The Mixed Signal Optimum Energy Point: Voltage and Parallelism

Brian P. Ginsburg
Texas Instruments
Dallas, TX
bginzz@ti.com

Anantha P. Chandrakasan
Massachusetts Institute of Technology
Cambridge, MA
anantha@mtl.mit.edu

ABSTRACT

An energy optimization is proposed that addresses the non-trivial digital contribution to power and impact on performance in high-speed mixed-signal circuits. Parallel energy and behavioral models are used to quantify architectural tradeoffs across the analog/digital boundary. An interleaved ADC is optimized as a case study to demonstrate this approach. The chosen operating point of 36 channels and 700mV operation gives a 3× improvement in energy compared to the seed of the model. The model matches closely the measured results of an ADC testchip implemented in a 65nm CMOS process.

Categories and Subject Descriptors

B.7.2 [**Integrated Circuits**]: Design Aids

General Terms

Design, Theory

Keywords

Mixed-signal circuits, Optimization, Analog-to-digital converters, Low-power

1. INTRODUCTION

Over time, most analog-to-digital converter (ADC) research has focused on architectures and techniques to reduce analog power. This has been encouraged by the scaling associated with Moore's law that reduces digital power in comparison with analog and enables greater digital integration on chip. As a result of this research trend, digital power is now a significant portion of the total dissipation in many ADCs, particularly those operating at low resolutions and high speeds. Figure 1 shows the contribution of digital circuits to the overall power consumption for recently published converters. While digital power can be significant across all the resolutions, it is particularly so at low resolutions and cannot be ignored, even though it is rarely mentioned in the literature. For the best energy efficiency, both digital and analog power consumption must be addressed.

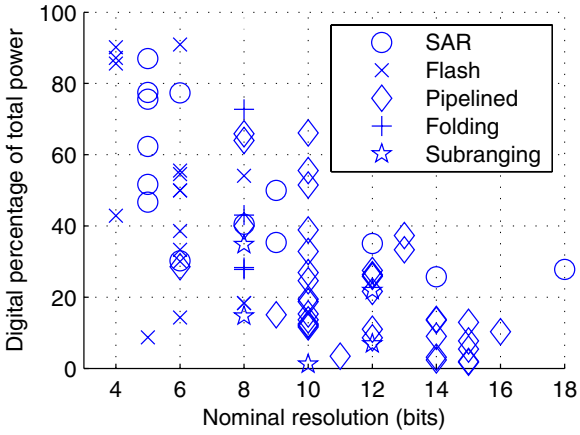

Figure 1: Digital contribution to overall power for published Nyquist ADCs that explicitly separate the analog and digital power consumptions.

Much analog optimization has been performed in the context of ADCs. For instance, many papers have been published discussing optimum scaling between successive stages in pipelined converters [1, 2]. A more comprehensive analog optimization for pipelined converters is presented in [3]. There, the specifications for the operational amplifiers and comparator are optimized for a combination of energy and area requirements using geometric programming, a type of convex optimization. Of these works, only [1] addresses the digital circuitry, and only in a limited fashion.

There are many well known techniques for energy optimization in digital circuits. Perhaps the most straightforward way to save energy is to reduce the power supply voltage. The switching energy in digital circuits is proportional to V_{DD}^2, and so significant energy savings can result. The drawback is the decreased operating speed, which is inversely proportional to V_{DD}. Parallelism or pipelining can be used to maintain constant throughput at lower operating voltages at the expense of increased area and overhead in terms of multiplexing circuits or registers, respectively [4]. More complex algorithms use sensitivity functions to choose optimal device sizing [5] and select between different architectures [6].

DAC 2008, June 8–13, 2008, Anaheim, California, USA

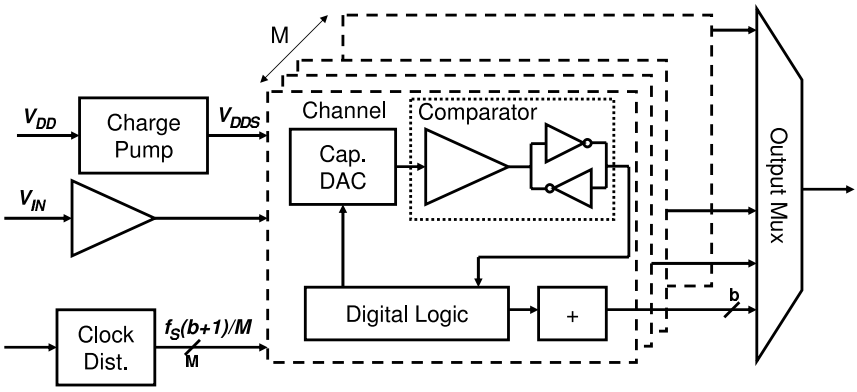

Figure 2: Blocks included in the comprehensive energy model.

The lower speed operation afforded by parallelism improves analog energy as well. For a typical CMOS analog amplifier the energy scales as

$$E_{amp} \propto \frac{V_{DD} C_L I_D}{g_m}, \tag{1}$$

where C_L is the total load capacitance. The energy is minimized when the transconductance efficiency (g_m/I_D) is maximized, which occurs in the sub-threshold region with $V_{GS} \leq V_T$. Unfortunately, with circuits operating at hundreds of MHz, the devices must be biased in strong inversion for large f_T. Parallelism relaxes the device speed constraints.

This paper describes a framework and model for comprehensive analog/digital optimization. Supply voltage reduction and increased parallelism are applied to an ADC energy model that includes optimization of transistor sizes and current biasing. It is therefore important that the effects of low voltage operation and overheads of increased parallelism are modeled accurately. The model is solved as a discrete optimization. While this approach leads to longer solution times and cannot guarantee reaching a global optimum without an exhaustive search, this is an acceptable tradeoff because of the ease of the model setup, and the optimization can be solved in 20–30 minutes on a modern microprocessor. While this is much longer than for convex optimizations solved by geometric programming [3], it is still manageable and permits reasonable design iterations.

A high-speed time-interleaved successive approximation register (SAR) ADC [7–10] is chosen as a test case for this optimization. This architecture has been chosen both because it has demonstrated excellent energy efficiency [7,9,10] and because it has comparable digital and analog complexity. Digital has a significant impact on both the power and, because it limits the critical feedback path during a conversion, speed. The target specifications and global optimization parameters are listed in Table 1. The chosen operating point of 5 bits and 500MS/s is consistent with an ADC for use in an ultra-wideband radio [11]. The remainder of the paper elaborates on the model setup, provides details of a few elements in the model, and presents the results of the optimization.

2. MODEL SETUP

This section describes the model setup and simulation methodology. The speed of solving the energy model is not a primary concern for this work. Instead, the model emphasizes accuracy, simplicity in formulation, and modularity for evaluating architectural and process changes. The final model setup is specific to the SAR topology, but the model formulation and some block descriptions (e.g., sampling switch optimization) are applicable to other mixed-signal systems.

Table 1: Global optimization parameters

Parameter	Description	Value/Range
b	Converter resolution	5
f_S	Sampling frequency	500 MHz
M	# time-interleaved channels	≥ 1
V_{DD}	Core supply voltage	0.5–1.2 V
V_{FS}	Full scale input voltage	0.1 V–V_{DD}
V_{DDS}	Boosted sampling voltage	V_{DD}–1.2 V

Parallel energy and behavioral models are both developed based on partitioning the ADC into the set of hierarchical blocks shown in Fig. 2. Within the dashed rectangle are the sub-blocks of a single SAR ADC channel. These include a capacitive digital-to-analog converter (DAC), comparator (preamplifier and latch), digital logic controller, and digital offset correction block. Wrapped around the M interleaved channels are a number of blocks that model the overhead associated with increased parallelism and low voltage operation. Analog overhead includes the input buffer to drive a potentially large input capacitance. Digital overhead includes an expanded clock network and an output mux to serialize all of the channels. Finally, a charge pump is introduced to allow independent setting of the core voltage from the sampling voltage, as described in Section 3.1. Both the energy and behavioral models have the same block structure. The energy model relates the top-level optimization variables to total ADC energy based on a set of design constraints (e.g., offset voltage, number of time constants required for preamplifier settling). The time-domain behavioral model can relate these constraints to final SNDR. For example, the settling time constraint for the energy model is derived by sweeping the settling time in the behavioral model and selecting a value that gives the proper ADC performance. The coupling of the models is further explored in Section 3.2, and details on the behavioral model are available in [12].

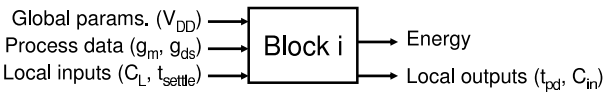

Figure 3: Interface structure of each energy model block.

Each block is implemented as a separate Matlab function, following the conceptual interface of Fig. 3. All blocks share the same set of global inputs, which are separated into the optimization variables listed in Table 1 and tables of process parameters (see below). In addition, each block has internal constraints taken from the behavioral simulations and local inputs, derived from the local outputs of other blocks, that are used to characterize inter-dependencies. The block's Matlab function then computes the energy per conversion and any local outputs. This computation is block- and circuit topology-specific. As a specific example, the comparator block takes as local inputs the available preamplifier settling time (derived from the digital logic's propagation delay), has internal constraints of target offset voltage and required number of time constants for settling, and then uses equations extended from those in [8] to calculate total comparator energy. Intermediate variables calculated within the function (e.g., number of preamplifiers, transistor bias points) are saved to give an initial starting point for transistor-level design. As every block conforms to a similar interface, it is straightforward to evaluate a block design change by re-writing only its equations; other blacks are unmodified.

Each block's model is parameterized and dependent on a limited set of process data, including basic transistor performance (g_m, g_{ds}) and delay/power of simple digital gates. The process data is extracted from SPICE simulations of the worst case design corners. Thus, when porting to a new process, the basic model setup, which emphasizes circuit topology, can be reused, and only the global process inputs must be re-simulated.

For a fixed set of global inputs (i.e., M, V_{DD}, etc.), the top-level SAR energy model returns the energy per conversion. The model, however, cannot find a solution for all sets of global inputs. For example, the specification $M = 1$ and $V_{DD} = 0.5$ means a single 500 mV channel operating at 500 MS/s, requiring an internal clock rate of 3 GHz; as the delay through a single digital gate at 500 mV is much greater than 300 ps, no possible solution exists. The optimization then searches the design space (Table 1) for the set of global inputs that yield a functioning design with the minimum modeled energy.

3. MODEL DETAILS

The complete description of the model is beyond the scope of this paper. In this section, three parts of the model are detailed that are useful to demonstrate the scope of this model across the analog/digital boundary and some of the overheads in a highly parallel system. The first part to be detailed is the sampling operation, which consists of digital signals driving analog circuits and is a particular concern at low voltages. Next, digital offset compensation is presented as a demonstration of the coupling of behavioral and energy models to quantify architectural tradeoffs between analog and digital processing. Finally, clock distribution

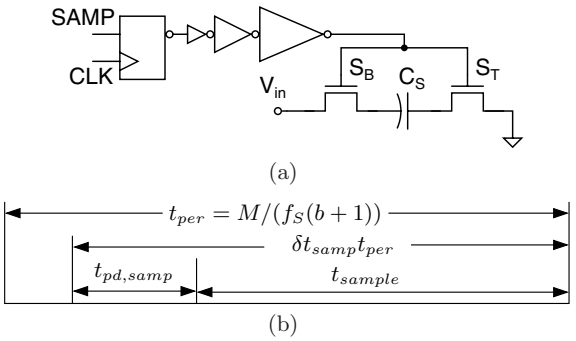

Figure 4: (a) Basic sampling network and (b) sampling period timing constraints.

is presented. As the largest digital overhead to increased parallelism, it limits the number of channels that can be efficiently integrated on chip.

While not discussed in detail, the remaining blocks of Fig. 2 are included in the optimization, with details in [12].

3.1 Sampling: Digital Driving Analog

While lowering the voltage supply reduces energy for the digital logic and even for the matching-dominated comparator or capacitive DAC [13], sampling a high-speed analog signal with a low supply voltage is non-trivial. Consider the sampling network in Fig. 4(a). The bottom plate switch S_B, implemented as an NMOS transistor, has conductance

$$g_{ds} \approx \mu C_{ox} W/L(V_{on} - V_{in} - V_T), \tag{2}$$

where μ is the mobility, C_{ox} is the gate oxide capacitance per unit area, V_T is the threshold voltage, and V_{on} and V_{in} are the gate drive and input signal voltages, respectively. For a fixed V_{on}, the conductance varies with the input voltage level, creating an input signal dependent (and nonlinear) attenuation and phase shift, and the switch fails to turn on for $V_{in} > V_{on} - V_T$. This is particularly a problem in deep sub-micron CMOS where the supply voltages and therefore the maximum $V_{DD} - V_T$ are limited. The top plate switch, S_T, does not suffer from the same problem because its source voltage is always fixed at a DC reference point, which can be set low to maximize gate overdrive.

Several techniques have been proposed to address this problem. Bootstrapping or constant-V_{GS} sampling [14] sets V_{on} to track V_{in}, giving a constant gate overdrive across the input signal range. While required for higher resolution converters, at 5 bits the overhead in terms of area and power per channel is large. Switched-RC sampling [15] replaces the floating series switch with a linear resistor and shunt switch but does not scale well to high frequencies. Instead, in this model, a simple boosting circuit is modeled, where the sampling voltage V_{DDS} is set to an independent, higher value than the core voltage V_{DD}. While this voltage may already be present in the system (e.g., for I/O supplies), for completeness, this model assumes that V_{DDS} is generated by a charge pump that is shared between all channels.

The sampling network has two basic requirements. At the onset of sampling, a first order step response transient of the charge across C_S must settle within the available sampling window t_{sample}. In addition, the switch must be able to

accurately track the highest frequency input signal, $f_S/2$. These translate into the constraints (3)–(4).

$$t_{sample} > k_{slew}(b+1)\ln(2)\tau_{samp}. \quad (3)$$

$$\frac{1}{\sqrt{1+(f_S\tau_{samp}/2)^2}} > 1-2^{-(b+1)} \quad (4)$$

The factor k_{slew} models the nonlinear part of the settling transient and has been determined through simulations to be 1.9. The effective sampling time constant, $\tau_{samp} \approx C_S \cdot (1/g_{dsB,min}+1/g_{dsT})$, is taken for its worst case point, when $V_{in} = V_{FS}$. At this maximum input voltage, the conductance is minimized (see (2)). The widths of S_B and S_T are chosen to satisfy (3)–(4), and simulations confirm that this leads to a conservatively sized sampling network exceeding THD specifications.

The resultant switch size presents a rather large capacitance $C_{SW,samp}$ that requires buffering from the output of a standard size digital gate. This buffering is automatically inserted by the model assuming each stage in the buffer can drive a fan out of 4. One potential problem is that the buffering increases the digital delay $t_{pd,samp}$ and reduces t_{sample} in (3) (see Fig. 4(b)). The timing constraints are automatically enforced by the model. The total sampling energy for a conversion, including the overhead of a charge pump with efficiency η_{CP}, is

$$E_{samp} \approx \frac{(C_{SW,samp}+C_{buf})V_{DDS}^2}{\eta_{CP}}. \quad (5)$$

3.2 Digital Offset Correction

Offset has two deleterious effects in interleaved ADCs. The first can be seen by considering the digital output of an ADC

$$y[n] = \mathcal{Q}\left(v_{IN}[n] + v_{OS}[n \bmod M]\right), \quad (6)$$

where $\mathcal{Q}(\cdot)$ represents the quantization function and $v_{IN}[n]$ is the sampled analog input signal. For an interleaved ADC, the offset voltage v_{OS} is periodic with the number of channels, producing spurs in the output spectrum at multiples of f_S/M, the channel sampling frequency.

Offset produces a second effect, namely clipping of large input signals. The maximum input dynamic range of an ADC is thereby reduced by the peak-to-peak offsets across the channels. Analog-only offset compensation [8] sets the comparator's offset requirement to reduce the power of the periodic term in (6) below the quantization noise power, namely

$$E[v_{OS}^2] = \sigma_{v_{OS}}^2 < \frac{V_{LSB}^2}{12}. \quad (7)$$

This condition forces the offset of every channel to be small ($\ll V_{LSB}$), and the reduction in dynamic range due to clipping is negligible.

Another approach to reducing the power of the periodic term in (6) is to correct it digitally, which simply requires an adder[1] [16]. Then the analog offset requirements can be set only to keep clipping within an acceptable limit. The effect of digital offset correction is quantified using a coupling of behavioral and energy models.

[1]While medium complexity calibration circuitry is required to determine the proper offset correction value, once calculated, the continual correction only requires an adder.

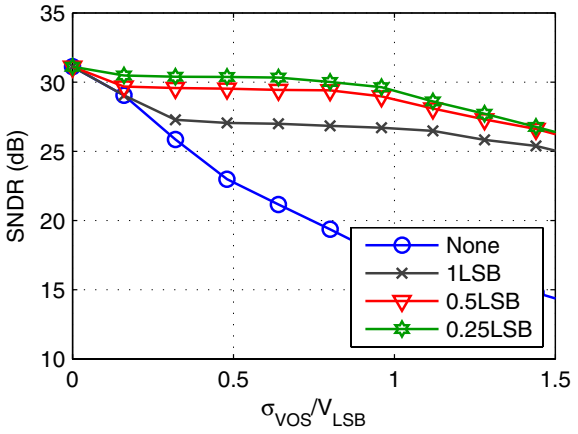

Figure 5: Behavioral simulations showing the performance degradation of the ADC versus the analog offset voltage for increasing degrees of digital offset correction.

Figure 5 shows an interleaved ADC's signal-to-noise-plus-distortion ratio (SNDR) versus the analog offset voltage for varying degrees of digital offset correction. The lower plot, with no offset correction, tails off directly due to the effect of the periodic offset term in (6), while the upper curves, with digital offset correction, only drop when clipping becomes significant. For the same 1 dB degradation in SNDR, digitally correcting the offset to within a half an LSB permits using a 6× larger analog offset voltage, although 3× is used to be conservative in this model.

The results of the behavioral model can then be fed back to the energy model. Specifically, the comparator offset constraint is set as

$$V_{OS,in} = \begin{cases} \dfrac{2^{-b}V_{FS}}{6} & \text{no digital offset correction,} \\ \dfrac{2^{-b}V_{FS}}{2} & \text{with digital offset correction.} \end{cases} \quad (8)$$

The increased comparator offset voltage permits a reduction of the number of preamplifiers from 2 to 1, and the energy model predicts an overall energy savings of 30% (after the energy of an adder is included) from using the digital offset correction. This demonstrates both the effectiveness of digital correction for analog impairments and the ability for the coupled behavioral and energy models to quantify architectural tradeoffs across the analog/digital boundary.

3.3 Clock Distribution

Clock distribution is one of the principle overheads of highly interleaved ADCs because both the number of clocks and the die area across which the clocks are distributed increase. For the proposed energy model, the clock distribution is implemented using a simple clock tree model. To minimize timing skew, the clock paths to every channel should be balanced, which means that the paths to the closest and farthest channels should have the same length. This can be accomplished by an H-tree [17] or similar network. The

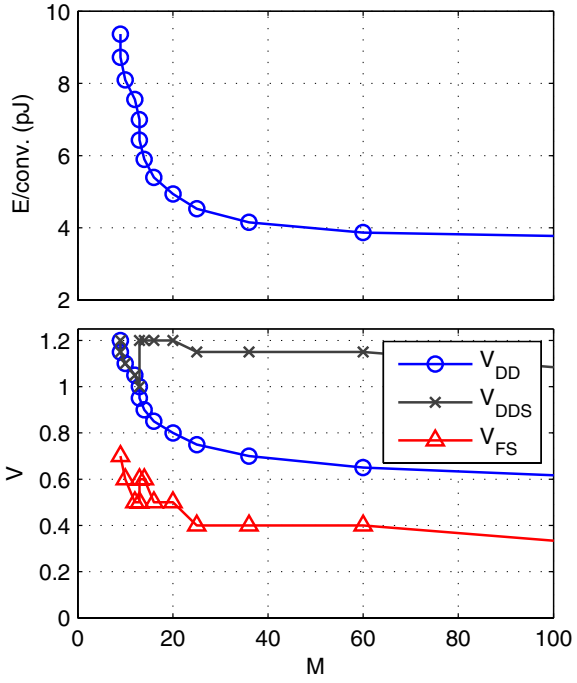

Figure 6: Results of the SAR energy optimization, with energy per conversion (top) and optimum voltages (bottom) versus number of parallel channels M.

length of the clock path to any channel therefore grows as the length from the center of the die (presumed to be the origin of clock distribution) to the corner of the die (the location of the farthest channel). If the die area grows with M, then the length of the clock path to a single channel grows with the diagonal, and the per-channel clock capacitance $C_{clk} \propto \sqrt{M}$. This scales similarly to that of an H-tree's capacitance, $C_{clkwire}$, with the number of gates, N_g, in a digital circuit, $C_{clkwire} \propto \sqrt{N_g}$ [18].

The total energy per conversion is

$$E_{clk} = C_{clk} V_{DD}^2 (b+1). \quad (9)$$

The final term is included because a single conversion requires $(b+1)$ clock periods. It is important to note that (9) only applies if the minimum frequency clock is distributed to every channel. If, instead, the sampling clock is distributed in a balanced path to every channel, the number of clock edges per conversion increases from $b+1$ to M, giving a total clock energy that grows as $M^{3/2}$ instead of $M^{1/2}$. Clearly, minimizing the frequency of the clock distribution is critical in interleaved ADCs.

4. MODEL RESULTS

The energy model has been implemented using Matlab with the global optimization parameters defined in Table 1. Figure 6 plots the predicted energy per conversion and operating voltages versus the number of parallel channels. The optimum energy point occurs with 121 channels, but the curve is fairly shallow above 30 channels. The small additional energy savings would not be worth the great increase in area and complexity. The kink in the curve for small values of M occurs because of the overhead of the charge pump. When $M < 13$, the core voltage is sufficiently high that the charge pump, with its non-ideal efficiency η_{CP}, is not necessary. A summary of the model results at 36 channels, which is within 10% of the optimum energy point, is presented in Table 2, and the breakdown of energy consumption among the various blocks is shown in Fig. 7. As expected from (1), the model selected preamplifier current densities at the edge of weak inversion near the maximum transconductance efficiency. Compared to a non-optimized SAR ADC in the same technology [9], the model yields an ADC with 3× lower energy per conversion.

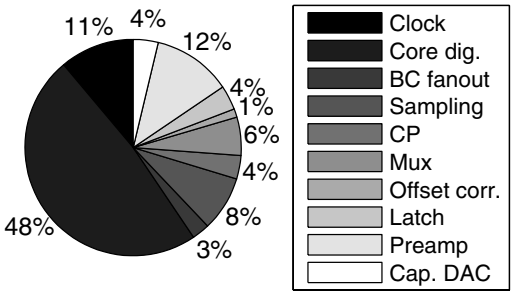

Figure 7: Predicted energy breakdown of optimized 36-channel SAR ADC.

Table 2: Summary of optimum and 36 channel operating points

Parameter	36 Channel
M	36
V_{DD}	0.7 V
V_{FS}	0.4 V
V_{DDS}	1.15 V
E/conversion	4.15 pJ

The flexibility of the model is demonstrated by evaluating the use of a boosted sampling supply. Re-running the optimization with the additional constraint that $V_{DDS} = V_{DD}$, yields the curve shown in Fig. 8. At low voltages, the sampling switch size rapidly increases. Even with the non-ideal charge pump, boosting the sampling supply can produce an overall 22% energy savings with a smaller sampling switch and lower core voltage.

5. DISCUSSION AND CONCLUSION

The traditionally digital energy saving methods of supply voltage reduction and parallelism have been applied to the interleaved SAR ADC and have, combined with an optimization for analog transistor sizing and bias current densities, led to an operating point with 3× energy savings. A 36-way interleaved ADC prototype has been implemented in a 65 nm CMOS technology using the transistor sizes from the model as a starting point and further refining them through transistor-level simulations [19]. Figure 9 compares the power breakdown of the chip with the power expected from the model. Due to a limited number of supply pins, the granularity of these results is not as fine as the breakdown

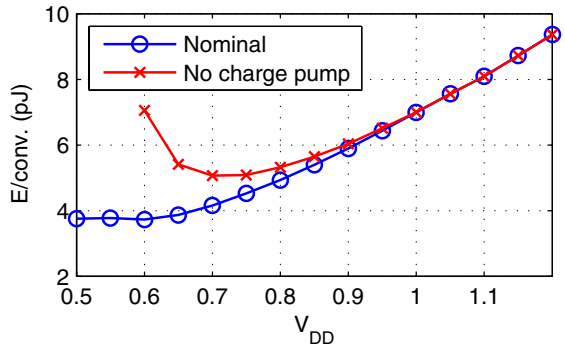

Figure 8: Model results comparing the energy versus supply voltage with and without the charge pump and boosted V_{DDS}.

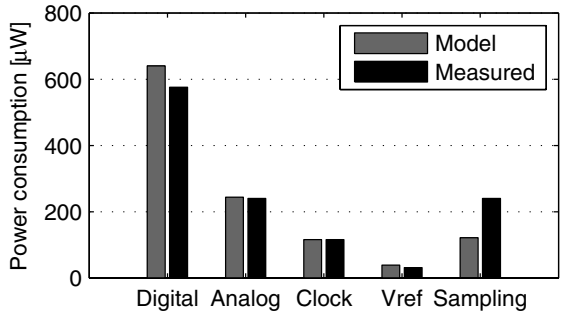

Figure 9: Comparison of the measured ADC power from a 65 nm CMOS testchip [19] with the results from the energy model.

in Fig. 7. The measured power numbers are fairly close to the modeled ones. The increase in sampling power is due to an architectural change where the sampling supply is used to drive the first stage of clock buffers across the chip.

While relatively undiscussed among ADC literature, digital energy is and will continue to be non-trivial in medium-to-low resolution converters and must be considered when targeting the best possible performance in a mixed-signal circuit or system. The results presented above demonstrate that significant mixed-signal energy savings are achievable through joint optimization of analog and digital circuits.

6. REFERENCES

[1] J. Goes, J. Vital, and J. Franca, "Systematic design for optimization of high-speed self-calibrated pipelined A/D converters," *IEEE Trans. Circuits Syst. II*, vol. 45, no. 12, pp. 1513–1526, Dec. 1998.

[2] Y. Lin *et al.*, "kT/C constrained optimization of power in pipeline ADCs," in *Proc. of the IEEE Int. Symp. on Circuits and Systems*, vol. 3, May 2005, pp. 1968–1971.

[3] M. del Mar Hershenson, "Design of pipeline analog-to-digital converters via geometric programming," in *IEEE/ACM Int. Conf. on Computer Aided Design*, 10-14 Nov. 2002, pp. 317–324.

[4] A. Chandrakasan, S. Sheng, and R. Brodersen, "Low-power CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 27, no. 4, pp. 473–484, Apr 1992.

[5] D. Markovic *et al.*, "Methods for true energy-performance optimization," *IEEE J. Solid-State Circuits*, vol. 39, no. 8, pp. 1282–1293, Aug. 2004.

[6] V. Zyuban and P. Strenski, "Unified methodology for resolving power-performance tradeoffs at the microarchitectural and circuit levels," in *Proc. of the Int. Symp. on Low Power Electronics and Design*, 2002, pp. 166–171.

[7] D. Draxelmayr, "A 6b 600MHz 10mW ADC array in digital 90nm CMOS," in *ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 264–265.

[8] B. P. Ginsburg and A. P. Chandrakasan, "Dual time-interleaved successive approximation register ADCs for an ultra-wideband receiver," *IEEE J. Solid-State Circuits*, vol. 42, no. 2, pp. 247–257, Feb. 2007.

[9] ——, "500-MS/s 5-b ADC in 65-nm CMOS with split capacitor array DAC," *IEEE J. Solid-State Circuits*, vol. 42, pp. 739–747, Apr. 2007.

[10] S.-W. M. Chen and R. W. Brodersen, "A 6b 600MS/s 5.3mW asynchronous ADC in 0.13μm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. of Tech. Papers*, vol. 49, Feb. 2006, pp. 574–575.

[11] P. P. Newaskar, R. Blazquez, and A. P. Chandrakasan, "A/D precision requirements for an ultra-wideband radio receiver," in *IEEE Workshop on Signal Processing Systems*, Oct. 2002, pp. 270–275.

[12] B. P. Ginsburg, "Energy-efficient analog-to-digital conversion for ultra-wideband radio," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2007.

[13] S. Gambini and J. Rabaey, "Low-power successive approximation converter with 0.5 V supply in 90 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 11, pp. 2348–2356, Nov. 2007.

[14] T. Brooks *et al.*, "A cascaded sigma-delta pipeline A/D converter with 1.25 MHz signal bandwidth and 89 dB SNR," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 1896–1906, Dec. 1997.

[15] G.-C. Ahn *et al.*, "A 0.6-V 82-dB delta-sigma audio ADC using switched-RC integrators," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2398–2407, Dec. 2005.

[16] D. Fu *et al.*, "A digital background calibration technique for time-interleaved analog-to-digital converters," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 1904–1911, Dec. 1998.

[17] J. M. Rabaey, A. Chandrakasan, and B. Nikolić, *Digital Integrated Circuits*, 2nd ed. Prentice-Hall, 2002.

[18] D. Liu and C. Svensson, "Power consumption estimation in CMOS VLSI chips," *IEEE J. Solid-State Circuits*, vol. 29, no. 6, pp. 663–670, Jun. 1994.

[19] B. P. Ginsburg and A. P. Chandrakasan, "Highly-interleaved 5b 250MS/s ADC with redundant channels in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. of Tech. Papers*, vol. 51, Feb. 2008, pp. 240–241.