

# Explaining the Gap Between ASIC and Custom Power: A Custom Perspective

Andrew Chang  
Cadence Design Systems, Inc.  
2655 Seely Avenue, San Jose, CA 95134  
408-570-3714  
chang@cadence.com

William J. Dally  
Stanford University  
Gates CS Bldg. 3A-301, Stanford, University 94305  
650-725-8945  
billd@csli.stanford.edu

## ABSTRACT

Power dissipation is now both a key constraint and an application driver in VLSI systems. For a specific application, the energy efficiency of different implementations can differ by multiple orders of magnitude. This work surveys a range of techniques available to improve energy efficiency and highlights their cumulative benefit. Understanding, adopting and adapting selected techniques from full-custom solutions can help bridge the efficiency gap for the ASIC designs. Architecture and micro-architecture choices yield multiple-order of magnitude improvements in power dissipation by matching the structure of the design to the structure of the application and by providing multiple operating and power-down modes. The combination of methodology and full-custom circuit techniques and libraries provide benefits primarily due to reduced parasitic loading enabling the improved performance to be translated into the potential for factor-of-3 to factor-of-10 improvements in power.

## Categories & Subject Descriptors:

B.7.0 [Integrated Circuits]: General.

**General Terms:** Design, Experimentation, Performance

**Keywords:** ASIC, Custom Circuits, EDA, Energy Efficiency, Low Power, Normalized Metrics, Technology Scaling.

## 1. INTRODUCTION

Selective application of custom techniques can significantly reduce the power required by ASIC designs. For a specific application, the energy efficiency and resulting power dissipation of different implementations can differ by multiple orders of magnitude. Full custom solutions benefit from the ability to optimize across domains as a holistic combination of architecture, micro-architecture, design methodology, circuit styles and libraries and fabrication process leads to overall system efficiency. In contrast, ASIC and ASP solutions are traditionally constrained. While ASIC and ASP solutions can adopt architectures and micro-architectures similar to full-custom solutions, practical con-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DAC 2005, June 13-17, 2005, Anaheim, California, USA  
Copyright 2005 ACM 1-59593-058-2/05/0006...\$5.00.

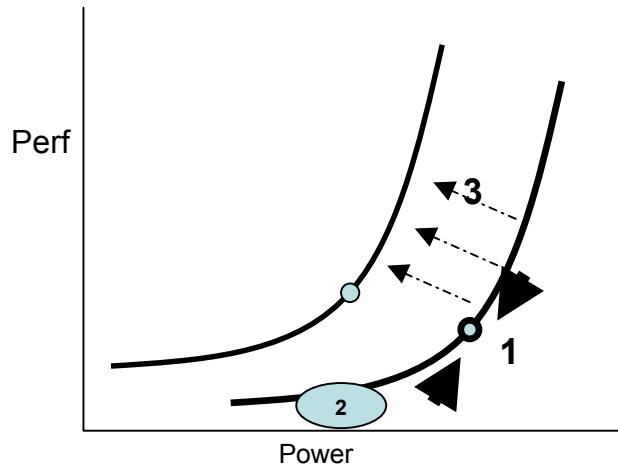


Figure 1. Basic Power Improvement Options.

siderations in execution, validation, and characterization currently prevent full realization of this potential. Custom designers have had three key advantages relative to their ASIC counterparts: they explicitly handle interconnect design, they have more flexibility in circuit styles and techniques to realize opportunities enabled by the architectural choices; and they already allocate substantial effort for characterization and verification of circuit operation at reduced supply voltages.

Every design has a unique power versus performance characteristic. Maximizing the energy efficiency of a design enables the minimization of power dissipation by creating the largest range of trade-offs between performance and power. Increasing the efficiency of a design and reducing the power necessary to deliver the required application performance are achieved by accomplishing one or more of the following three basic goals shown in Figure 1:

- 1) Moving along the curve towards a more efficient operating point.
- 2) Reducing power dissipation by operating at a lower-performance and lower-dissipation point.
- 3) Moving to a different power-performance curve by either changing the architecture or the process node.

**Table 1 E<sub>bit</sub> Energy**

Energy	180nm	130nm	90nm	65nm
E <sub>bit</sub> (fJ)	3.3	1.4	0.5	0.36
Relative	180nm	130nm	90nm	65nm
E <sub>bit</sub>	1	1	1	1
1b FO4	~10	~10	~10	~10
1b SP-SRAM	0.3-7	0.3-7	0.3-7	0.3-7
1b RF	4-20+	4-20+	4-20+	4-20+
1b DFF	20-30+	15-30+	10-30+	10-30+
1b Nand2	11-30 (typ 19)	5-30 (typ 14)	5-30 (typ 14)	5-30 (typ 14)
Move 1b 1000χ	~100	~100	~100	~100
Move 1b 1.5mm	268	367	467	714

## 2. NORMALIZED ENERGY METRIC AND A LOW POWER EXAMPLE

Two of the challenges in studies of low power designs are how best to compare different designs created with different implementation styles on different processes and how best to identify the full range of available and achievable power savings.

### 2.1 Basic Normalized Energy Metric

Throughout this work, we employ a normalized energy metric -E<sub>bit</sub> - as a reference unit. This metric is proportional to the energy required to store a binary value on a minimum sized SRAM bit cell for a given semiconductor process and can be estimated (with C<sub>bit</sub> approximated as 4 \* 2fF/um \* W<sub>min</sub> for the process).

$$E_{bit} = C_{bit} * V_{dd}^2$$

The first row of Table 1 summarizes E<sub>bit</sub> for four technology nodes from 180-nm to 65-nm. The remaining rows show the typical relative energy required for various simple operations: data storage (RF, SP-SRAM, and DFF), data transformation (Nand2) and data movement (1b move over either a normalized distance of 1000χ or over a fixed distance of 1.5-mm). The range of energies for the data storage (SP-SRAM, and RF) is based on size of the specific array as smaller arrays have larger relative E<sub>bit</sub> as there are fewer total bits to amortize the energy cost of accessing the array. The range for the logic gates (DFF and Nand2) is due to range of sizes available in commercial cell libraries.

### 2.2 Low Power 16b 1024-point FFT Example

The energy efficiency and energy-delay-product (EDP) for seven implementations of a 16b 1024-point FFT is provided to show that an almost five order-of-magnitude difference in power and over three order-of-magnitude difference in energy efficiency and EDP can exist between implementations of the same function. Note, the best performing design depends on the specific optimization goal (MIT FFT is the most energy efficient but Spiffee has the highest EDP). The custom MIT FFT processor employs subthreshold circuit techniques, libraries and design methodology [14]. The low power Spiffee FFT processor [2] employs high performance algorithm/architecture and low supply voltages. The StrongArm SA-1100 processor [7] employs custom circuits, clock gating and reduced supply voltages. The Stratix is

**Table 2 Energy and EDP 16b 1024-pt FFT**

Design	Fab	V <sub>dd</sub>	MHz	mW	Cycles
MIT FFT	180	1.8	0.01	1.6	95
Spiffee	700	3.3	173	845	5190
SA-1100	350	2	74	39	31500
Imagine	150	1.5	232	4000	3708
Stratix	130	1.3	275	884	1291
Intel P4	130	1.2	3000	51200	71680
TI 'C6416	130	1.2	720	1200	6526
Design	EDP (rel norm)	E <sub>bit</sub> (fJ)	E <sub>eff</sub> (nJ)	Normalized to E <sub>bit</sub> (1e6)	Energy Ratio
MIT FFT	143	3.3	154	47	1
Spiffee	1	91	25350	277	6
SA-1100	283	4.2	16601	3953	85
Imagine	148	2.2	63931	29726	637
Stratix	24	1.4	4149	2964	64
Intel P4	12548	1.4	1E+06	873813	18591
TI 'C6416	27	1.4	10877	7769	166

an FPGA with dedicated embedded FFT logic [10]. The Intel Pentium-4 [11] is a standard general purpose microprocessor. The Imagine [12] is a media processor and the TI 'C6416 [1] is a digital signal processor. Both the Imagine and the 'C6416 were created using pseudo-custom datapath tiling. In addition, the TI 'C6416 employs pass-gate multiplexor circuits. As shown in Table 2, the actual efficiency differences between implementations is smaller than the power dissipation difference once the designs are normalized for process technology. Nevertheless, a large range of variation still remains and provides the opportunity for improvements.

## 3. TECHNIQUES FOR ENERGY EFFICIENCY AND POWER REDUCTION

All low power techniques either reduce the dynamic energy dissipated by the system and/or minimize the static current. Architectural choices yield the greatest benefit, providing multiple-orders of magnitude improvement. While specific implementation choices yield less dramatic benefits, they still can provide up to a factor-of-10 improvement in energy efficiency.

### 3.1 Dynamic Energy Efficiency

The basic equation for digital circuit dynamic power consumption (assuming constant frequency clock and balanced number of 0-to-1, 1-to-0 transitions) is:

$$P_{dyn} = \alpha CV_{dd}^2 f = \alpha E_{circuit} f$$

Where α is the activity factor, E<sub>circuit</sub> is the average energy per operation of the circuit and f is the switching frequency. Specific techniques:

Reduce  $V_{dd}$  by: (1) static lowering of supply voltage, (2) dynamic lowering of supply voltage, (3) creation of distinct voltage islands and (4) supply gating.

Reduce  $\alpha$  and  $f$  by: (1) Explicitly disabling unnecessary portions of the chip through clock-gating and/or block enables, (2) dynamic frequency scaling, (3) bus bit encoding to reduce transitions and (4) glitch identification and elimination.

Reduce  $E_{circuit}$  by: (1) Minimizing parasitics by explicitly engineering the interconnect and matching loads with drive, (2) increasing efficiency of circuits (circuit techniques, cell libraries and memories), (3) reducing required energy of circuits by employing subthreshold circuit techniques.

### 3.2 Static Power Dissipation: Leakage

At semiconductor technology nodes below 180-nm, leakage power is an increasingly important contributor to overall design power and at nodes below 130-nm, leakage power can be the dominant component of power consumption in specific applications. Two main contributors are subthreshold leakage current ( $I_{sub}$ ) and gate-oxide leakage current ( $I_{ox}$ ). The basic equations for digital circuit static power consumption [3] are:

$$P_{static} = V_{dd} * (I_{sub} + I_{ox})$$

$$I_{sub} = K_1 W e^{-V_t/nV_\theta} (1 - e^{-V_{gs}/V_\theta})$$

$$I_{ox} = K_2 W (V_{gs}/t_{ox})^2 e^{-\alpha t_{ox}/V_{gs}}$$

where  $K_1$ ,  $K_2$ ,  $\alpha$  and  $n$  are experimentally determined and  $W$  is the transistor width,  $V_{dd}$  is the supply voltage,  $V_{gs}$  is the gate-to-source voltage,  $V_t$  is the threshold voltage, and  $V_\theta$  is the thermal voltage ( $kt/q$ , 25mV at 25°C). Specific techniques:

Reduce  $V_{dd}$  (same approach as in dynamic power reduction) by: (1) static lowering of supply voltage, (2) dynamic lowering of supply voltage, (3) creation of distinct voltage islands, and (4) supply gating.

Increase effective  $V_t$  by: (1) substituting high threshold devices in non-critical logic paths (MT-CMOS), (2) employing transistor stacking to generate negative body-to-source voltages, negative  $V_{gs}$  and reduce the effect of Drain-Induced Barrier Lowering (DIBL) on  $V_t$  and (3) introducing body-bias (either static or active) to increase the effective  $V_t$ .

Reduce effective  $W$  by: reducing the number and size of transistors within the design.

### 3.3 Summary of Potential Contribution of Low Power Techniques

Architectural choices have the greatest impact on the system's energy and power efficiency as they potentially enable the design to operate on an improved power-performance curve. Once the architecture is selected, careful implementation allow the efficiency gains and power savings to be fully realized.

#### 3.3.1 Architecture

An optimized chip architecture minimizes the energy overhead by using the minimum required resources for each operation and matches both the computational intensity and data/control movement of the design to the requirements of the specific application. Traditional application-specific designs hardwire the connection of these computation and communication resources. However, hardwired designs do not extend easily for alternate applications. Software solutions on general purpose processors

**Table 3 Correlation between Estimated and Reported CV/I**

Technology Node	CV/I est (ps)	CV/I reported (ps)	$t_{FO4}$ est (ps)
Foundry A 180-nm	3.94	3.70	53
Foundry A 130-nm	2.55	2.17	34
Foundry A 90-nm	1.85	2.04	25
Foundry A 65-nm	1.45	1.00	20

provide both application flexibility and time-to-market benefits but are the least energy efficient as exemplified in Table 2. Recent work in stream [12] architectures combines programmability with the energy efficiency of hardwired solutions. In addition, proactively disabling unnecessary parts of the design during operation and carefully selecting power down modes further improve energy efficiency.

#### 3.3.2 Implementation

Eliminating parasitic loading, optimizing interconnect and maximizing the energy efficiency of the underlying circuits are all keys in both improving overall performance [5][6] and enabling trade-offs to reduce power dissipation.

The importance of power dissipated to drive on-chip interconnects increases with technology scaling [9]. In microprocessor designs, up to 50% of the power is dissipated in the interconnect [9]. Detailed floorplanning and placement and explicit planning of routing can result in a factor-of-1.4 increase in performance (30% reduction in interconnect capacitance) due to the elimination of parasitic loading [5][9].

Custom designs benefit from both more efficient circuits and better load matching between circuits. There is a factor-of-1.7 improvement in performance due to circuit styles and techniques. Detailed attention to sizing in custom libraries results in an additional factor-of-1.4 improvement in loading over the standard cells used in ASIC circuits [5]. Similarly, SRAM Arrays can have over a factor-of-2 difference for the same array size based general or low-power implementation.

#### 3.3.3 Power versus Performance

Energy efficient design enables the trade-off of potential performance for reduced power as lowering the supply voltage results in a quadratic reduction in dynamic power and a linear reduction in static power with only a near linear ( $V_{dd-new}/V_{dd-orig}$ )<sup>1.25</sup> reduction in performance. The  $I_{dstat}$  of a foundry process largely determines the speed of the process. Below 180-nm, the  $I_{dstat}$  is limited by short channel effects and velocity saturation. In [4] the authors develop a simple model to estimate  $I_{dstat}$  under these additional constraints.

$$I_{dstat} = K_3 L_{eff}^{-0.5} t_{ox}^{-0.8} (V_{gs} - V_t)^{1.25}$$

The CV/I [13] of the process can be used to form an approximation that ties  $V_{dd}$  and  $V_t$  to  $t_{FO4}$ . In this estimate,  $K_4$  is 13.5 [8],  $C_{eff}$  is approximated to 2fF and  $V_{gs}$  is assumed to be, in the worst-case equal, to  $V_{dd}$ . The correlation of the estimate for CV/I and reported CV/I for a range of foundry processes is shown in Table 3.

$$t_{FO4} = K_4 [C_{eff} V_{dd} / I_{dstat}]$$

Custom-specific techniques can yield between a factor-of-1.5

**Table 4 Power Improvement from Implementation Techniques**

Technique	Type	Custom vs. ASIC	Energy	Type
Circuit Styles and Flops	Dynamic	1.7	0.815	Logic
Libraries + V <sub>dd</sub> Scaling		1.4	0.855	Logic
SRAM Circuits		2	0.95	SRAM
Interconnect + V <sub>dd</sub> Scaling		1.4	0.855	Interconnect
Bit Encoding		1	0.84	Interconnect
Clock Gating		1	0.84	Chip
Frequency Scaling		1	0.5	Chip
Subthreshold Circuits		N/A	0.062	Chip
V <sub>dd</sub> Scaling	Static	1	0.79	Chip
MT-CMOS		1	0.5	Chip
Stacking and input state vector		1.4	0.7	Chip (typically only one of these three is applied)
Body Bias		2	0.5	
Supply Gating		10	0.1	
Type	Tech	ASIC (Cust)	Tech	ASIC (Custom)
Net Dyn	130-nm	45% (32%)	90-nm	28%(20%)
Net Static		8% (4%)		20%(10%)
Total		53% (36%)		48%(30%)

and factor-of-2 reduction in energy relative to ASIC designs due to the additional options for circuits and explicit interconnect optimization. In addition, use of subthreshold circuit techniques and supply-gating can further extend the differences in achievable power savings to over an order-of-magnitude additional savings.

The power improvements from a range of techniques is surveyed in Table 4 and is organized into three parts – dynamic power reduction, static power reduction and combined impact for example designs in 130-nm and 90-nm. For dynamic power, the corresponding performance differences between Custom and ASIC [5] are shown, followed by the resulting power improvement due to V<sub>dd</sub> scaling (while maintaining a fixed performance). The fifth column indicates the specific power component reduced: logic, interconnect, or full-chip. The second section of the table provides similar data for static power reduction. The final section combines the dynamic and static savings in the context of two microprocessor chips – 130-nm with 80% dynamic and 20% static and 90-nm with 50% dynamic and 50% static power dissipation excluding subthreshold circuits and

supply gating. At 130-nm, the dynamic dissipation is reduced to 45% (32% custom) and 53% (36% custom) of the original. At 90-nm, the dissipation is reduced to 28% (20% custom) dynamic and 20% (10% custom) static.

#### 4. SUMMARY AND CONCLUSIONS

Custom designers can employ the full range of optimizations from architecture, microarchitecture, through circuits and process to improve the energy and power efficiency for the complete design by at least a factor-of-3 and with the potential of over a factor-of-10. Unlike ASIC designers, they have flexibility in circuit styles and techniques and the pre-existing practice of detailed circuit-level characterization and verification. Selective application of custom circuit techniques and explicit interconnect design combined with tools to automate the verification of operation at lower supply voltages can enable ASIC designers to bridge the gap between ASIC and Custom power.

#### 5. REFERENCES

- [1] Agarwala, S., et al. A 600MHz VLIW DSP. *IEEE Journal of Solid-State Circuits*, 37, 11 (November, 2002), 1532-1544.
- [2] Baas, B. A Low-Power, High-Performance 1024-point FFT Processor. *IEEE Journal of Solid-State Circuits*, 34, 3 (March, 1999), 380-387.
- [3] Chandrakasan, A., Bowhill, W., and Fox, F., *Design of High-Performance Circuits*. IEEE Press 2001.
- [4] Chen, K., et al. Predicting CMOS Speed with Gate Oxide and Voltage Scaling and Interconnect Loading Effects. *IEEE Transactions on Electron Devices*, 44, 11 (November 1997), 1951-1957.
- [5] Chinnery, D. G. and Keutzer, K., *Closing the Gap Between ASIC and Custom*. Kluwer Academic Press, Norwell, MA 2002.
- [6] Dally, W. J. and Chang, A. The Role of Custom Design in ASIC Chips. In *Proceedings of the 37th Design Automation Conference*, Los Angeles, CA, June 5-9 2000. 643-647.
- [7] Intel. *StrongARM SA-1100 Microprocessor for Portable Applications Brief Datasheet*. Intel, Chandler, AZ 1999.
- [8] ITRS. *International Technology Roadmap for Semiconductors 2001 Edition – System Drivers*. ITRS. 2001.
- [9] Magen, N. et al. Interconnect-Power Dissipation in a Microprocessor. In *Proceedings of the 2004 International Workshop on System-Level Interconnect Prediction* (Paris, France). 7-13.
- [10] Lim, S.Y. and Crosland, A. Implementing FFT in an FPGA Co-Processor. In *The International Embedded Solutions Event (GSPx)*. Santa Clara, CA, September 27-30, 2004.
- [11] Rahal-Arabi, T. et al. Designing a 3GHz, 130nm, Intel Pentium 4. In *Digest of Technical Papers, Symposium on VLSI Circuits* (June 13-15, 2002), 130-133.
- [12] Rixner, S. et al. A Bandwidth-Efficient Architecture for Media Processing. In *Proceedings of the 31<sup>st</sup> Annual International Symposium on Microarchitecture (MICRO 31)* (Dallas, TX). 3-13.
- [13] Taur, Y., and Ning, T. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, Cambridge, CB2 1RP, United Kingdom 1998.
- [14] Wang, A., and Chandrakasan, A. A 180-mV Subthreshold FFT Processor Using a Minimum Energy Design Methodology. *IEEE Journal of Solid-State Circuits*, 40, 1 (January, 2005), 310-319.