19.6 A 0.27V 30MHz 17.7nJ/transform 1024-pt Complex FFT Core with Super-Pipelining

Mingoo Seok[1], Dongsuk Jeon[1], Chaitali Chakrabarti[2], David Blaauw[1], Dennis Sylvester[1]

[1]University of Michigan, Ann Arbor, MI, [2]Arizona State University, Tempe, AZ

Recently, aggressive voltage scaling was shown as an important technique in achieving highly energy-efficient circuits. Specifically, scaling V_{dd} to near or sub-threshold regions was proposed for energy-constrained sensor systems to enable long lifetime and small system volume [1][2][4]. However, energy efficiency degrades below a certain voltage, V_{min}, due to rapidly increasing leakage energy consumption, setting a fundamental limit on the achievable energy efficiency. In addition, voltage scaling degrades performance and heightens delay variability due to large I_d sensitivity to PVT variations in the ultra-low voltage (ULV) regime. This paper uses circuit and architectural methods to further reduce the minimum energy point, or E_{min}, and establish a new lower limit on energy efficiency, while simultaneously improving performance and robustness. The approaches are demonstrated on an FFT core in 65nm CMOS.

Pipelining is a well-known method to improve performance or to enable limited energy savings by trading gained performance through lowering V_{dd}. However, in this paper, we make the counterintuitive observation that inserting additional pipeline latches improves both energy efficiency and performance in the ULV operating regime. Since pipelining shortens the clock period, it limits leakage energy consumed by idling gates, which reduces energy consumption and allows further voltage scaling. Simulations of inverter chains show that reducing stage depth from 65 to 11 fanout-of-four (FO4) delays yields 36% energy savings and a V_{min} reduction from 0.37 to 0.26V (Fig. 19.6.1 upper left). By applying this "super-pipelining" approach to the multipliers in an FFT core, we find that it consumes minimum energy when pipelined in 6 stages at a stage depth of 17 FO4 delays. This design approach differs radically from conventional ULV designs, which tend to use limited pipelining and typically have cycle times in the 50-to-200 FO4 range [1,2]. In this paper, we also show how clocking overhead can be reduced through circuit techniques to facilitate super-pipelining while process variation is addressed through the use of latch-based design. Additionally, architecture modifications are proposed to improve energy efficiency and throughput. Measurements show that the FFT core consumes 17.7nJ per 1024-pt complex FFT while operating at 30MHz at V_{dd}=0.27V, demonstrating an improvement over the FFT energy efficiency reported in [2-4].

An important principle driving our ULV design methodology is to suppress leakage energy, allowing for larger potential energy savings by enabling further voltage scaling. We first address this by architectural modifications through minimizing idling modules (Fig. 19.6.2). In a traditional memory-based FFT (Fig 19.6.2, bottom right), most memory cells idle while a single butterfly unit processes data word-by-word over many clock cycles. These idling cells increase leakage energy, harming energy efficiency and voltage scalability. On the other hand, conventional pipeline architectures such as MDC (Multi-path Delay Commutator) have high memory utilization but instead suffer from low butterfly unit activity [5]. We therefore propose a modified MDC that accepts 4 inputs concurrently using a commutator configuration, enabling full utilization of both butterflies and memory elements. Additionally, we use two of the modified MDC lanes to double throughput and halve memory counts per lane, reducing leakage energy consumption from memories. As shown in Fig. 19.6.1, these modifications improve energy efficiency and throughput by 2.8× and 6.2×, respectively, compared to a radix-4 memory-based FFT core.

Multipliers in the FFT are super-pipelined as shown in Fig 19.6.3. To successfully employ super-pipelining, sequential element overhead must be limited. Six latches share a local clock driver to reduce clock load. The drivers also use minimum-width fingers that enhance drivability at iso-input capacitance due to smaller V_{th} from inverse narrow width effects. Two latches are embedded in a mirror adder to save two transistors per latch. Latches are upsized from min-width for robustness such that they pass corners and 2 million Monte-Carlo mismatch simulations, providing an estimated 99% chip-level yield with 10k latch instances per chip at 0.2V. We implement the multiplier along with an unpipelined baseline multiplier, separately from the FFT core. Measured results (Fig. 19.6.1 upper right) show that the super-pipelined multiplier operates at 18MHz at 0.225V. It is 1.6× faster while consuming 30% less energy than an unpipelined multiplier. At iso-V_{dd}, it operates 3.6× faster and consumes 18% less energy.

The FIFOs in the commutators contribute as much as 29% of the total FFT energy consumption in this architecture. To address this, we replace the address decoder with a cyclic address generator for reduced energy and use logic-based readout paths for improved performance, as shown in Fig. 19.6.4. Simulation results show that the FIFO design consumes 12% lower energy while improving performance by 20% over a memory with MUX-based readout. Positive-edge read and negative-edge write operations for preventing hold time violation are used.

Although the techniques above improve energy efficiency and performance, we must pay attention to delay variability and overall design robustness given the ULV design point. To this end, we use 2-phase latches rather than flip-flops. Although the stage depth is drastically reduced in super-pipelined designs, time borrowing removes hard boundaries in the pipeline, re-establishing averaging of process variations along long paths that are present in unpipelined designs. Fig. 19.6.1 shows Monte Carlo simulations on latch and flip-flop pipelined multipliers indicating that a latch-pipelined multiplier can absorb delay variations, leading to higher performance yield. In addition, variability-induced hold time violations must also be avoided to ensure functionality. We identify short paths, aided by the regular structure of multipliers, and add delay elements that incur a marginal energy overhead of 2.4% per multiplier. Padded short paths were verified to satisfy hold times using 150k Monte-Carlo simulations under random process variations and corners.

The clock distribution network is designed to suppress process-variation-induced skew and its resulting hold time violations. Conventionally, many clock buffers are used to mitigate RC mismatch. However, at low V_{dd} the mismatch in these buffers is exacerbated and contributes significant skew, while RC delay is small compared to gate delay. Therefore, we design a 3-level clock network where a reduced number of large buffers and matched RC interconnect are used. The lowest and middle levels of the clock network are implemented with minimum width thin interconnect while the top level uses fish-bone shaped, thick metal interconnect for low RC delay and good slew. Fig. 19.6.4 shows that the simulated worst-case RC mismatch is less than 0.15ns (0.14×FO4 at V_{dd}=0.27V).

The FFT core is fabricated in 65nm CMOS using the above circuit and architectural techniques. Measurements in Fig. 19.6.6 show that it computes 234k 16b 1024-pt complex FFTs per second. The clock frequency is measured as 30MHz with V_{dd}=0.27V compared to frequencies of 10's of kHz for typical ULV designs at the same supply voltage. The FFT consumes 17.7nJ/transform, which is 4× smaller than prior work when scaled for word width, FFT size (2× from real to complex and 4× from 256 to 1024pt) and technology [3].

Acknowledgement:
The IC fabrication support of STMicroelectronics is gratefully acknowledged. Authors also acknowledge the Multiscale Systems Center and Army Research Laboratory for their support in this work.

References:

[1] G. K. Chen et al., "Millimeter-Scale Nearly-Perpetual Sensor System with Stacked Battery and Solar Cells," *ISSCC Dig. Tech. Papers*, Feb. 2010.
[2] A. Wang et al., "A 180mV FFT Processor using Subthreshold Circuit Techniques," *ISSCC Dig. Tech. Papers*, 2004.
[3] Y. Chen et al., "A 2.4-Gsample/s DVFS FFT Processor for MIMO OFDM Communication Systems," *IEEE J. Solid-State Circuits*, May 2008.
[4] S. Sridhara et al., "Microwatt Embedded Processor Platform for Medical System-on-Chip Applications," *Symp. on VLSI Circuits*, 2010.
[5] Y. Jung et al., "New Efficient FFT Algorithm and Pipeline Implementation Results for OFDM/DMT Applications," *Trans. Consumer Electronics*, Feb. 2003.

 978-1-61284-302-5/11/$26.00

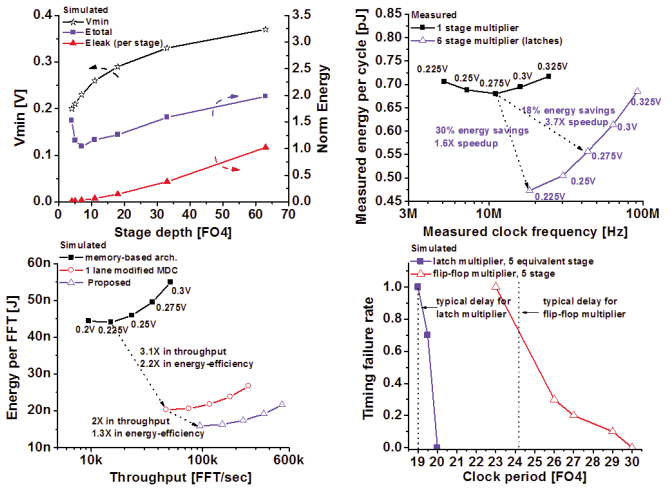


Figure 19.6.1: Inverter chain experiments, energy and performance of multipliers, architecture modification benefits, and multiplier variability.

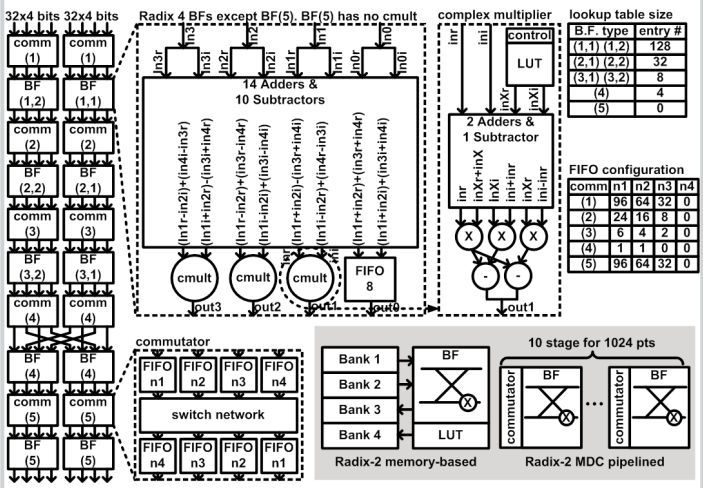


Figure 19.6.2: Pipelined, 8×32b input, radix-4, 2-lane, 1024-pt, complex FFT along with conventional architectures.

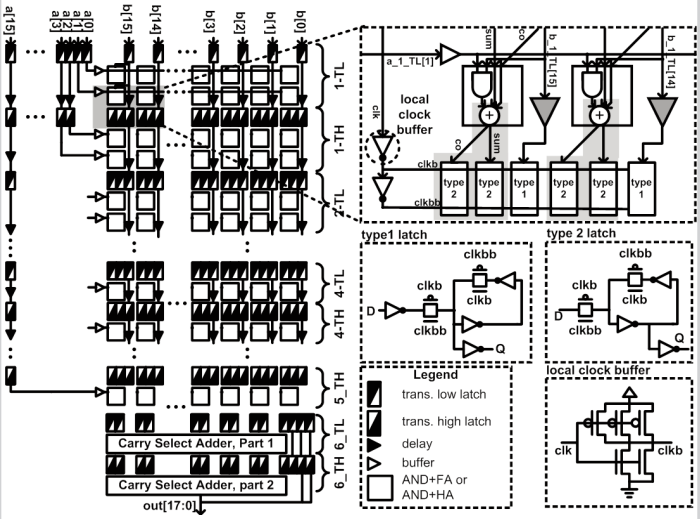


Figure 19.6.3: A 16b BW multiplier is pipelined with 12 banks of 2-phase latches. 5-4-3-2-2 length carry select adder is used for accumulation.

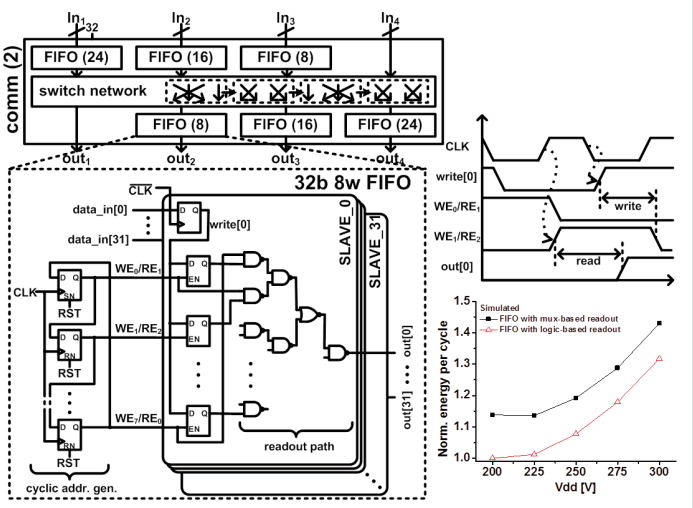


Figure 19.6.4: A commutator consists of a switch network and FIFOs. Positive-edge read and negative-edge write are described.

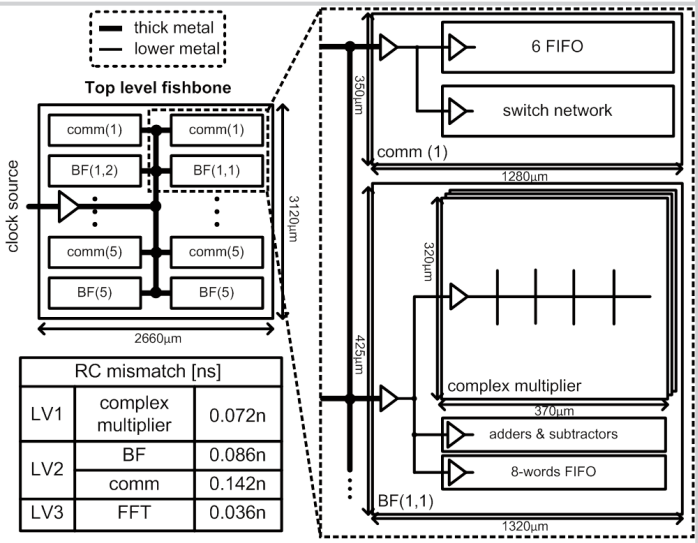
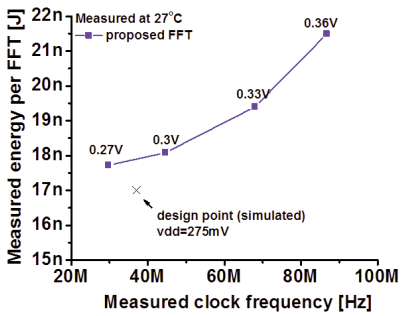
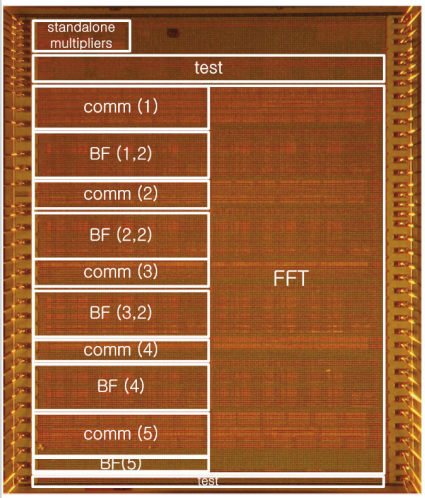


Figure 19.6.5: The clock network is designed with a limited number of buffers and matched interconnect to address key ULV skew sources.



	Proposed	[2]	[3]	[4]
Technology	65nm	180nm	90nm	130nm
Low power config	1024-pt CV 0.27V, 30MHz	1024-pt RV 0.35V, 10KHz	256-pt CV 0.85V, 300MHz	256-pt RV 0.5V, 7KHz
Energy/FFT	17.7 nJ	155 nJ	12.8 nJ	100 nJ
Norm. energy/FFT	**17.7 nJ**	**111.9 nJ**	**71.0 nJ**	**400 nJ**

Figure 19.6.6: The measured energy efficiency and performance of the FFT and comparisons normalized to technology and FFT type.



Type	1024pts complex
Vdd	270mV
Clock Freq.	30MHz
Energy / FFT	17.7nJ
Throughput	234k FFT/sec
Area	2.66x3.12mm^2
Technology	65nm CMOS

Figure 19.6.7: Die photo of the FFT core implemented in 65nm CMOS with summary table.

978-1-61284-302-5/11/$26.00 ©2011 IEEE