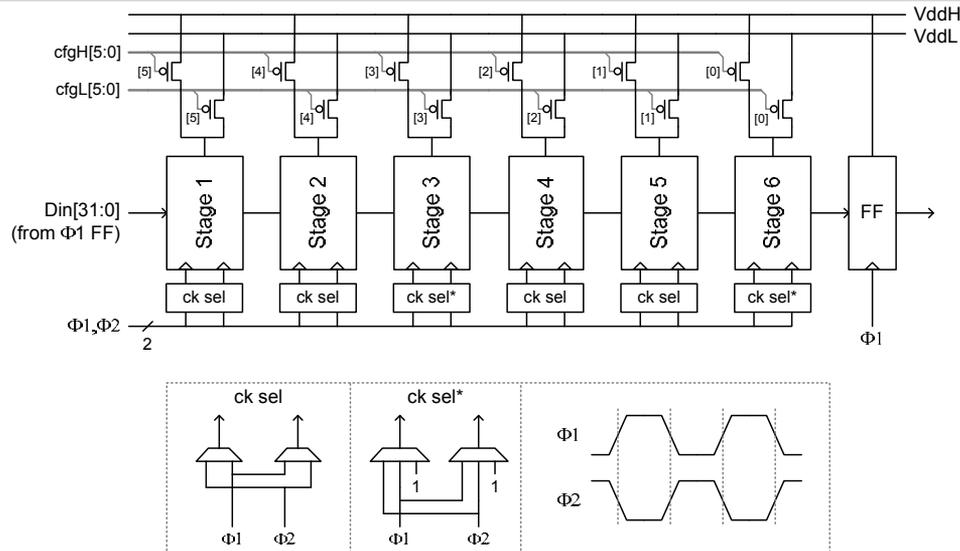


Figure 22.3.1: Pipelined FPU block diagram with per-stage Vdd and clock selection circuitry.

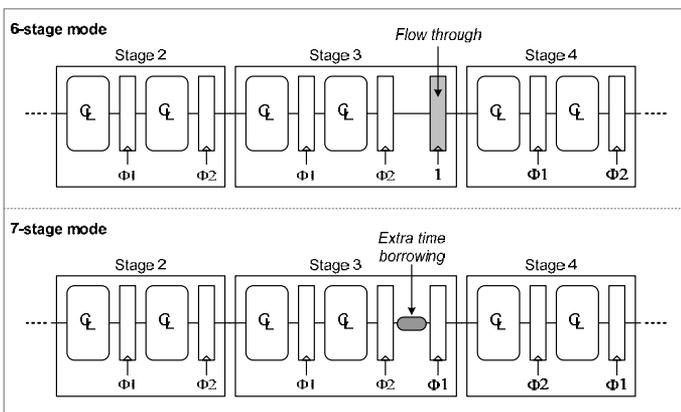


Figure 22.3.2: Variable latency clocking schemes for 6-stage and 7-stage modes. Illustrates extra time borrowing for 7-stage mode. Only 3 out of the 6 stages are shown.

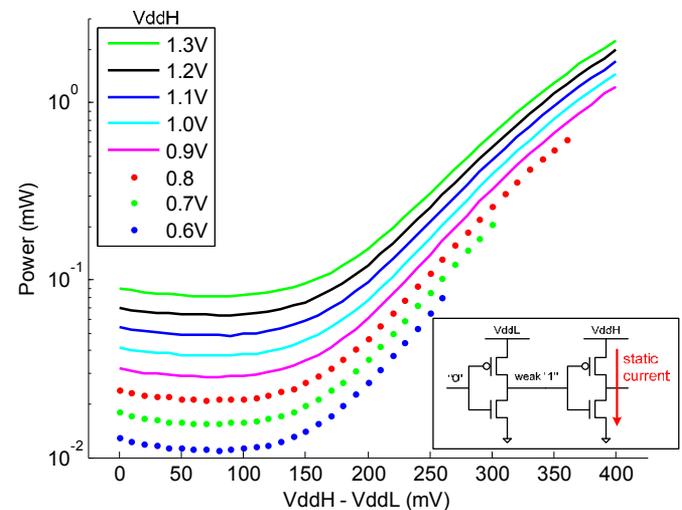


Figure 22.3.3: Static power vs. ΔV vs. VddH settings for worst-case voltage interpolation setting. Data points corresponding to inoperable voltage settings omitted.

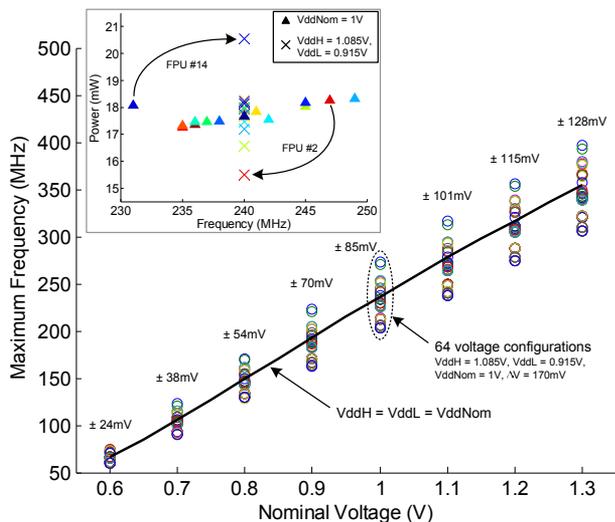


Figure 22.3.4: Maximum frequency vs. voltage with interpolation for 6-stage pipeline.

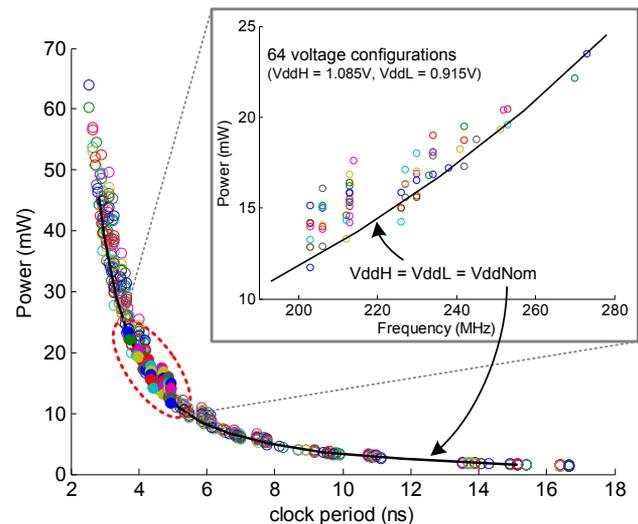


Figure 22.3.5: 6-stage pipeline power vs. clock period with voltage interpolation.

Continued on Page 623

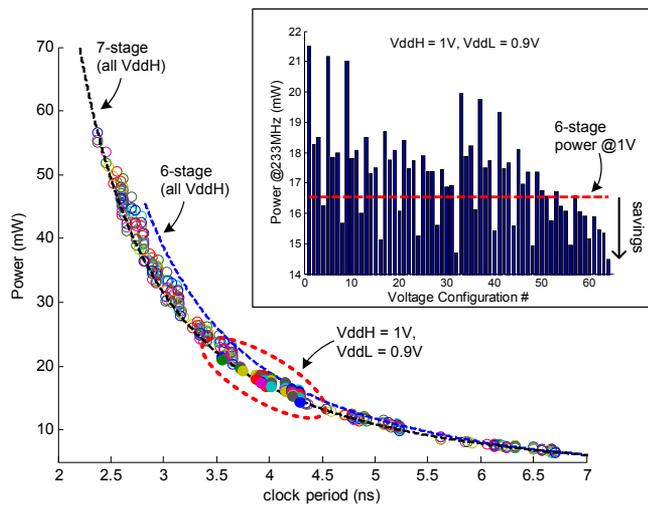


Figure 22.3.6: Power vs. clock period for all 7-stage voltage configurations across multiple voltage settings. Power savings shown for variable latency (7-stages) with voltage interpolation.

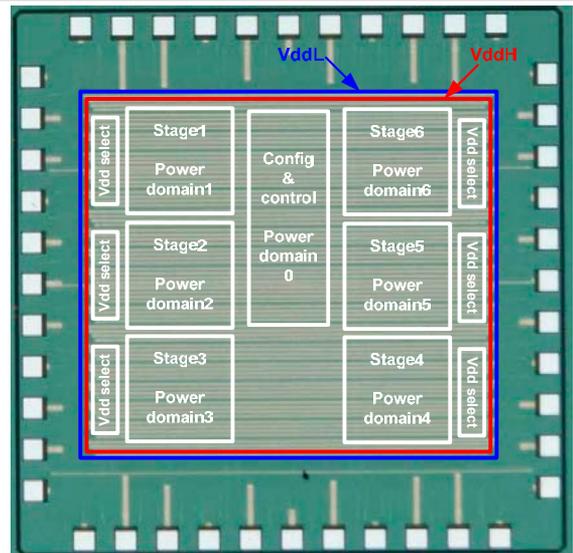


Figure 22.3.7: Die micrograph with floorplan overlay.