Resolving the Memory Bottleneck for Single Supply Near-Threshold Computing

Tobias Gemmeke[2], Mohamed M. Sabry[3], Jan Stuijt[2], Praveen Raghavan[1], Francky Catthoor[1], David Atienza[3]

[1] imec
Leuven, Belgium

[2] Holst-Centre / imec
Eindhoven, The Netherlands
tobias.gemmeke@imec-nl.nl

[3] ESL-EPFL
Lausanne, Switzerland

***Abstract*— This paper focuses on a review of state-of-the-art memory designs and new design methods for near-threshold computing (NTC). In particular, it presents new ways to design reliable low-voltage NTC memories cost-effectively by reusing available cell libraries, or by adding a digital wrapper around existing commercially available memories. The approach is based on modeling at system level supported by silicon measurement on a test chip in a 40nm low-power processing technology. Advanced monitoring, control and run-time error mitigation schemes enable the operation of these memories at the same optimal near-V_t voltage level as the digital logic. Reliability degradation is thus overcome and this opens the way to solve the memory bottleneck in NTC systems. Starting from the available 40 nm silicon measurements, the analysis is extended to future 14 and 10 nm technology nodes.**

Keywords—Near-threshold computing; memories;

I. INTRODUCTION

The increase in integrated circuit complexity and continued scaling of process technology has led to a tremendous wealth of new applications. Similarly to bipolar and later NMOS technology, the so-called power wall more and more limits the number of active elements in today's systems. This problem has started when voltage scaling flattened out in recent CMOS technology generations. A key game changer is the application of near-threshold computing (NTC) to systems being limited by their dissipated power or energy. Besides the various difficulties of scaling the supply voltage of digital logic efficiently, on-chip SRAM memories tend to dominate the overall power figures in such NTC systems (Section II). Hence, it is crucial to address this bottleneck in the overall NTC platform. This paper will present techniques to overcome the classic supply voltage boundaries of supply voltage scaling in on-chip memories. We address this challenging problem by proposing several contributions. We start with a thorough evaluation of the potential of NTC memory operation (Section III). Then we show that for cell-based memories, the cells can be optimized in the same way as digital logic to allow robust low power operation at the same low voltage level, in contrast to what is feasible with traditional custom SRAM design (Section III). In particular, for the L1 memory layer, their limited sizes motivate this choice as the area overhead is still affordable. Actual measurements in a 40nm technology allowed us to accurately characterize the reliability degradation and the expected number of simultaneous bitflip errors as function of the voltage. Next, we discuss our modeling approach for two distinct NTC memory design flows (Section IV), and we show subsequently that significant power savings can be achieved by exploiting a novel low-overhead run-time scheme to mitigate these multiple errors (Section V). These memory-oriented techniques are fully compatible and complementary to the emerging methods that are focused on the processor cores. The latter NTC methods are outlined in the contributions on architectural implications for best-effort extreme scale many-core processors [1] and NTC many-core processors with performance guarantees [2]. The overall result of this work is a new multi-core computing architecture, where the memory organization can operate in the same voltage range as the processor data-path, thus saving energy up to 2x compared to the traditional Error-Correcting Coding (ECC) approaches, and 3x compared to no mitigation. We complete this paper by a study on the impact of future technologies on NTC memories (Section VI).

II. RELEVANCE OF MEMORIES AT NTV

Reduction of power consumption in SRAM memories is a key topic of continuous research in the era of low-power computing. With the introduction of near-threshold or even sub-threshold operation to digital circuits, memory design has gained renewed attention as being more susceptible to the side effects of low voltage operation. This can mainly be attributed to the fact that the classic 6T SRAM cell is a ratioed circuit relying on the relative drive strength of the transistors involved. Parametric variations of the individual devices can lead to functional failures of the cell. One apparent option is the use of different supply voltages for the digital domain and memories. This approach entails additional complexity on system level (requiring the generation and distribution of multiple supply voltages) as well as in the backend (implementing level shifting and multi-voltage timing closure).

978-3-9815370-2-4/DATE14/©2014 EDAA

Certain parts of the standard SRAMs are already optimized to avoid full swing operation limiting the benefit of voltage reduction in such cases. However, other portions of the memory still feature full-swing dynamic and static operation that benefits from power reduction with supply voltage.

Another important aspect is active leakage power. It is to the first order proportional to the total transistor count which is dominated by the memories. Additionally to the existing wealth of design techniques targeting leakage reduction, supply voltage is a leverage achieving up to 10x better static power. What is more, applications benefitting from NTC typically have significant standby times. Whereas digital logic can largely be powered off, memories have to retain their contents. In this case supply voltage scaling achieves a significant leakage power reduction complementary to other design techniques as detailed in the next section.

Figure 1 highlights the different contributors to total energy per operation as measured for a signal processor [3].Remarkably, the dissipation in the memories actually increases on an energy per cycle measure at reduced voltages. This is due to the fact that supply scaling of the commercial memories is stopped at 0.7V, i.e. any further reduction of performance in the digital domain leads to a high impact of the unscaled leakage power in the memories. Also, the share of leakage power dominating the energy per cycle figure below 0.6V becomes apparent. In this paper we want to alleviate this bottleneck through novel memory design and error mitigation techniques.

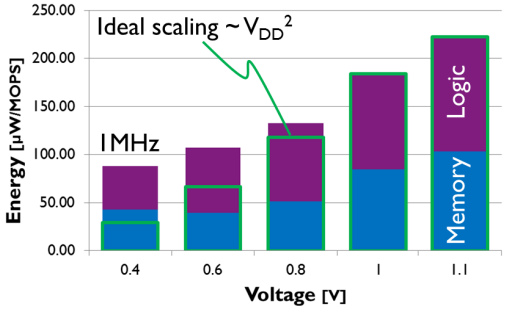

Figure 1. Energy per cycle vs. supply voltage based on measurements of an advanced signal processor in a 40nm low-power technology [3].

III. Memory Design For Low-Power

The three modes of operation of an SRAM are read, write, and data retention, each having their specific minimal supply voltage. The dynamic read and write operation can be improved by a variety of assist techniques realized in the periphery of the actual cell array. One field of techniques weaken (write) or strengthen (read) the cell during the access by (temporarily) deviating from the nominal voltage levels on the supply rails, bit-lines, and/or word-lines. Another field is the reduction in variations between the complementary bit values starting at the memory cell including the bit-line wiring (length & routing), and including compensation schemes for the mismatch in the sense-amplifier. An extreme case is the use of (single-ended) full-swing logic avoiding the mismatch in the read-out altogether.

On the other hand, new bit cells are used to allow independent write and read operation at the cost of a higher transistor count. This leads to an increase in area, which becomes even worse if the new cell has to follow standard digital design rules as opposed to the tighter SRAM design rules optimized for a specific 6T cell layout. This drawback has to be carefully weighed against the improvement in stability and read/write performance.

The hold stability [5] limits the voltage scaling during retention and therefor largely impacts the standby leakage power. Optimization focuses on improving the static noise margin subject to mismatch variations [6]. It covers the core cell with the cross-coupled inverter pair and includes the effect of the access transistor.

All of these techniques improve SRAM operation in view of increased variability due to reduced supply voltage and/or modern technologies. Low-power dynamic access is best achieved by hierarchical subdividing the memory as to limit switching activity to short local bit and/or word-lines as well as limiting the swing on the local and global lines (an early overview is given in [7]; recent results can be found in [8]).

An extreme case is the operation of the SRAM in the near- or sub-threshold regime [9][10] during standby [6] as well as during dynamic access [11], which was preceded by a register file like implementation based on combinational logic [4].

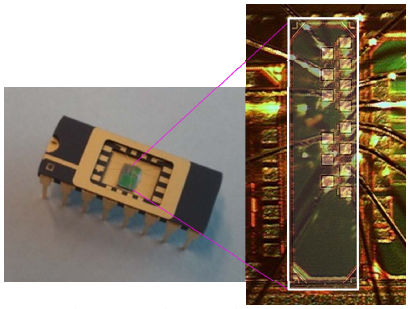

Figure 2. Micrograph of test chip.

IV. Memory Performance at NTV

Two designs were integrated on a test chip (cf. Figure 2) and measured in silicon. They cover the two extremes of design space of NTV memory solutions. The analysis focuses on implementations that do not require extensive custom re-design effort. This has the additional benefit of avoiding major NRE (non-recurring engineering) cost which is already an increasing limitation in the adoption of modern technologies. The NTV compatibility is rather achieved by adding suitable error mitigation techniques at and above the RTL level.

On one side, a cell based memory is used consisting of standard cells, only. Whereas [13] uses sequential elements (latches), in our case the bit cell is composed of a cross-coupled pair of AND-OR-INVERT gates resulting in better area efficiency (cf. Table 1).

Apparently, such kind of memory still suffers an area penalty as compared to SRAMs based on highly optimized 6T cells. However, the cell-based approach enables operation at the same voltage level as the common digital logic without any additional custom design effort.

On the other side, a commercial memory IP is characterized. As a matter of fact even such memory can run at a much lower supply voltage than the one specified by the IP provider. This is due to the fact that the provider's limits have to account for all PVT variations and ageing over the lifetime of a product.

In both cases measuring actual silicon reveals the margin that can be exploited with techniques as discussed in Section V. Apparently, the minimal voltage will change over lifetime of a product requiring a monitoring and control loop that adjust run-time knobs such as the supply voltage level.

Table 1. Comparison of different implementations scaled to 1k x 32b memory (40nm, TT corner, 1.1V, 25C).

Feature	Unit	COTS 40nm	Custom SRAM [12] 40nm	Cell-based SRAM [13] 65nm	Cell-based imec 40nm
Dyn. power	pJ	12	3.6*[2]		1.4
(red. voltage)		-	-	0.93@0.4V*[2,4]	0.18@0.4V
Act. Leakage	µW	2.2	11*[2]	-	5.9
(red. voltage)		-	-	8 @ 0.35V*[2]	-
Area	mm^2	0.01*[3]	0.024*[3]	0.19*[3,4]	0.058
Retention	V	0.85*[1]	-	0.25	0.32*[1]
Performance	MHz	820	454	9.5@0.65V	96*[1]
(red. Voltage)		-	-	0.1@0.45V	0.4*[1]@0.45V

*[1] Measurement results
*[2] Scaled to same number of bits
*[3] Scaled ∝ total bits
*[4] Scaled ∝ technology $(40nm/65nm)^2$

The first measurement captures the minimal retention voltage for both memories. Individual bit failures as a function of the minimum supply voltage are shown in Figure 3.

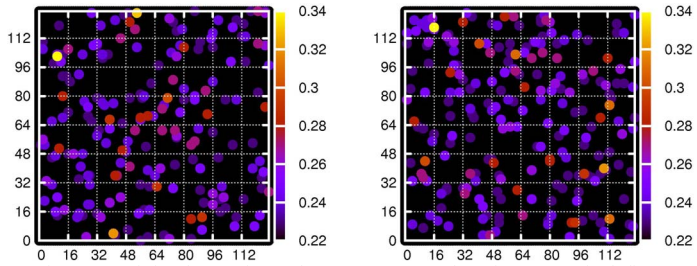

Figure 3. Minimal retention voltage vs. memory location (x,y axis define memory location) of one instance of commercial memory IP (left) and cell-based memory (right).
(Errors in individual bits cells are magnified for better legibility)

The cumulative bit failure probability for all 9 tested dies is shown in Figure 4. Each bit cell has its specific noise margin *NM* following a Gaussian distribution. The estimation of the *NM* in Eq. (2) [14] as a function of the standard variation σ allows to predict the number of failing bit cells assuming that all bit cells fail at the same *NM*.

$$NM = c_0 \cdot V_{DD} + c_1 + c'_2 \cdot \sigma. \quad (2)$$

The standard deviation σ defines the limiting variation of any memory cell that feature the specified or a better noise margin. Fixing the *NM* in Eq. (2) at the level of failure, any change in V_{DD} is related to a specific change in standard deviation Δσ in the distribution of *NM* of the core cells Figure 4:

$$\Delta V_{DD} / \Delta\sigma = c'_2 / c_0 = \text{const.} \quad (3)$$

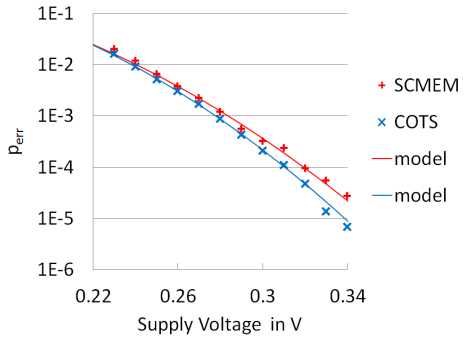

Figure 4. Retention bit error rate vs. supply voltage.

The limiting standard deviation σ can be related to the bit error probability, using the cumulative density function of the corresponding Gaussian distribution. This results in the following model of bit error probability as function of supply voltage:

$$p_{\text{bit,err}} = 0.5 \cdot \left[1 + \text{erf}\left(\frac{V_{DD}/d_0 - d_1}{\sqrt{d_2^2}}\right)\right] \quad (4)$$

with $d_{0\ldots2}$ being fitted to the data.

The second measurement reveals the minimal supply voltage for read & write operations. The testing is done as quasi-static operation masking the change in performance as far as it is not part of self-timed circuits within the memory instance.

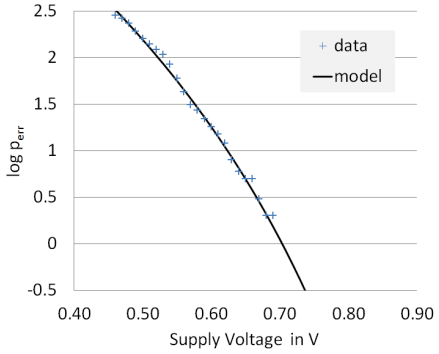

Figure 5. Error probability of RW access vs. supply voltage.

The measured data is fitted to the following empirical model

$$p_{\text{bit err}} = A \cdot (V_0 - V_{DD})^k. \quad (5)$$

For the commercial memory, the fit parameters are A=6, k=6.14, and V_0=0.85. In case of the cell based memory, the minimal

access voltage is V_0=0.55 (in the worst-case) going down to a few 10mV above the retention voltage for most parts.
The silicon measurement data is also used to calibrate circuit simulation, which we use to model the behavior of such memory over a wide voltage range. The resulting model is integrated in a memory calculator estimating key figures of merit over a wide range of input parameters.

V. Power Reduction Applying Error Mitigation

In the previous sections, we have highlighted the impact of NTC on the bit error rate. In fact, the significant error rate increase (cf. Figure 5) undermines NTC effectiveness in energy savings due to the resulting performance degradation. Thus, it is essential to integrate an error mitigation module that recovers from bit errors. A crucial aspect in the integrated module is the energy overhead introduced. The energy spent in recovering from increased error rates should not overcome the savings achieved from NTC. In this respect, we have explored several error recovery mechanisms to assess the corresponding impact on the overall energy consumption. In particular we have examined the following mechanisms:
SECDED Hamming Code [15][16]**:** ECC is one of the most popular error mitigation mechanisms implemented in hardware and it is widely used in industry. In particular, a memory code is generated for each memory word write access, i.e. this code is sufficient to detect a double-bit and correct a single-bit error.
OCEAN [17][18]: This error correction mechanism is categorized as a cross-layer hybrid SW/HW technique where the mitigation happens demand-driven at run-time. This class of techniques is introduced to minimize the Energy-Performance-Area (EPA) overheads required to mitigate the increased error rates. In particular, OCEAN applies nonlinear programming to achieve the minimal energy overhead possible.
For each of the mentioned techniques, we apply an upper bound on the acceptable failure rate (FIT) of 10^{-15} (i.e. the maximum failure rate where no error correction is possible is 10^{-15} faults per read/write transaction). This upper limit imposes an upper bound on the minimal voltage applicable depending on the error mitigation mechanism. Indeed each error protection mechanism has an upper limit on the number of bit errors that can be corrected. In the case of SECDED, a triple-bit error would lead to system failure, while in OCEAN a quintuple (5 bits) error is needed for system failure. In the following subsection we elaborate more on the platform setup used in this exploration, as well as the examined application.

A. *Experimental Setup*

To explore the potential gains of NTC when error mitigation is applied, we use a simulated single-core platform that includes a 32-bit ARM 9 processor, 4 KB instruction memory and 8 KB scratchpad data memory. This architecture is similar to the NXP system-on-chip (SoC) platform [19]. Figure 6 shows the schematic diagram of the simulated platform. In order to estimate the energy consumption of this platform, we use the power figures, both leakage and dynamic, of building blocks in a 40nm low-power technology that uses our internal memory database to calibrate the energy vs. performance trends in the CACTI tool [20][21]. CACTI is then used in the experiments of this work to estimate the power consumption of the target platform, including the error mitigation overheads, as the absolute figures for the commercial memory generator data are confidential.

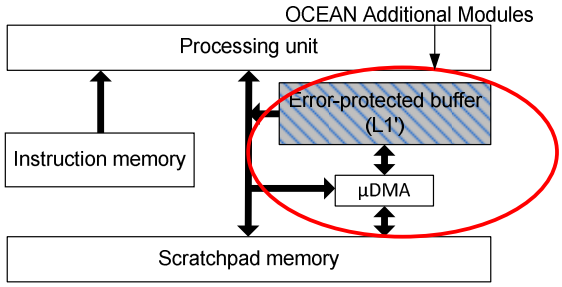

Figure 6. Schematic diagram of the evaluated architecture. Additional modules due to OCEAN[18] are encircled in red color.

We simulate this SoC on a cycle-accurate simulator, namely MPARM [22]. MPARM is a multi-processor virtual platform, which includes SystemC models of different modules (such as processing units, cache memories, SPMs, DMA) interconnected with different interconnection protocols (AMBA-AHB, AMBA-AXI, NoC, etc.).
To incorporate the mentioned error mitigation mechanisms, we have extended MPARM to support these in the following way:

1) **SECDED***:* We use the (39, 32) SECDED code implementation to cope with the memory word width. To achieve an accurate simulation, we have added the energy overhead required for this module. In particular, we augment the read/write energy to account for the fact that we read/write 39 bits instead of 32. Moreover, we add the energy overhead needed to generate the code word, to check for an error, and to correct the error if it occurs [16].

2) **OCEAN**: As mentioned earlier OCEAN is a hybrid SW/HW mitigation mechanism. In a nutshell, OCEAN achieves low-energy error mitigation by exploiting a dynamic check-pointing and rollback mechanism at finer granularity. As shown in Figure 7, OCEAN splits a computation tasks into a set of equivalent phases. Each phase generates a chunk of data that is required for the subsequent phases to be error-free. In OCEAN, these data chunks are stored in error-protected buffer, with quadruple error correction capability, such that they can be used in case of error detection in the data memory. Thus, OCEAN requires additional HW (cf. Figure 6) modules and SW routines. We have extended MPARM to provide the support needed for OCEAN. We refer the reader to [17][18] for more details.
We use in this analysis a 1K-point FFT, but the analysis is applicable to other streaming applications as well.
In order to have an extensive evaluation, this study examines several performance requirements for the FFT, which span a

wide range of operating frequencies, namely 290 KHz (the minimum allowable frequency at the lowest voltage) and 1.95 MHz. These performance requirements impose constraints on the minimum operating voltage for each case, in addition to the error rate requirements. Table 2 shows the operating voltage for each performance requirement. For the highest frequency the gains are very limited because we cannot reduce the voltage compared to the nominal one for the ECC reference. This motivates the use of parallelism to allow reducing the required frequencies and to exploit the quadratic voltage gains at a quasi-linear parallelization cost (applications like FFT support this).

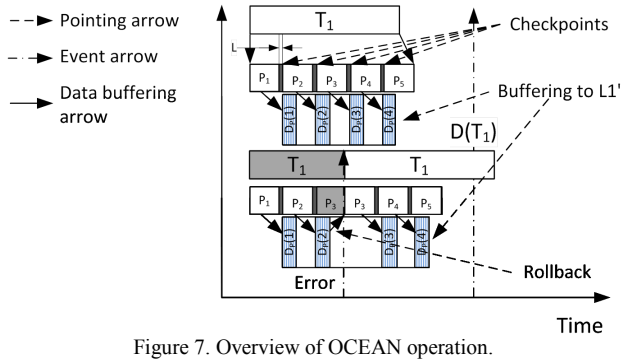

Figure 7. Overview of OCEAN operation.

Table 2. Minimum voltage requirement to achieve the desired FIT (10^{-15}).

Frequency	Voltage		
	No Mitigation	ECC	OCEAN
290KHz	0.55V	0.44V	0.33V
1.96MHz	0.55V	0.44V	0.44V

B. *Experimental Results*

We report in this analysis the dynamic and leakage energy for all of the modules in Figure 6. In particular, the following results include the power consumption of the processing unit (core), instruction memory (IM), scratchpad data memory (SP), and protected memory (PM).

Figure 8 shows the power consumption for the 290 KHz case. From these figures, we can observe the following. First, adding an error mitigation technique definitely saves power for the required error rate, as the energy overhead for protection is superseded by the gains from lowering the operational voltage. In fact these power savings can be up to 70% in the lowest performance requirements (290 KHz) case and 37% in the highest performance requirements (1.96 MHz).

OCEAN also saves up to 48% more power than ECC. This is related to the protection mechanism of OCEAN that can allow much higher error rates, hence reducing the voltage more than the case of ECC. Moreover, the OCEAN protection mechanism is more power efficient than ECC, which is reflected in the 7% increased power savings achieved, compared to ECC, when the supply voltage is similar in ECC and OCEAN (cf. Table 2).

In addition to the previous design point, we examine the system with higher operating voltage as well. In this case, the nominal voltage where there is no error mitigation is shifted from 0.5V to 0.85V (cf. Section IV). In this case, the operating frequency is increased to 11MHz, and the corresponding operating voltages, to maintain the desired FIT, are as follows; 0.88V for the no mitigation case, 0.77V for ECC, and 0.66V for OCEAN.

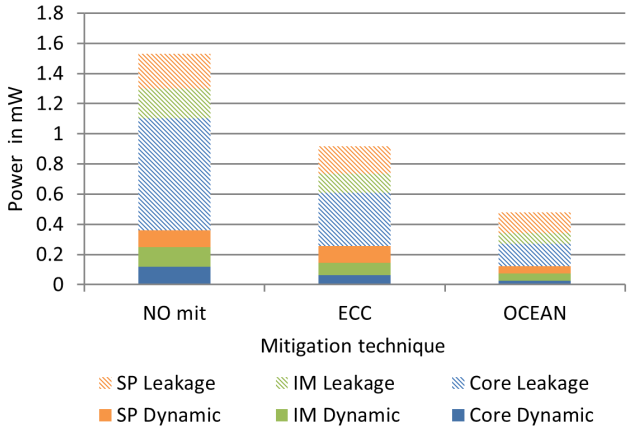

Figure 8. Power consumption in the case 290 KHz frequency.

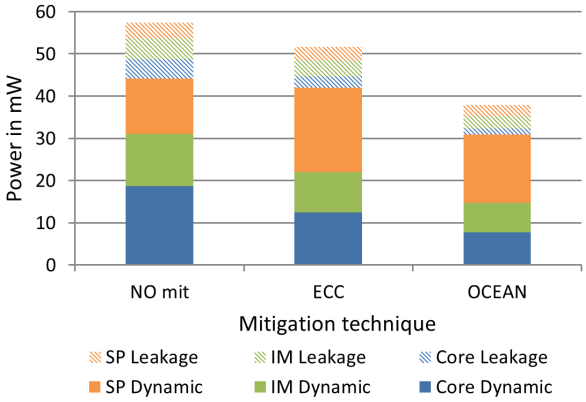

Figure 9. Power consumption in the case of 11 MHz frequency.

In this new case, OCEAN achieves comparable power savings as in the former case (see Figure 9). In particular, we achieve 34% power savings between OCEAN and no mitigation, and 26% power savings in OCEAN compared to ECC. It is important to mention however that the power consumption values are one order of magnitude higher than the values in Figure 8. For instance, the power consumption in the no mitigation (No mit) case is 57 mW.

These results confirm the crucial need to deploy error mitigation mechanism, and preferably the energy-optimized technique, to harness the full potential of NTC energy savings.

VI. NTV Memories in Future Technologies

Moving from planar devices to finFET has a major advantage of having better sub-threshold slope; therefore the use of NTV in finFET techniques is expected to bring higher gains. However there is need to keep the Avt of the device under control. The authors of [23] show the gains at the SoC level of finFET.

Figure 10 shows the impact on a 14nm finFET and a 10nm multi gate device for the example of an inverter delay as the voltage is scaled to near threshold regime. The figure shows two aspects, the scaling of the mean delay, which is determined by the sub-threshold slope of the device, as well as the sigma spread of the delay which is determined by process control and random variations.

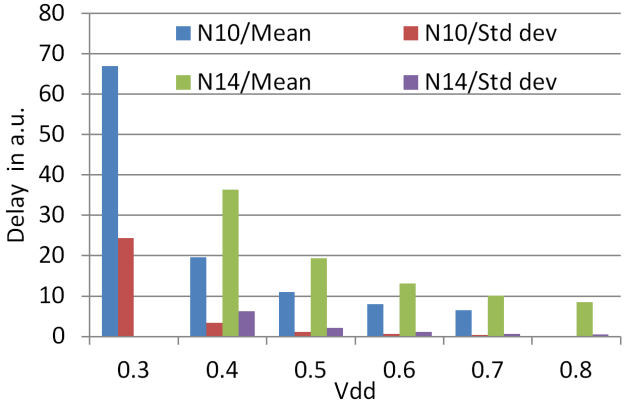

Figure 10. Inverter delay in finFETs.

The benefits for reducing the power consumption using such modern technologies are three-fold. Firstly, there is the reduction of wiring capacitance with reduced feature size that impacts dynamic energy per operation directly, but also indirectly as the smaller capacitance requires less drive strength to achieve a certain voltage step in a specific time frame.

Secondly, there is the significant improvement in performance due to higher drive currents in the smaller geometries. Going from 14nm to 10nm results in a 2x speed-up (cf. Figure 10).

Finally, the standard deviation is under tight control. As shown in Section IV, this impacts timing closure but also the minimal operational voltage of the memory. The results in Figure 10 not only reveal the small standard variation employing a finFET device, but also highlight additional improvements in the transition from 14nm to 10nm. Overall, it is clear that the gains with OCEAN and other NTV methods would largely benefit by the use of modern finFET devices.

VII. Conclusion

Power consumption is a hard constraint for a variety of applications for many years. Rising performance and feature requirements have led to an ever extending need for high efficient computing. Reducing the supply voltage is considered one key enabler – limited by various challenges. In case of the on-chip memories, high susceptibility to process variations makes them especially vulnerable to low-voltage operation.

In this paper, the wide variety of techniques to achieve ultra-low power memory operation, ranging from technology selection to error mitigation in middleware was outlined. Together, these techniques pave the path to low-voltage memory operation overcoming reliability issues of modern technologies at the same time. Depending on the memory style, be it highly optimized 6T cells or cell based, significant savings are achieved by scaling the memory voltage together with the digital domain. Applying the presented scheme of monitoring, control and mitigation a 3.3x lower dynamic power is achieved beyond the voltage limit for error free operation.

Acknowledgment

This work was partially supported by the EU FP7 Project Phidias (GA n. 318013) and IMEC's IIAP program.

References

[1] J. Torrellas, "Extreme-Scale Computer Architecture: Energy Efficiency from the Ground Up," to appear DATE, 2014.
[2] C. Silvano, et al., "Voltage Island Management in Near Threshold Many-core Architectures to Mitigate Dark Silicon", to appear DATE, 2014.
[3] M. Konijnenburg et al., "Reliable and energy-efficient 1MHz 0.4V dynamically reconfigurable SoC for ExG applications in 40nm LP CMOS," ISSCC, 2013.
[4] A. Wang and A. Chandrakasan, "A 180 mV FFT processor using subthreshold circuit techniques," ISSCC, pp. 292-3, 2004.
[5] H. Kato et al., "SRAM cell stability under the influence of parasitic resistances and data holding voltage as a stability prober," JSSC, 1997.
[6] H. Qin et al., "SRAM leakage suppression by minimizing standby supply voltage," ISQED, pp. 55-60, 2004.
[7] B.S. Kiyoo Itoh, K. Sasaki, Y. Nakagome, "Trends in low-power RAM circuit technologies," Proc. of the IEEE, pp. 524-43, 1995.
[8] V. Sharma et al., "Ultra low-energy SRAM design for smart ubiquitous sensors," IEEE Micro, pp. 10-24, 2012.
[9] E. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operation," JSSC, pp. 224-31, 1997.
[10] C. Piguet et al., "Low-Power Electronics Design," CRC Press, 2005.
[11] B. Calhoun and A. Chandrakasan, "A 256 kb sub-threshold SRAM in 65 nm CMOS," ISSCC, pp. 628-9, 2006.
[12] B. Rooseleer, W. Dehaene, "A 40 nm, 454MHz 114 fJ/bit Area-Efficient SRAM Memory with Integrated Charge Pump," ESSCIRC, 2013.
[13] O. Andersson et al., "Dual-VT 4kb Sub-VT Memories with <1 pW/bit Leakage in 65 nm CMOS," ESSCIRC, pp. 197-200, 2013.
[14] T. Gemmeke et al., "Cell libraries for robust low-voltage operation in nanometer technologies," JSSE, pp. 132-41, 2013.
[15] L. D. Hung et al,. "Utilization of SECDED for Soft Error and Variation-Induced Defect Tolerance in Caches", DATE, pp 1-6, 2007.
[16] Z. Wang et al., "Design of Memories with Concurrent Error Detection and Correction by Nonlinear SEC-DED Codes", Journal of Electronic Testing, pp. 559-580, 2010.
[17] M. M. Sabry et al., "A Hybrid HW-SW Approach for Intermittent Error Mitigation in Streaming-Based Embedded Systems", DATE, 2012.
[18] M. M. Sabry et al., "OCEAN: An Optiized HW/SW Reliability Mitigation Approach for Scratchpad Memories in Real-Time SoCs", ACM TECS (to appear), 2014.
[19] NXP. NXP ARM-based microntrollers. www.nxp.com/documents/data sheet/LH7A400 N.pdf.
[20] S. J. E. Wilton, N. P. Jouppi, "CACTI : An enhanced cache access and cycle time model," JSSC, pp. 677-88, 1996.
[21] CACTI 6.0: "An integrated cache access time, cycle time, area, leakage, and dynamic power model for uniform and non-uniform cache artchitectures,". http://www.cs.utah.edu/~rajeev/cacti6.
[22] L. Benini et al., "MPARM: Exploring the Multi-Processor SoC Design Space with SystemC", Journal of VLSI Signal Processing Systems, 2005.
[23] A. Theon et al., "Impact of Multi-Gate Device Architectures on Digital and Analog Circuits and its Implications on System-On-Chip Technologies," IEDM, 2013.