# Hybrid Memory Architecture for Voltage Scaling in Ultra-Low Power Multi-Core Biomedical Processors

Daniele Bortolotti[†], Andrea Bartolini[†§], Christian Weis[‡], Davide Rossi[†] and Luca Benini[†§]

[†]DEI, University of Bologna, Italy - Email: {daniele.bortolotti, davide.rossi}@unibo.it

[§]Integrated Systems Laboratory, ETH Zurich, Switzerland - Email: {barandre, lbenini}@iis.ee.ethz.ch

[‡]Microelectronic Systems Design, University of Kaiserslautern, Germany - Email: weis@eit.uni-kl.de

*Abstract*—Technology scaling enables today the design of sensor-based ultra-low cost chips well suited for emerging applications such as wireless body sensor networks, urban life and environment monitoring. Energy consumption is the key limiting factor of this up-coming revolution and memories are often the energy bottleneck mainly due to leakage power. This paper proposes an ultra-low power multi-core architecture targeting eHealth monitoring systems, where applications involve collection of sequences of slow biomedical signals and highly parallel computations at very low voltage. We propose a hybrid memory architecture that combines 6T-SRAM and 8T-SRAM operating in the same voltage domain and capable of dispatching at high voltage a normal operation and at low voltage a fully reliable small memory partition (8T) while the rest of the memory (6T) is state-retentive. Our architecture offers significant energy savings with a low area overhead in typical eHealth Compressed Sensing-based applications.

## I. Introduction and Related Work

Emerging and future healthcare policies are fueling up an application driven shift toward long term monitoring of bio-signals by means of embedded ultra-low power (ULP) devices. Modern human behavior-related diseases, such as cardiovascular diseases, require accurate and continuous medical supervision, which is unsustainable for the traditional healthcare system due to increasing costs and medical management needs [1]. Personal health monitoring systems are able to offer large-scale and cost-effective solutions to this problem.

Wearable health monitoring systems, enabled by Wireless Body Sensor Networks (WBSNs), face contrasting requirements such as a continuously tighter power budget and an increasing demand of computation capabilities to pre-process locally the sensors information so as to reduce the amount of data transmitted, as well as response time. To ensure minimal energy several aspects must be considered, combining optimizations of the signal processing aspects and of the technological layers of the ULP architecture.

Several works in literature [2], [3] show that embedded feature extraction algorithms and data compression schemes greatly contribute to minimizing energy. Compressed Sensing (CS) signal acquisition/compression paradigm has recently proved to be effective in reducing energy consumption in embedded ECG monitors. Enabling a sub-Nyquist sampling rate for sparse signals, authors in [3] show a 37.1% improved lifetime compared to state-of-the-art compression techniques.

At the architectural level, voltage scaling has been widely used and proved its effectiveness though it faces several challenges. Supply voltage has remained essentially constant beyond 65nm and dynamic energy efficiency improvements have stagnated, while leakage currents continue to increase.

Motivated by the inherent parallel nature of medical grade ECG monitoring, where multi-channel signal analysis is often embarrassingly parallel, multi-core architectures proved their efficiency compared to single-core solutions [4], [5]. In [4] authors introduced a multi-core architecture where individual leads are processed on different cores in parallel. Parallel processing enables more aggressive voltage-frequency scaling than single-core solutions, though at low workload requirements the single-core solution proved to be more efficient. The efficiency of the multi-core architecture was further extended in [6], by deploying broadcast mechanism in the instruction memory and clock gating on memories, achieving extra 39.5% power savings at high workload requirements. While at low workload requirements leakage power, mainly due to data and instruction memories, has a big impact and aggressive voltage scaling cannot be applied due to reliability issues for the memories.

Unfortunately, the failure probability of the conventional 6-Transistors (6T) SRAM cell increases considerably as the supply voltage is scaled down [7]. Read failure, due to the lack of Static Noise Margin (SNM), is one of the major failure factors, limiting the efficiency of dynamic voltage scaling. The usage of more reliable SRAM bit-cells, such as 8-Transistors (8T) or 10-Transistors (10T) cells, allows scaling to lower supply voltage, however, such solutions incur in large area penalties (at least 30% overhead for 8T compared to 6T bit-cells [11]).

In the context of CS algorithm, the reliable memory footprint requirement greatly varies according to the different phases of the execution: the *sensing* phase requires the system enough memory to store the sampled data, while the *compressing* phase has a bigger memory footprint to correctly access the data structures used for computation and temporary storage. A typical system performing CS on biomedical signals in real-time, spends most of the time in low workload phases (sensing), while a small portion of its time is spent in high workload phases (compression). In [13], where a single-core CS is implemented in real HW, the ratio between high

workload and low workload phases is below 5%.

These considerations motivate the idea of the present work: using a hybrid memory architecture, combining classic 6T-SRAM cells with 8T-SRAM cells, we are able to offer reliable operation at lower supply voltage. In the sensing phase of the CS execution, the system works in a low-power state (600mV), where only the memory (8T) needed to store sampled data is active and reliable [8], while the other portion (6T) is idle. In this phase the 6T memory has enough hold SNM to be in data-retentive mode [7] though it cannot be correctly accessed. When compression is performed, the system increases its performance, operating at a higher voltage (1.2V) and the whole 6T/8T memory is active and reliable.

The concept of hybrid memory has already been introduced in literature [9], [11]. The work presented in [11] tolerates an error on the computation related to the 6T memory when operating at low voltage, while in our architecture such behavior would compromise execution correctness. Moreover, their approach is highly customized for the specific application, avoiding the usage of standard memory compilers. In [9] authors propose a cache architecture with ways capable of operation at near-threshold voltage. The usage of separate voltage domains for cores, 6T and 8T cache ways has a non negligible overhead on the area, making it not feasible for scratchpad memories [10]. Our architecture can therefore benefit from using a single voltage domain, adapting its operating point to different workload scenarios.

The main contributions of this paper are the following:

- a novel hybrid memory architecture for ULP multi-core biomedical processors is proposed. The combination of 6T and 8T-SRAM banks enables aggressive power management during workload phases with low memory usage and low computational requirements.
- the proposed architecture leads to a significant improvement in energy saving ($\approx 25\%$ in a typical scenario) when compared to a standard architecture that uses solely 6T-SRAM banks.
- we demonstrate that our solution has a negligible area overhead ($\approx 2\%$) with respect to the baseline solution making it preferable to a solution with only 8T-SRAM due to its higher area overhead.

The rest of the paper is organized as follows. In Section II the baseline architecture is introduced. Section III discusses the main features of CS algorithm and execution and describes the proposed hybrid memory architecture for ULP biomedical processors. Next, in Section IV we describe the experimental setup and the results of the comparative study of our architecture with the baseline in terms of energy efficiency and area overhead. Finally, the conclusions of this work are presented in Section V.

## II. CS ARCHITECTURE

We consider a baseline architecture similar to several current multi-core architectures targeting biomedical signals

processors [6], [5]. The considered architecture, presented in Figure 1, features 8 Processing Elements (PEs) each one with a private Instruction Cache. The PEs do not have private data caches, therefore avoiding memory coherency overhead, while they all share a L1 multi-banked tightly coupled data memory (TCDM) acting as a shared data scratchpad memory. The TCDM has a number of ports equal to the number of banks to have concurrent access to different memory locations.

Intra-cluster communication is based on a high bandwidth logarithmic interconnect (LIC). It consists of a Mesh-of-Trees (MoT) interconnection network able to support single-cycle communication between PEs and memory banks (MBs), resembling the hardware module presented in [14]. In case of multiple conflicting requests, for fair access to memory banks, a round-robin scheduler arbitrates the accesses. To ease the negative impact of banking conflicts we consider a banking factor of 2 (16 banks). Moreover, to reduce memory access time and increase shared memory throughput, PEs can benefit from the broadcast mechanism of the interconnect.

The DMA shown in Figure 1 is in charge of periodically moving the data sampled by the analog front-end (AFE) buffer to the TCDM making it available to the multi-core processor to perform compression.

## III. HYBRID MEMORY ARCHITECTURE

In this section the baseline multi-core ULP architecture to perform Compressed Sensing (CS) on biomedical signals is presented. We introduce then the CS phases with a qualitative analysis on their characteristics in terms of memory footprint and processing requirements. Finally the proposed memory architecture is presented.

### A. Compressed-Sensing Application

Typical WBSNs-based biomedical applications require to sense biological signals from the patient (i.e. ECG, EMG, EEG) and send them to a more powerful computing node for further analysis. The recently-developed Compressed Sensing (CS) theory states that sparse (and thus compressible) signals can be reconstructed from a smaller number of samples than required by Nyquist sampling frequency [3]. By deploying this sparsity property, which applies to many classes of biomedical
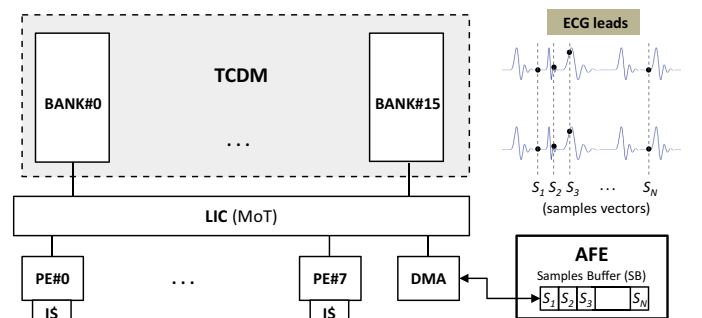


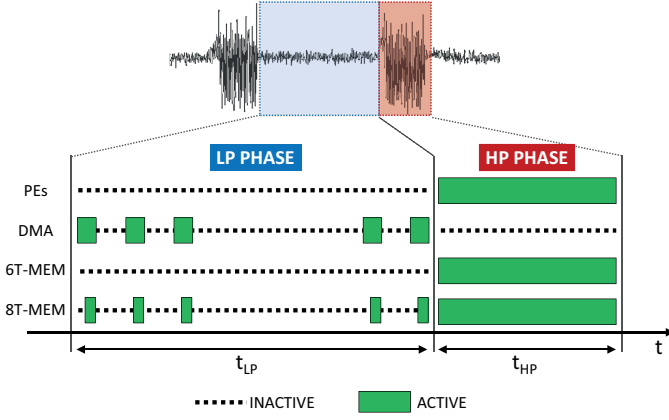Fig. 1: Baseline multi-core architecture for CS

Fig. 2: Active/inactive architectural elements during CS execution (LP and HP phases)

signals, the CS paradigm can be suitable for implementing low-resource sensor applications [2], since it reduces the amount of samples required in processing and storage.

In the hereby considered CS architecture, the input multi-channel signal is sampled by the analog front-end (AFE), with a sampling frequency ($f_s$) according to the dynamics of the signal to analyze and the accuracy needed. The samples ($s_i$), corresponding to different leads, are stored in a buffer inside the AFE. Once the values are sampled, the DMA is triggered to move the samples from the buffer to the local memory of the CS multi-core processor. Then CS compression algorithm starts, where each core operates on its own subset of the sampled data. We assume that the computation phase must be completed before the first sample of the next window ($N+1$) is available to avoid double buffering overhead.

Such CS application, similarly to other sensor-data based computation, is composed of two phases: *data collection* and *computation*. The first phase is characterized by low-workload/low-memory requirements and a long duration, thus it will be referred as *LP Phase* (Low Performance). The latter instead will be named *HP Phase* (High Performance). This concept is depicted in Figure 2 where data collection and computation are shown.

*Data Collection (LP Phase):* During the data collection phase the ULP processor waits for the number of samples ($N$) required to perform CS computation. Considering typical sampling frequencies for biomedical signals, this phase exceeds in time the phase of computation. For instance, with $f_s = 250Hz$ and $N = 512$, the data collection phase lasts 2048 ms. During data collection the only requirement for the architecture is to make available enough memory to store locally the data sampled by the AFE. It is clear that during this phase for most of the time the system is idle thus requiring a ultra-low power state to avoid unnecessary consumption. Figure 2 shows a timing diagram of the status of the architectural elements during the LP phase. The only active elements are the DMA and the portion of the TCDM memory where samples are

moved for future elaboration. The required active memory, varies according to system specification (sampling frequency, compression algorithm).

*Computation (HP Phase):* Once the data collection phase is over, the DMA has already copied the buffer with $N$ samples to the local (TCDM) memory and the computation phase starts. As introduced before the considered architecture performs a burst of computation on the available data for future transmission. During this phase the system is in an operating point characterized by high workload requirements and high memory footprint. All the processing elements are active and working on the data sampled during the last observation window. The amount of active memory required in HP phase is higher then in LP Phase because of all data structures needed to perform the convolution kernel of the CS algorithm (Section IV-A). Moreover, considering that the compression kernel is memory-bound by nature, the bandwidth requirements in core-memory bandwidth implies higher supply voltage for the memory in order to sustain the throughput.

### B. 6T/8T Hybrid Architecture

Considering the limitation imposed by classic 6T-SRAM memory when operating aggressive voltage scaling and the characteristics of biomedical applications, as outlined in the previous section, we consider an alternative memory architecture. By combining 6T and 8T-banks the reliable operating range is further extended to lower supply voltage. The proposed 6T/8T hybrid architecture is schematized in Figure 3 and compared to the baseline architecture introduced in Section II, it features:

- single voltage domain for the whole architecture. This reduces area overheads and design complexity.
- 8T portion of the TCDM (*LP memory*) able to offer reliable operation down to 600mV.
- 6T portion of the TCDM with reliable access down to 800mV but able to operate in data retentive mode (sufficient hold SNM) at 600mV.
- at voltages higher than 800mV all the TCDM (6T + 8T) operates correctly (*HP memory*).
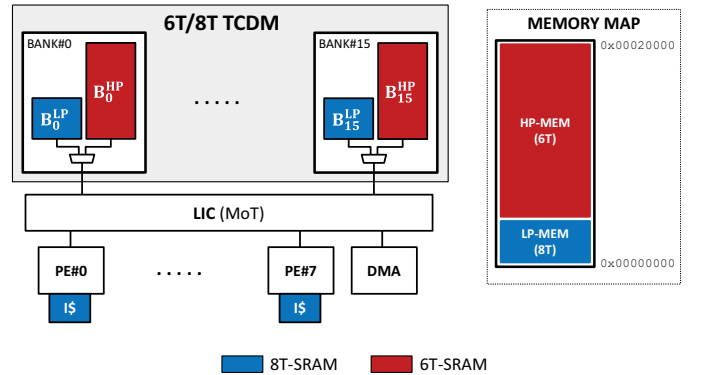


Fig. 3: Hybrid 6T/8T memory architecture and memory map

- the interleaving on different banks operated by the logarithmic interconnect (Section II) enables to have a contiguous memory map among the 6T and 8T portions. This concept is depicted in Figure 3.

## IV. EXPERIMENTAL SETUP AND RESULTS

In this section we present the experimental setup and the results of the evaluation of the proposed hybrid memory architecture in terms of energy efficiency and area overhead.

### A. CS Algorithm Analysis

The reference benchmark considered in this work is a real-time multi-lead ECG processing application composed of two main kernels: *Compressed Sensing* (CS) and *Huffman Coding* (HC). The CS kernel [3] performs compression (50% ratio) on a block of 512 samples of ECG data per lead with a sampling period of 4ms. The HC kernel performs the Huffman encoding on the compressed data, reducing its footprint further for wireless transmission [3]. The CS algorithm operates on 8 leads in parallel where each Processing Element (PE) works on a separate lead data-set. The CS part has a constant program flow without any dependence on the input data, while the HC part adds a short section of data-dependent program flow. Considering a single lead ECG, the memory footprint of the CS algorithm consists of 648 bytes for instructions and 16 KB for data. The data section consists of two contributions: working data (the samples) and read-only data with a memory footprint of 2048 and 14336 bytes respectively. More in detail the read-only data consists of 3 Look-Up Tables (LUTs), i.e. a vector of random coefficients for the CS kernel (12288 bytes) and two data dependent LUTs (1024 bytes each) for the HC kernel.

Such CS algorithm analysis was used at design time to choose the appropriate memory cuts, for both baseline and hybrid architectures, and to statically allocate the memory structures. The TCDM size is assumed to be 128KB in both architectures, while an instruction cache of 1KB (per-core) is chosen. Considering that during compression every core operates on 512 samples, the 8T-SRAM memory (where sampled data is stored during LP phase) is chosen to be 16KB with 16 banks of 1KB each.

### B. Hybrid Memory Analysis

Table I shows the power numbers (dynamic and leakage) considered for the evaluation of the proposed architecture. For 6T/8T memories the power values were extracted from the data-sheets of the respective SRAM architectures for a low power 65nm technology library. The memory numbers reported here refer to 1024x32 bits arrays (mux column = 4). The idle power is the standby power of the SRAM, where only the clock and address pins are toggling. Write and read power were measured with 100% activity (back to back cycling), with half of the address and data inputs (only for write) toggling. All inputs are stable (no toggling) for deriving the leakage power. We further assumed the worst case for leakage (i.e. best

TABLE I: 6T/8T memories and PE energy numbers

| DYNAMIC [µW/MHz] | | | | | |
|---|---|---|---|---|---|
| 6T-MEM | | 8T-MEM | | PE | |
| HP | LP | HP | LP | HP | LP |
| IDLE 2.20 | 0.54 | 2.32 | 0.56 | | |
| READ 11.79 | 2.87 | 12.04 | 2.93 | 68.76 | 16.74 |
| WRITE 13.88 | 3.38 | 14.11 | 3.43 | | |
| LEAKAGE [µW] | | | | | |
| 6T-MEM | | 8T-MEM | | PE | |
| HP | LP | HP | LP | HP | LP |
| -40 C 0.61 | 0.31 | 0.27 | 0.13 | 0.63 | 0.32 |
| 25 C 11.56 | 5.89 | 5.35 | 2.63 | 11.18 | 5.69 |
| 125 C 326.77 | 166.23 | 158.77 | 80.77 | 338.44 | 172.17 |

case for the technology). 8T cells considered here are Low-Leakage (LL) cells, a register-file architecture, which offer better performance and reliability. On the other hand, for the 6T-SRAM, the LL cells incur in reliability problems when reducing the supply voltage to 600mV [7].

For the Processing Element (PE), we considered an average active energy of 68.76 µW/MHz and 16.74 µW/MHz when operating at 1.2V and 0.6V, respectively. These numbers are based on post-synthesis characterization of an openRISC core. For the DMA and the logarithmic interconnect our characterization estimates 63.13 µW/MHz and 54.73 µW/MHz respectively at 1.2V as average active energy (15.37 µW/MHz and 13.13 µW/MHz, respectively, at 0.6V). Comparing the number of NAND equivalent gates of the DMA and the 8x16 interconnect with respect to a single PE, we derived corrective factors for the leakage power equals to 0.92x and 2.19x, respectively. Leakage power is scaled to 0.6V considering the relation expressed in [12].

### C. Area Overhead (iso-size)

To evaluate the area overhead of our solution, in an *iso-size* comparison, we quantified the overhead introduced by the 8T memory portion in the hybrid architecture compared to the baseline (6T-only) solution. Table II shows the impact of each element on total area.

The overhead of extra-circuitry for the hybrid memory, required by the separation of logical banks in 6T and 8T banks is negligible, leading to a total overhead below 2%. If instead we consider an architecture with only 8T-SRAM, the overhead

TABLE II: Area comparison (Hybrid vs Baseline)

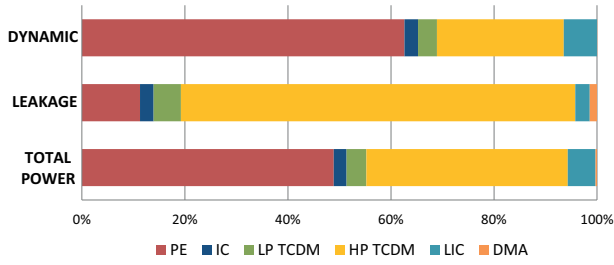| ELEMENT | HYBRID [$mm^2$] | BASELINE [$mm^2$] |
|---|---|---|
| PEs | 0.85408 | 0.85408 |
| 6T TCDM | 0.70652 | 0.80746 |
| 8T TCDM | 0.13323 | - |
| 6T I$ (DATA) | - | 0.05047 |
| 8T I$ (DATA) | 0.06662 | - |
| DMA | 0.09801 | 0.09801 |
| LOGINT 8X16 | 0.23348 | 0.23348 |
| TOTAL | 2.09194 | 2.04349 |

Fig. 4: Power breakdown for HP phase (hybrid, T=25°C)



Fig. 6: Power breakdown for LP phase (hybrid, T=25°C)

on the overall system would be non negligible ($\approx 14\%$) and leakage contribution would affect the energy efficiency.

### D. Hybrid Memory Efficiency

To evaluate the energy efficiency of the proposed architecture, the power numbers of Section IV-B have been integrated in a SystemC-based cycle-accurate virtual platform [15]. The architecture was configured with 8 cores, 1 DMA, an 8x16 logarithmic interconnect and 6T/8T portions as determined in Section IV-A. The HP phase is performed in 94.56k clock cycle, while the LP phase takes 24.12k clock cycles (sum of all DMA data movements in an observation window).

### HP Phase

The first set of experiments was aimed at comparing the energy efficiency of the proposed 6T/8T hybrid memory architecture to the baseline case of an ULP multi-core architecture where all the TCDM is composed of 6T-SRAM cells. During the HP phase, all cores are active and executing the CS kernels described in Section IV-A operating in parallel on its separate data set. On the memory side, the whole TCDM memory is active, as well as the I-caches. The DMA is idle, contributing only for leakage power. The operating point considered in this experiment is a clock frequency of 100 MHz and a supply voltage of 1.2V. In Figure 4 a power breakdown for the hybrid architecture (T=25°C) is shown. Total power consumption has two main contributions: PEs and HP TCDM (6T-SRAM) as
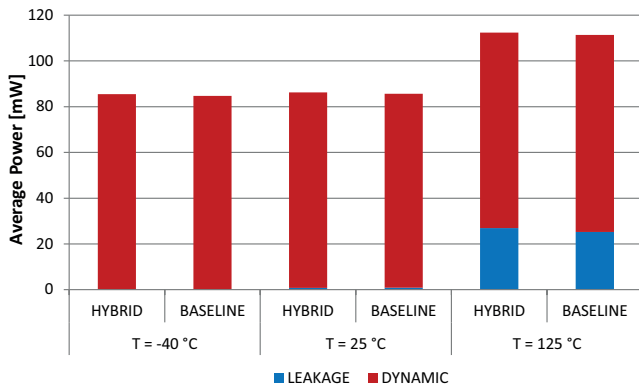
expected. The number of accesses in the HP portion of the TCDM exceeds the number of accesses in the LP portion, mainly due to data structures of the CS kernels and stack. For completeness a separated breakdown for dynamic and leakage is presented, though the dynamic power contributes for 99% to total power. Figure 5 shows the average power during HP phase for the baseline and the proposed architecture. At different temperature leakage contribution (exacerbated at T=125°C) impact both architectures, though the 8T Low Leakage (LL) cells can amortize this effect. As expected in the HP phase our solution has a lower energy efficiency compared to the baseline, mainly due to the higher contribution of dynamic power for the 8T-memory. The impact of the hybrid architecture in the HP phase is very low, being below 1% for all the considered temperatures.

### LP Phase

As a second experiment we compared the energy efficiency of our solution and the baseline during the data collection phase. During the LP phase all cores are idle waiting for the sampled data to be ready. Only the amount of memory needed to store the samples is active, while the other portion of memory is clock gated, contributing only for leakage. The DMA is in charge of moving the sampled data from the AFE buffer to the LP-portion of the memory. The operating frequency considered in this phase is 10 MHz. For a fair comparison with the baseline, we consider only 16KB active of



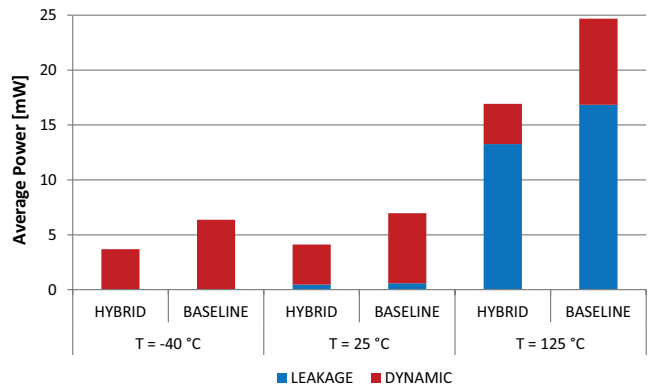Fig. 5: Power comparison for HP phase (baseline vs hybrid)



Fig. 7: Power comparison for LP phase (baseline vs hybrid)

TCDM, with the other portion being clock-gated. Considering reliability issue for 6T-SRAM [7], the baseline has a supply voltage of $V_{dd} = 0.8V$, while our solution thanks to the higher reliability of 8T-SRAM memory can operate at 0.6V. For completeness, in Figure 6 is shown a breakdown of total power for the hybrid architecture at the temperature of 25°C.

Figure 7 shows the average power during the LP phase for the baseline and the proposed architecture at different temperatures. These results confirm the effectiveness of the proposed solution: thanks to the extended voltage scaling range offered by the reliability of 8T-SRAM the dynamic component can be greatly reduced. At T=25°C the overall reduction of power compared to the baseline is 24.5%.

*Overall Efficiency:* The last set of experiments was intended to evaluate the efficiency of the 6T/8T hybrid memory architecture varying the amount of time spent in LP and HP phases. The average power consumption shown before demonstrates a good improvement for the proposed solution in the HP phase and a small penalty in the HP phase but is not taking into consideration the time spent in the two phases during a period of Compressed Sensing. The results of this analysis are presented in Figure 8, where on the x-axis is shown the ratio between HP and LP phases and on the y-axis is shown the energy efficiency of the hybrid architecture with respect to the baseline.

The proposed solution improves energy efficiency of the system for the range 0-90% of HP/LP ratio, with a crossing point at $\approx 90\%$ where the baseline architecture outperforms the hybrid solution. Considering a typical scenario with a 5% ratio between HP and LP phases [13], the proposed solution proves to be $\approx 25\%$ more efficient than the baseline architecture. This result is valid on the whole temperature range considered. The quadratic trend in efficiency validates the motivation behind the work. Power consumption has a quadratic dependency on supply voltage for the dynamic component and increasing the amount of time spent in LP phase, the more effective becomes the aggressive voltage scaling that can be operated on our hybrid architecture.
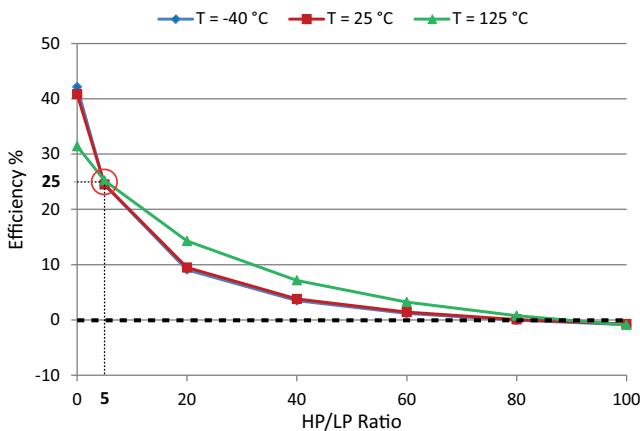


Fig. 8: Hybrid vs Baseline efficiency varying HP/LP ratio

## V. CONCLUSIONS

In this work we introduce a 6T/8T hybrid memory architecture for multi-core biomedical processors. Classic memory architectures composed of 6T-SRAM memories face reliability issues when reducing supply voltage to threshold. Static noise margin for such memory cells compromise execution correctness making aggressive voltage scaling not feasible. The proposed architecture greatly benefits from the varying workload/memory footprint requirements of biomedical processing, adapting in a reliable way to different operating points. Our solution offers significant improvements in energy saving ($\approx$ 25% in a realistic scenario) when compared to a 6T-only architecture with a negligible ($\approx 2\%$) area overhead.

### REFERENCES

[1] World Health Organization [Online] URL: http://www.who.int/mediacentre/factsheets/fs317/en/index.html.

[2] F. Rincon et al., *"Development and evaluation of multilead wavelet-based ECG delineation algorithms for embedded wireless sensor nodes"*, In: TITB, vol. 15, no. 6, pp. 854–863, Nov. 2011.

[3] Mamaghanian, H. et al., *"Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes"*, In: IEEE Transactions Biomedical Engineering, vol. 58, no.9 pp. 2456–2466, 2011.

[4] Dogan A.Y. et al., *"Power/Performance Exploration of Single-core and Multi-core Processor Approaches for Biomedical Signal Processing"*, In: Proc. of PATMOS, 2011.

[5] Dreslinkski, R. G., et al., *"An energy efficient parallel architecture using near threshold operation"*, In: Proceedings of the 16th International Conference on Parallel Architecture and Compilation Techniques. IEEE Computer Society, 2007.

[6] Dogan A.Y. et al., *"Multi-core architecture design for ultra-low-power wearable health monitoring systems"*, In: Proc. of the ACM/IEEE DATE, 2012.

[7] Calhoun, B. H. et al., *"Analyzing static noise margin for sub-threshold SRAM in 65nm CMOS"*, In: Solid-State Circuits Conference, 2005. ESSCIRC 2005. Proceedings of the 31st European. IEEE, 2005.

[8] Verma, N. et al., *"A 256 kb 65 nm 8T subthreshold SRAM employing sense-amplifier redundancy"*, In: Solid-State Circuits, IEEE Journal of 43.1 (2008): 141-149.

[9] Dreslinkski, R. G. et al., *"Reconfigurable energy efficient near threshold cache architectures"*, In: Microarchitecture, 2008. MICRO-41. 2008 41st IEEE/ACM International Symposium on. IEEE, 2008.

[10] Mak W. et al., *"Voltage island generation under performance requirement for soc designs"*, In: Proc. of the Asia and South Pacific Design Automation Conference (ASPDAC), IEEE Computer Society, 2007.

[11] Chang I.J. et al., *"A Priority-Based 6T/8T Hybrid SRAM Architecture for Aggressive Voltage Scaling in Video Applications"*, In: IEEE transactions on circuits and systems for video technology, vol. 21, no. 2, Feb 2011.

[12] Roy, K. et al., *"Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits."*, In: Proceedings of the IEEE 91.2 (2003): 305-327.

[13] K Kanoun, et al., *"A Real-Time Compressed Sensing-Based Personal Electrocardiogram Monitoring System"*, In: Proc. of the ACM/IEEE DATE, 2011.

[14] Rahimi A. et al., *"A Fully-Synthesizable Single-Cycle Interconnection Network for Shared-L1 Processor Clusters"*, In: Proc. of the ACM/IEEE DATE, 2011.

[15] Bortolotti D. et al., *"VirtualSoC: a Full-System Simulation Environment for Massively Parallel Heterogeneous System-on-Chip"*, In: Proceedings of IPDPWS 2013.