

## 16.4 A 180mV FFT Processor Using Subthreshold Circuit Techniques

Alice Wang, Anantha Chandrakasan

Massachusetts Institute of Technology, Cambridge, MA

The key design metric in emerging applications such as wireless sensor networks, is the energy dissipated per function rather than clock speed or silicon area. The author's previous energy-scalable FFT ASIC uses an off-the-shelf standard-cell logic library and memory only scaled down to 1V operation [1]. This paper describes a custom real-valued FFT processor that operates over a variety of operating scenarios (programmable FFT length and bit precision) and employs circuit techniques that allow the supply voltage to be deeply scaled into the subthreshold regime for minimal energy dissipation.

As processing speed requirements are relaxed, the supply voltage can be scaled down well below the threshold voltage to minimize switching energy. However, at low clock frequencies, leakage energy dissipation can exceed active energy, leading to an optimal operating frequency and voltage that minimizes energy consumption. To investigate the optimal operating point for the FFT, logic and memory design techniques allowing subthreshold operation are needed. Previous research demonstrates the functionality of logic circuits at 200mV using low threshold devices [2]. This FFT processor operates at 180mV using a standard CMOS 0.18 $\mu$ m logic process with threshold voltages of around 450mV.

The 16b architecture of the FFT is shown in Figure 16.4.1. After the input data is reordered and clocked into the data memory, one N-point real-valued FFT is performed. In one clock cycle, two 32b complex values (A,B) are read from the data memory, and the datapath outputs (X,Y) are written back to the memory. In addition, one 32b complex twiddle factor (W) is read from the ROM. The 512-Word, 32b memory bank for the FFT is segmented by address parity and MSB to avoid read/write memory hazards. Additionally, the memory is configured to allow for variable FFT lengths. The 16b hardware for both memory and datapath logic is reused for 8b processing. In Fig. 16.4.1, the LSB inputs to the 16b Baugh-Wooley multiplier are gated to configure the multiplier for energy-efficient 8b processing.

For ultra-low voltage operation, there are new circuit design considerations. As the supply voltage decreases, a CMOS inverter may not achieve rail-to-rail output voltage swing due to reduced  $I_{on}/I_{off}$ . An increase in  $W_p/W_n$  causes larger PMOS drive currents and improves the output-high swing but degrades the output-low voltage level by increasing PMOS leakage currents. This effect is further compounded by process variations. At the FS corner, the Fast NMOS is more leaky than the Slow PMOS leading to a higher bound on  $W_p(\min)$ . Similarly, the SF corner sets  $W_p(\max)$ . Figure 16.4.2 shows  $W_p(\min)$  and  $W_p(\max)$  of the inverter at process corners assuming a 10-90% output voltage swing and for  $W_n=0.44\mu$ m. The worst-case minimum supply voltage for this cell is estimated at 195 mV when  $W_p=4.8\mu$ m where the two curves intersect.

Parallel leakage, sneak leakage paths, and stacked devices, all create problems for traditional logic circuits in deep subthreshold operation. For example, Fig. 16.4.3 shows the operation of a standard tiny XOR logic gate. When operating at normal voltage levels, for A=1, B=0, the output node, Z, is high. However, when the voltage is scaled down to 100mV, the output voltage is

degraded by three leaking devices and reduced swing input voltages (due to imperfect inverters). Alternatively, a transmission gate XOR has fewer parallel devices which improves subthreshold performance at worst-case input vectors. Additionally, having both NMOS and PMOS in the pull-up and pull-down reduces the effects of process variations on minimum voltage operation. Sneak leakage paths between standard cells are minimized by introducing inverters and buffers and by carefully analyzing interfaces between standard cells. In multiple stacked devices, the drive current is significantly reduced in subthreshold operation, so subthreshold transmission-gate MUXes cannot be directly cascaded. Datapath and control circuits for the subthreshold FFT processor are developed by minimizing stacked devices, reducing parallel leakage, and avoiding sneak leakage paths.

Memory design using subthreshold operation is challenging. Conventional SRAM designs will not function at low voltage due to reduced  $I_{on}/I_{off}$  and bitline leakage that depends on the values stored in memory. For read access in deep subthreshold operation, the bitline is segmented by using a MUX-based hierarchical approach (Fig. 16.4.4). The selectors to the muxes are the read-address inputs, and the data from the memories is hierarchically passed through the MUXes to the output. The MUXes are designed to ensure a high  $I_{on}/I_{off}$  at each level of hierarchy by avoiding parallel leakage and stack effects. The simulation in Fig. 16.6.4 contrasts operation of the hierarchical read bitline with a conventional read bitline. The MUXes can be daisy-chained and arrayed for compact layout. The same hierarchical design is used to create subthreshold Twiddle ROMs. A latch-based circuit is used for reliable write access at very low voltages and process corners (Fig. 16.4.4).

The low-voltage FFT containing 627k transistors is fabricated in a standard 0.18 $\mu$ m 6M CMOS process. It is fully functional at 128 to 1024 FFT lengths, 8 and 16b precision, for voltage supplies 180 to 900mV and for clock frequencies of 164Hz to 6MHz. The minimum supply voltage is 180mV where it dissipates 90nW. Figure 16.4.5 is a oscilloscope plot of outputs from the FFT chip functioning at 180mV. The optimal operating point is where energy is minimized and is a function of activity factor and process technology. The optimal operating point is at 350mV with a clock frequency of 9.6kHz and is shown in Fig. 16.4.6. This figure is a plot of the energy and the performance for a 16b, 1024 point FFT as a function of  $V_{DD}$ . As previously reported, a low power FFT processor implemented in a 0.7 $\mu$ m process dissipates 3.4 $\mu$ J when performing one 1024-point CVFFT at 1.1V [3]. The energy used by this FFT processor to compute one 16b, 1024 point RVFFT at the optimal operating point is 155nJ. Figure 16.4.7 shows a die photo of the IC that occupies 2.6mm x 2.1mm.

### Acknowledgments:

The authors thank J. Cline for her help with the multiplier design. We also thank B. Calhoun and Prof. K.C. Smith for valuable feedback on the paper. This effort is sponsored by DARPA Power Aware Computing and Communications (PAC/C) and the Air Force Research Laboratory, under agreement number F33615-02-2-4005. A. Wang is supported by an Intel PhD Fellowship.

### References:

- [1] A. Wang and A. Chandrakasan, "Energy-Aware Architectures for a Real-Valued FFT Implementation," ISLPED 2003, pp. 360-365, August 2003.
- [2] J. Burr and J. Shott, "A 200mV Self-Testing Encoder/Decoder Using Stanford Ultra-Low-Power CMOS," *ISSCC Dig. Tech. Papers*, pp. 84-85, Feb. 1994.
- [3] B. Baas, "A Low-Power, High-Performance, 1024-Point FFT Processor," *IEEE J. Solid-State Circuits*, vol 34, no 3, pp. 380-387, March 1999.

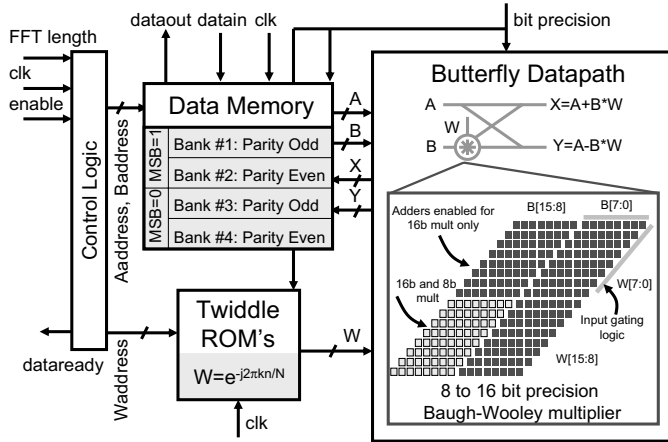


Figure 16.4.1: RVFFT architecture that enables scalability in bit-precision and FFT length, and includes circuits which can scale down to 180 mV operation.

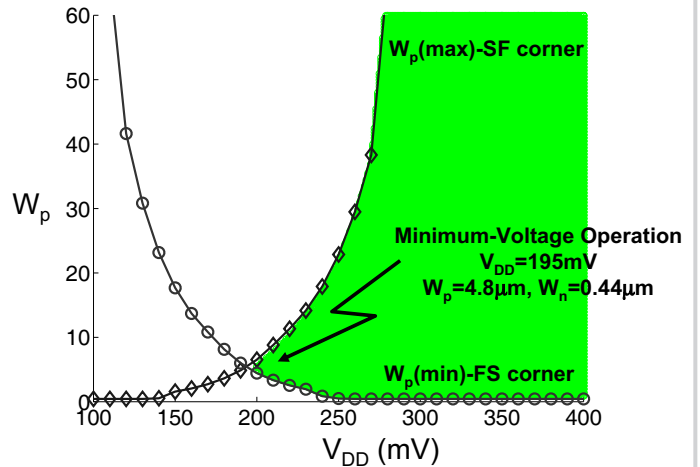


Figure 16.4.2: Sizing trade-off for an inverter at the minimum operating voltage with process variation considerations given  $W_n=0.44\mu\text{m}$  (simulation).

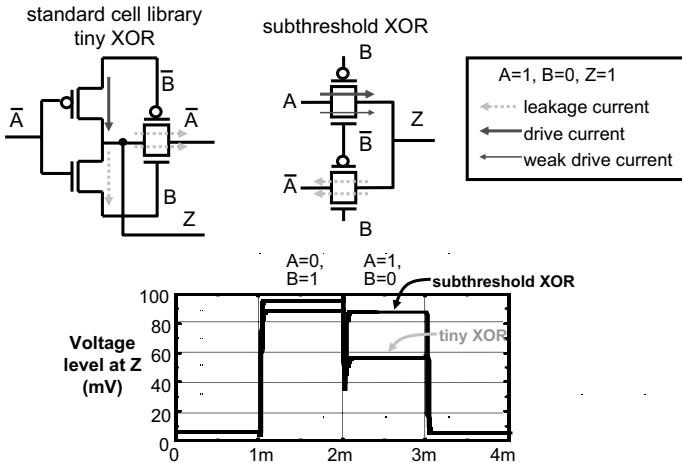


Figure 16.4.3: The effects of parallel leakage is compounded at ultra-low voltages as shown by the standard-cell tiny XOR gate for the inputs  $A=1$  and  $B=0$  at  $V_{DD}=100\text{mV}$ . Parallel leakage is reduced in the subthreshold XOR gate, which functions better at 100mV.

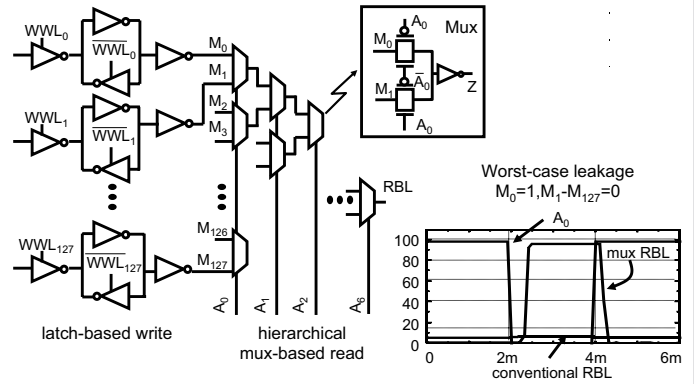


Figure 16.4.4: The MUX-based hierarchical-read access works reliably at 100mV in simulation compared to a conventional read bitline (RBL).

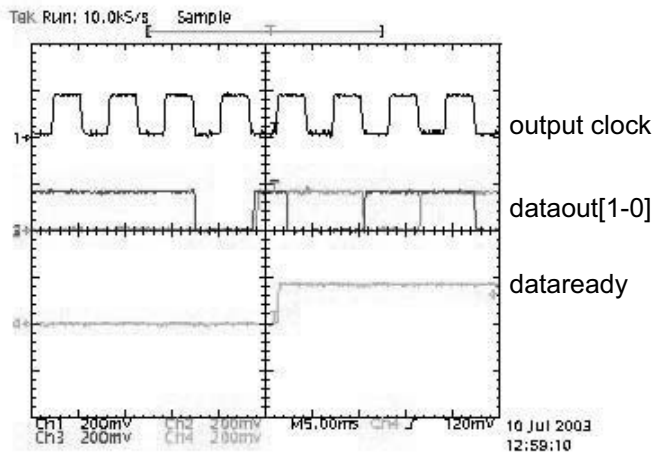


Figure 16.4.5: Oscilloscope plot showing outputs from the RVFFT chip at 180 mV operation.

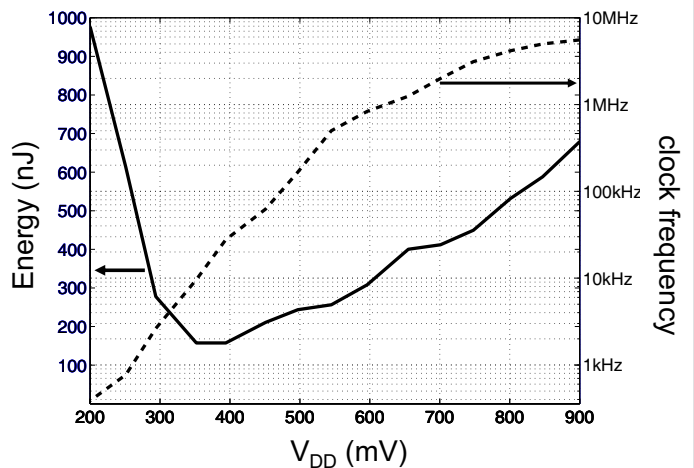


Figure 16.4.6: Energy and FFT clock frequency for 16b, 1024-point RVFFT as a function of  $V_{DD}$ .

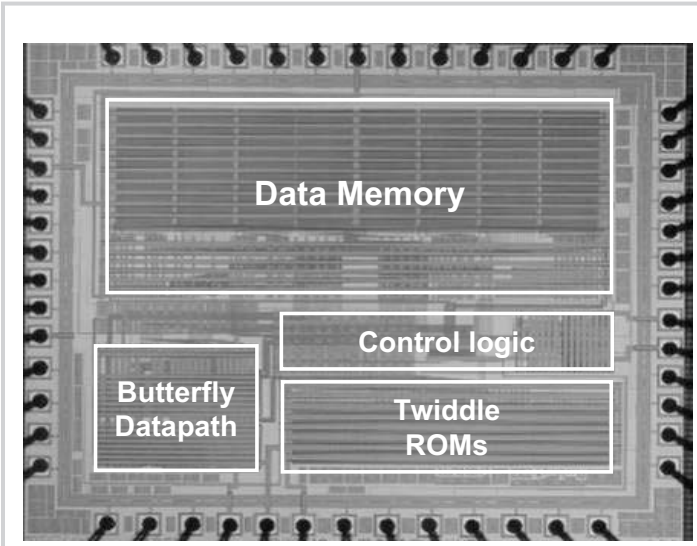


Figure 16.4.7: Die photograph of the 180mV real-valued FFT chip.

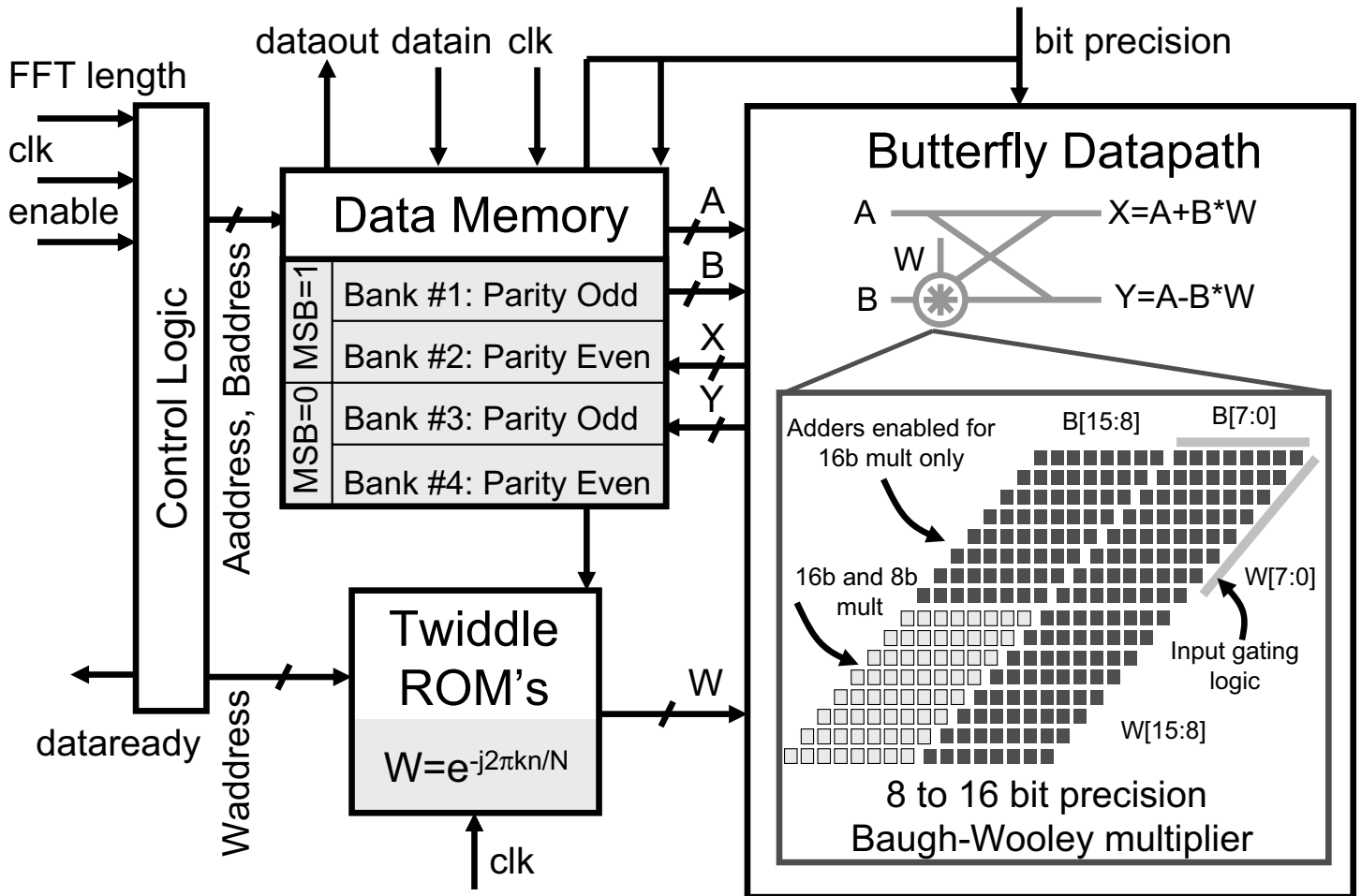


Figure 16.4.1: RVFFT architecture that enables scalability in bit-precision and FFT length, and includes circuits which can scale down to 180 mV operation.

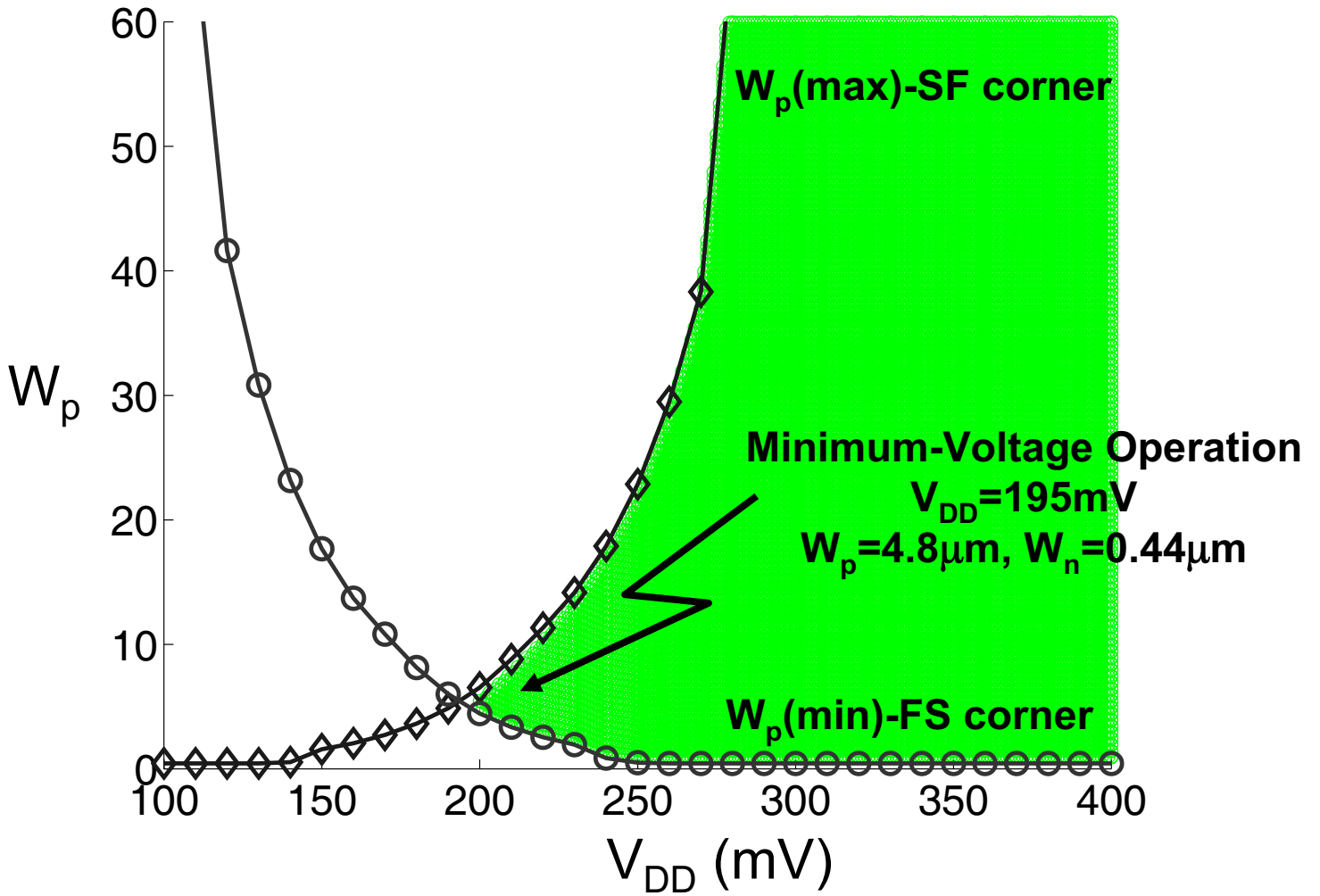


Figure 16.4.2: Sizing trade-off for an inverter at the minimum operating voltage with process variation considerations given  $W_n=0.44\mu\text{m}$  (simulation).

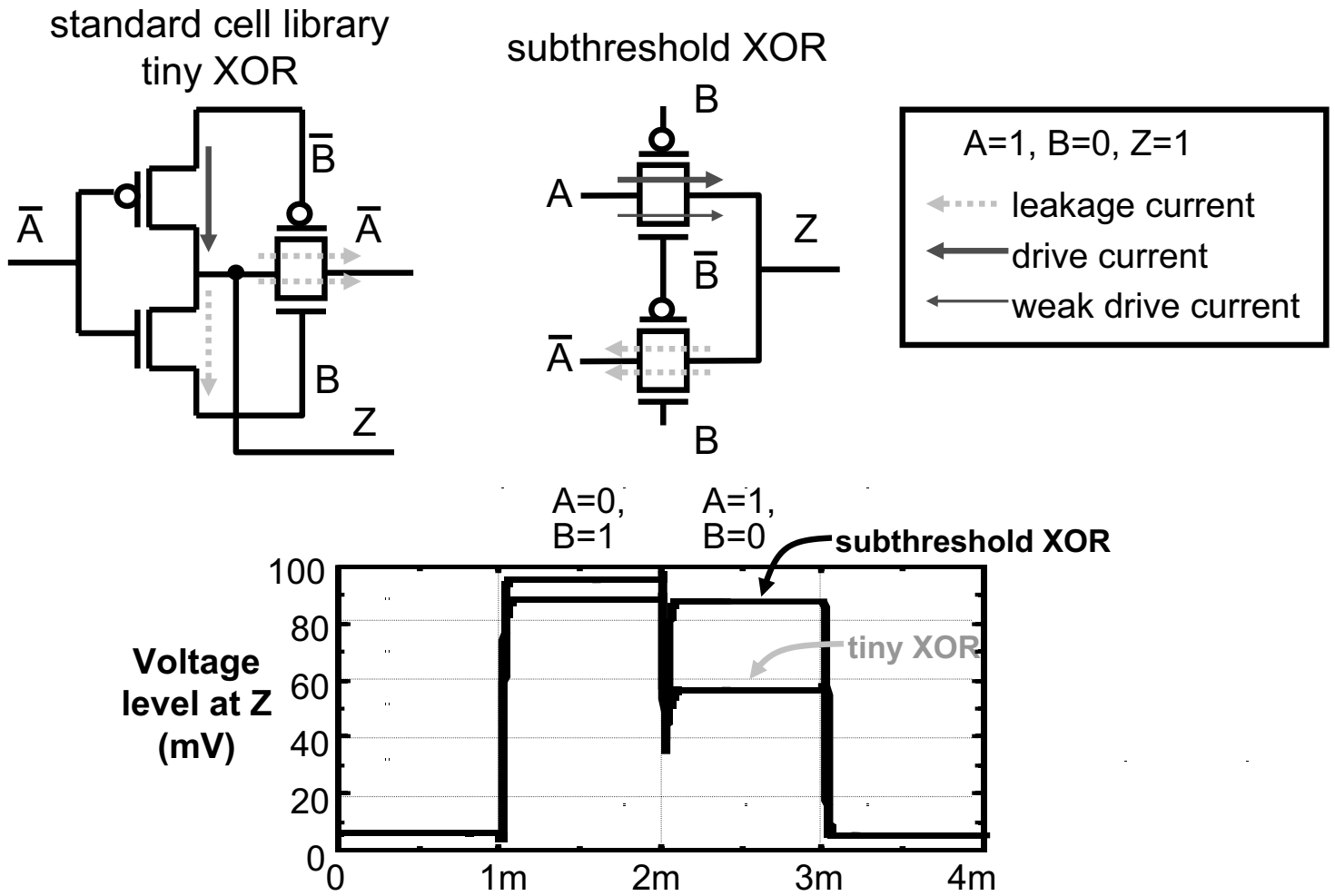


Figure 16.4.3: The effects of parallel leakage is compounded at ultra-low voltages as shown by the standard-cell tiny XOR gate for the inputs  $A=1$  and  $B=0$  at  $V_{DD}=100\text{mV}$ . Parallel leakage is reduced in the subthreshold XOR gate, which functions better at  $100\text{mV}$ .

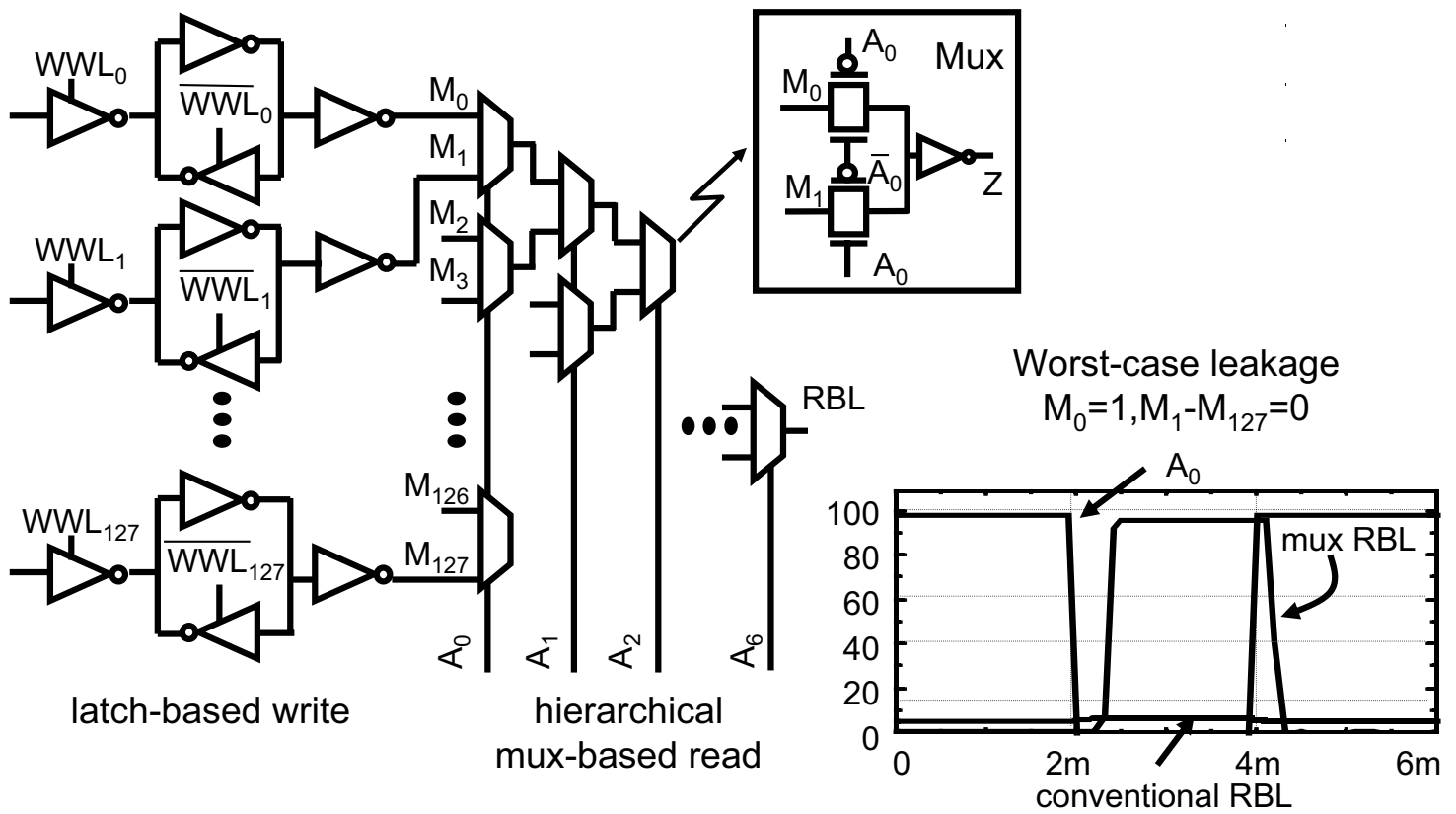


Figure 16.4.4: The MUX-based hierarchical-read access works reliably at 100mV in simulation compared to a conventional read bitline (RBL).

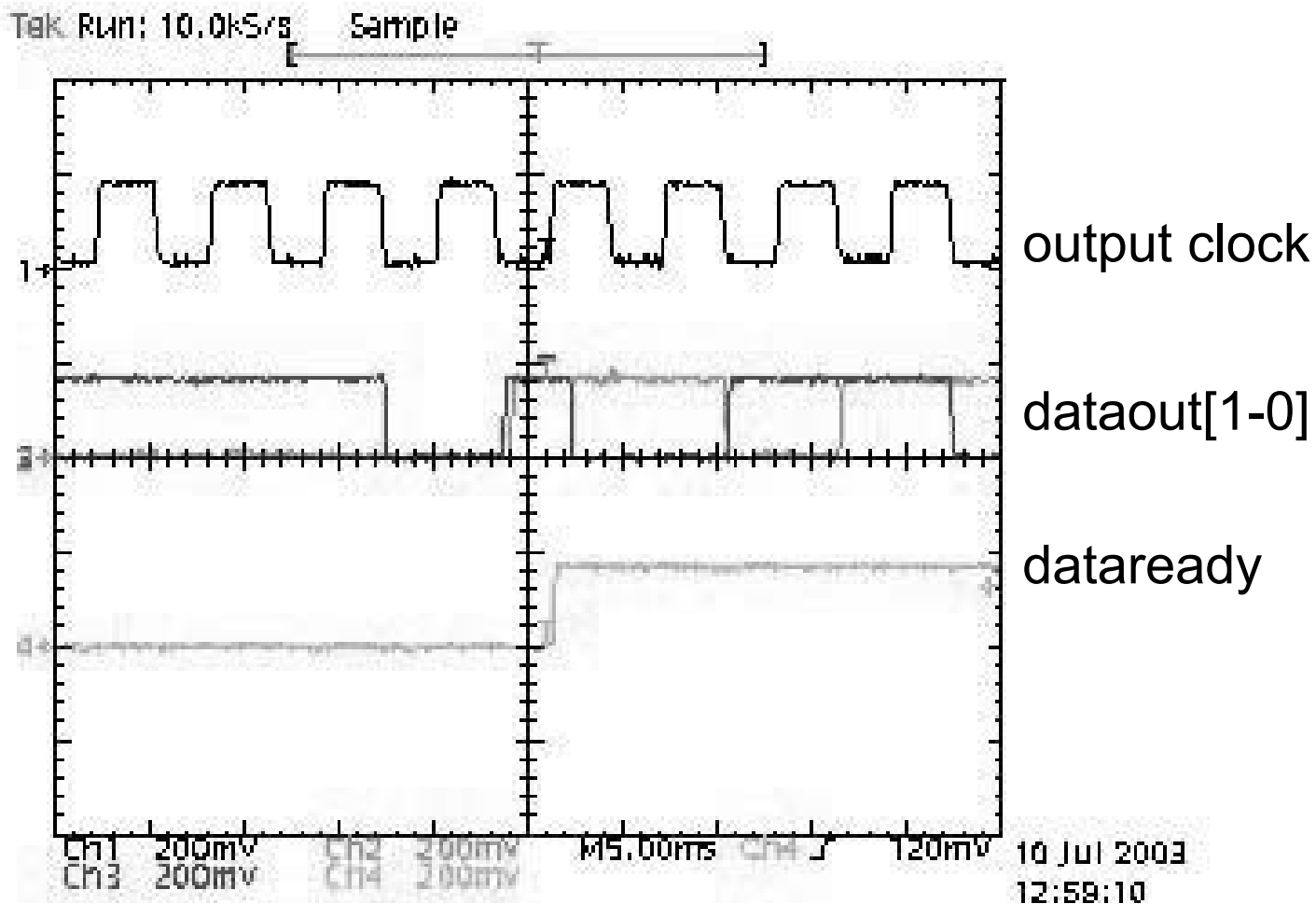


Figure 16.4.5: Oscilloscope plot showing outputs from the RVFFT chip at 180 mV operation.



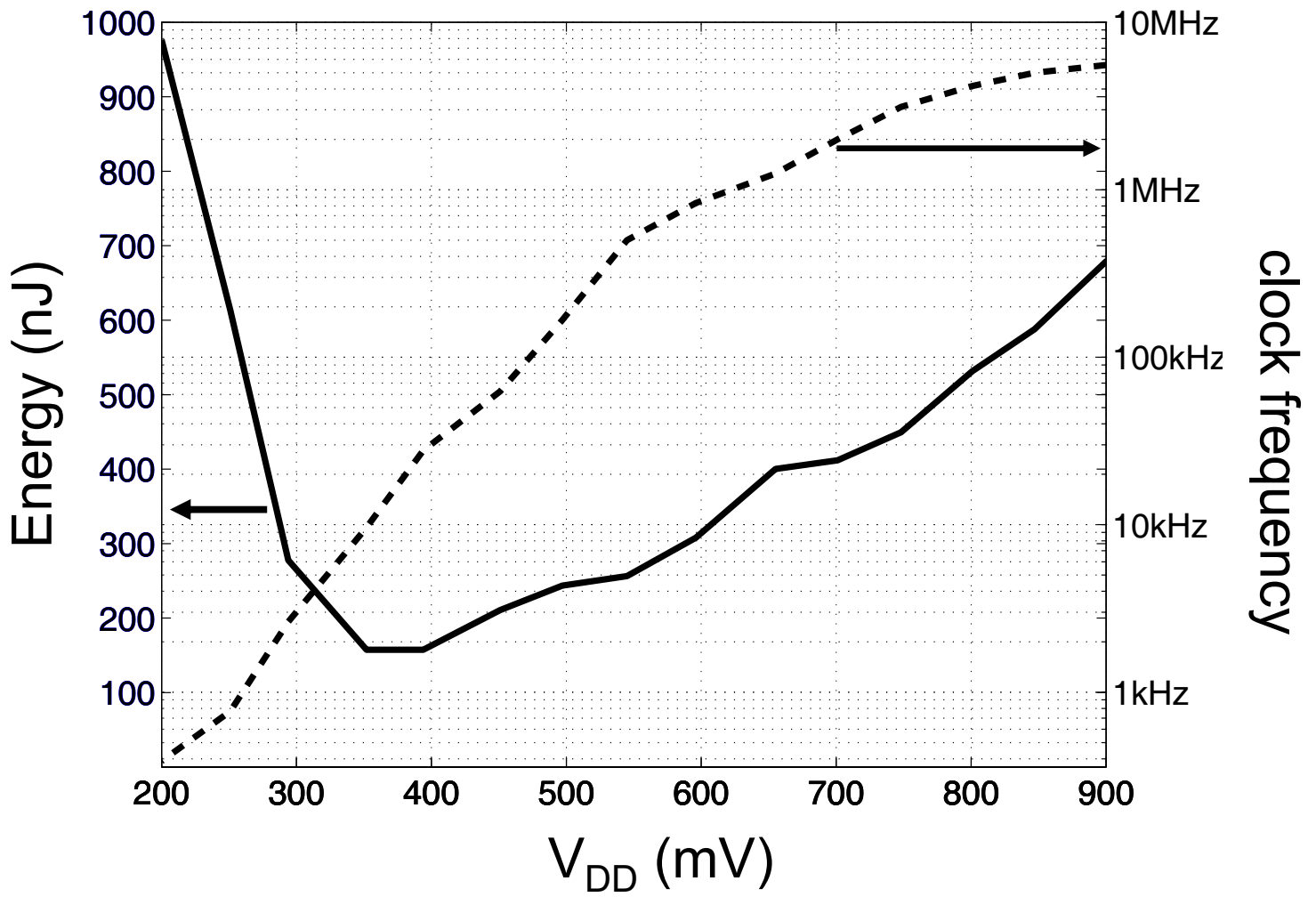


Figure 16.4.6: Energy and FFT clock frequency for 16b, 1024-point RVFFT as a function of  $V_{DD}$ .

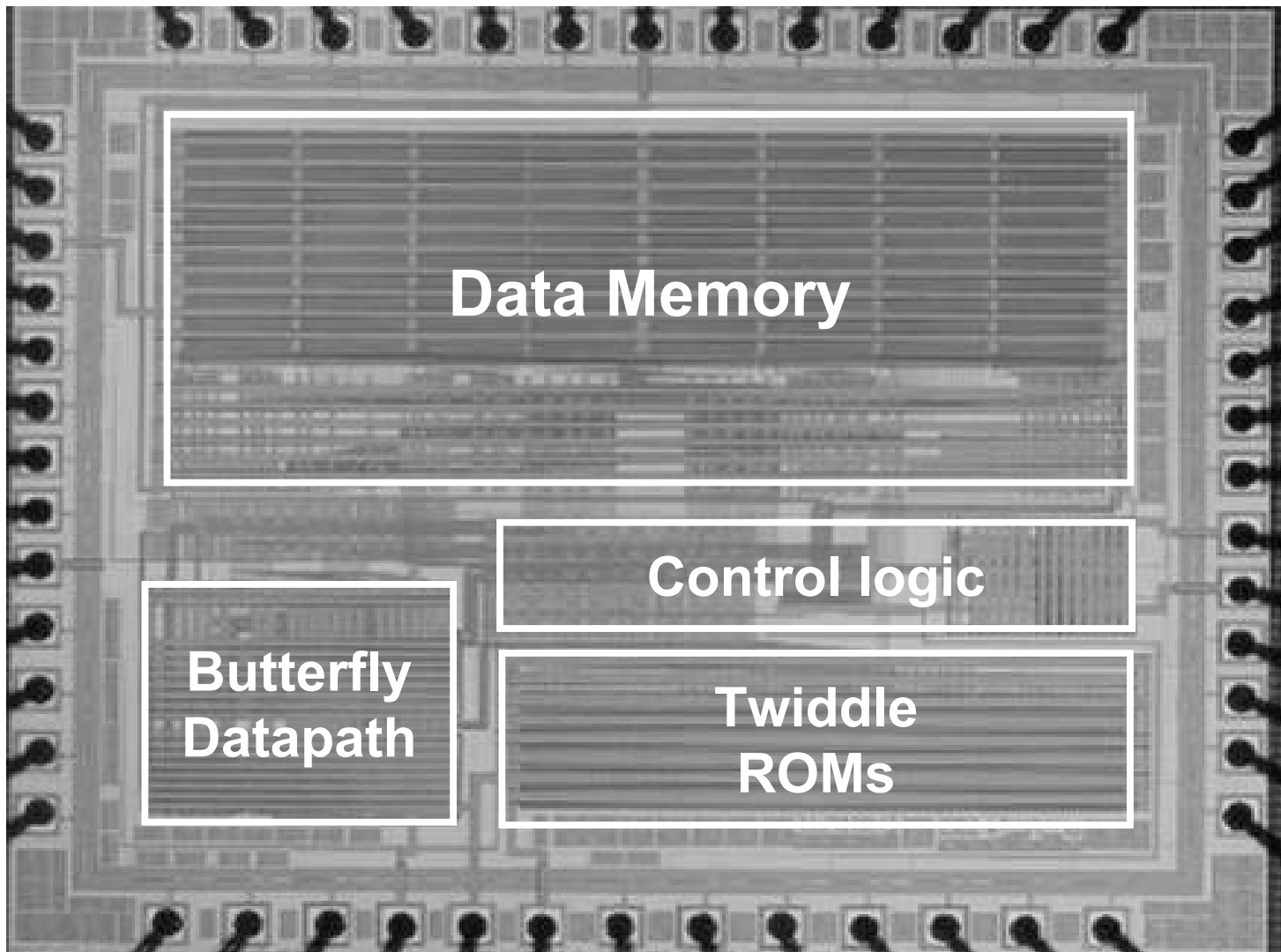


Figure 16.4.7: Die photograph of the 180mV real-valued FFT chip.