# A sub 10 pJ/cycle over a 2 to 200 MHz Performance Range RISC-V Microprocessor in 28 nm FDSOI

Roel Uytterhoeven and Wim Dehaene
KU Leuven, ESAT-MICAS
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
Email: Roel.Uytterhoeven@esat.kuleuven.be, Wim.Dehaene@esat.kuleuven.be

***Abstract*—This work presents a near-threshold microprocessor implementation that aims for ultra-high energy-efficiency, while providing MHz performances over a 100x frequency range. This allows the processor to target a wide range of applications and to provide edge computing capabilities to IoT-devices. The RISC-V IM32 ISA is implemented in 28 nm FDSOI technology. FDSOI's body-biasing capabilities are explored and result in performance specific energy optimization possibilities. Furthermore, near-threshold libraries that are based on the original technology's super-threshold standard-cells, allow for a swift implementation of the processor without the need to design a custom low-voltage standard-cell set. Measurements show that the processor core breaks the 10 pJ/cycle energy barrier over a frequency range from 2-200 MHz. The design remains functional down to 250 mV with a MEP of 4.18 pJ/cycle at 21.5 MHz (380 mV). The 2 × 32 kB on-chip SRAM consumes between 4.49 pJ/cycle and 9.58 pJ/cycle while operating at 0.5 V.**

## I. Introduction

The next powerful revolution in the electronic world is expected to be the explosion of the Internet of Things (IoT). Estimates suggest an increment of around 50 billion IoT-devices within the next few years [1]. Under such massive growth, conventional cloud computing suffers from a severe data bandwidth bottleneck and enormous energy costs related to raw-data transmissions. To resolve this issue, edge computing [1] tries to maximize local data-processing on the IoT-device. This drastically reduces the necessary communication bandwidth and associated energy losses.

However, edge computing significantly increases the required processing capabilities for microprocessors in IoT-devices. Furthermore, consumer-demands lead us towards ever smaller, more portable and faster devices. These two trends result in a strong interest towards more energy-efficient, yet high speed microprocessors.

Over the past years, several research examples have shown that the optimal energy consumption per operation in digital circuits can be achieved by using near- or sub-threshold supply voltages [2], [3]. However, at low voltage, both the circuit's delay and its process-variation resilience deteriorate strongly. This results in slower processors that waste a significant amount of time and energy in design margins. To overcome this, both variation resilient standard-cells designs [4] and timing-error detection techniques [5] have been explored. Yet, it remains challenging to achieve high processing speeds combined with high energy-efficiency while providing performance scalability.

This work implements a RISC-V microprocessor in 28 nm Fully Depleted Silicon on Insulator (FDSOI). The implementation aims at ultra-high energy-efficiency in a medium performance range with a large frequency scalability. The processor architecture is the open-source Zscale [6], which uses the RISC-V IM32 Instruction set Architecture (ISA). A near-threshold gate-level comparison between the Zscale and the commercial ARM Cortex-M0 provides benchmarking insights. Further, the Silicon on Insulator (SOI) body-biasing capabilities and their related effects on energy-efficiency are examined. Lastly, it is shown that typical super-threshold technology libraries can be enabled to implement a near/sub-threshold processor design.

This paper is organized as follows. First, section II discusses the Zscale implementation with a focus towards body-biasing impact, ultra-low voltage and system architecture. Second, section III presents the silicon measurements.

## II. Design implementation

### A. Body-biasing considerations

The high-level formulas below study the impact of body-biasing on the circuit's timing.

$$V_T \sim 85\,\mathrm{mV}/V_{bb} \tag{1}$$

$$I_{on} \approx I_0 e^{\frac{V_{dd}-V_T}{nV_{th}}} \tag{2}$$

$$I_{off} \approx I_0 e^{\frac{-V_T}{nV_{th}}} \tag{3}$$

$$T_d \approx \frac{CV_{dd}}{I_{on}-I_{off}} \approx \frac{CV_{dd}}{I_{on}} \tag{4}$$

Combining expressions 2 and 4 shows that the same critical data-path delay can be achieved with different combinations of $V_{dd}$ and $V_T$. This enables operation at a constant clock frequency while changing the supply voltage, when the body-biasing voltage is adapted correctly. The $V_{dd}$ and $V_{bb}$ combinations that maintain a constant clock frequency form an iso-performance line.

In order to study the sensitivity of $V_{dd}$, ie. the slope of the iso-performance lines, expression 5 sets $V_{dd}$ in terms of $V_T$, which has a known relation to $V_{bb}$ as shown in expression 1.

$$V_{dd} = V_T + nV_{th} \ln \frac{V_{dd}}{A} \text{ with } A = \frac{T_d I_0}{C} \tag{5}$$

From expression 2 and 4 it can be seen that $A \approx V_{dd}$ if $V_{dd} \approx V_T$. Hence, in near-threshold, the iso-performance lines are expected to behave almost linear with a slope of $85\,\mathrm{mV}/V_{bb}$.

978-1-5386-5404-0/18/$31.00 ©2018 IEEE

Next, the impact on energy consumption is considered.

$$E_{tot} = E_{dyn} + E_{leak} \tag{6}$$

$$E_{dyn} = \alpha C {V_{dd}}^2 \approx \alpha C {V_T}^2 \text{ in near-threshold} \tag{7}$$

$$E_{leak} = V_{dd} T_d I_{off} \approx V_t T_d I_0 e^{\frac{-V_T}{nVth}} \text{ in near-threshold} \tag{8}$$

Expressions 7 and 8 highlight a trade-off: forward body-biasing (FBB) reduces the dynamic energy, whereas reverse body-biasing (RBB) reduces the leakage energy. The exact optimum of this trade-off depends on the delay $T_d$ and the activity $\alpha$. Unfortunately, a choice must be made between FBB and RBB at design-time, as the body-biasing range is bounded by the transistors' well-structure. Since this design aims for efficiency at high speed, LVT-transistors are chosen with a resulting FBB range from $-0.3$ to $+1\,\mathrm{V}$.

### *B. Ultra low-voltage implementation*

The exponential term in expression 2 reveals near-threshold's inherent sensitivity to variations on $V_{dd}$ or $V_T$. As a consequence, random dopant fluctuations that result in small $V_T$ shifts, have a strong effect in bulk technologies [7]. However, FDSOI's fully depleted channel does not have such dopants, making the technology intrinsically more immune to variability and thus ideal for low-voltage designs.

Yet, technology vendors only provide standard-cell libraries in the super-threshold supply range. To avoid designing a new digital-cell library, this work does a spice-level low-voltage evaluation of the $28\,\mathrm{nm}$ FDSOI standard-cell netlists, assessing the near-threshold capabilities of these cells. This evaluation learns that these standard-cells remain fully functional down to $250\,\mathrm{mV}$. This makes them suitable for near-threshold operation without physical modifications. Libraries are obtained by an in-house characterization of the core standard-cells at multiple near-threshold voltages using *Siliconsmart*.

Besides the standard-cells, the design's timing constraints have an impact on both functionality and efficiency. First, setup-constraints are considered, next hold-constraints.

Setup-constraints can be set forward by the designer and mainly impact the optimizations performed by the implementation tools. Relatively lax setup-constraints allow for more area and power optimizations, whereas stricter constraints result in higher performance at the cost of power and area. The setup-constraints in this work try to optimize this trade-off by pushing the speed towards the point where the number of critical paths starts to become too excessive. This, in combination with the chosen LVT-transistors, allows high speeds at low-voltage without provoking too much energy-inefficient timing optimizations. Note that violations to setup-constraints do not necessarily impede circuit functionality.

Hold-constraints prevent destructive timing mismatch between the clock and fast propagating data-paths. At low-voltage, this timing mismatch grows significantly due to variability, which results in an increased amount of delay buffers associated with hold-fixing. Further, contrary to setup-issues, persisting hold-issues impede circuit functionality as reducing the clock speed cannot resolve this type of issue. Hence, a considerable degrading factor for hold-timing is taken in order to account for the extra uncertainty in near-threshold. This leads to even more additional delay-buffers. To minimize the energy impact of all these delay-buffers, this work implements dedicated ultra-low energy delay-buffers. These custom delay-buffers employ stacked transistors to achieve maximal delay with minimal additional switching energy.

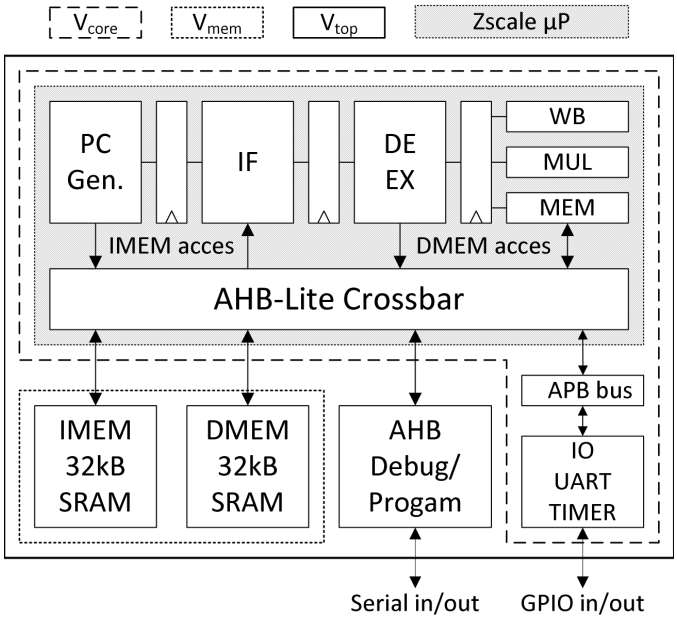



Fig. 1. Block diagram of the System on Chip.

### *C. System architecture*

Figure 1 gives an overview of the complete System on Chip. The Zscale core is extended with the following memory mapped peripherals: GPIO, uart and timer. These peripherals reside in the same near-threshold power domain as the Zscale processor's core. This power domain will from now on be considered as the system's core power domain. All logic in this power domain is implemented using the near-threshold libraries. Further, a debugger/programmer that can take over the AHB-bus, resides in a top-level super-threshold power domain. Finally, two equally sized SRAM blocks provide the Zscale with independent instruction and data memory access.

A gate-level comparison with the ARM Cortex-M0 (including a single cycle multiplier and similar perhiperals) allows to benchmark the Zscale architecture. Both designs are

TABLE I
COMPARISON CORTEX-M0 AND ZSCALE AT GATE-LEVEL

| | Cortex-M0 | Zscale | Difference |
|---|---|---|---|
| DMIPS | 0.9 | 1 | + 11% |
| Area | 11404 | 14497 | + 27% |
| Energy @MEP | $1.5\,\mathrm{pJ/cycle}$ | $1.9\,\mathrm{pJ/cycle}$ | + 27% |

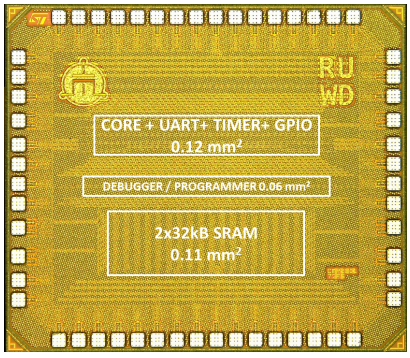



Fig. 2. Die photograph with area markings.

synthesized at 100 MHz using a 0.5 V near-threshold library. Post-synthesis gate-level simulations with the full set of near-threshold libraries locate the MEP of both designs at 350 mV. Table I compares the designs in this MEP. The Zscale achieves a slightly higher Dhrystone mark at the cost of a 27 % increase in both area and energy consumption.

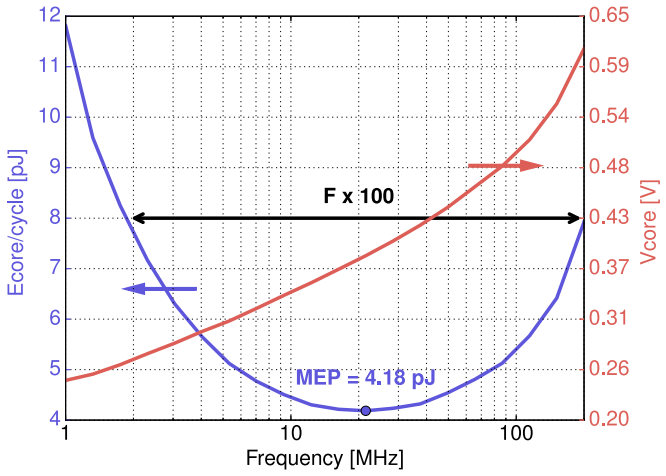
## III. Silicon Measurements

Figure 2 shows the die photograph of the fabricated chip in 28 nm FDSOI. A Xilinx zedboard drives the measurement setup. This board controls the clock, the programmer, a dedicated supply board and ADCs measuring the energy consumption. This allows to automate the measurement procedure, ensuring equal treatment of all samples. A total of 20 silicon samples is measured at 21.5 °C. All 20 samples show correct functionality, ruling out the possibility of hold-issues in the design. Reported numbers are averages over these samples. For all measurements the Dhrystone benchmark is used as test program on the Zscale.

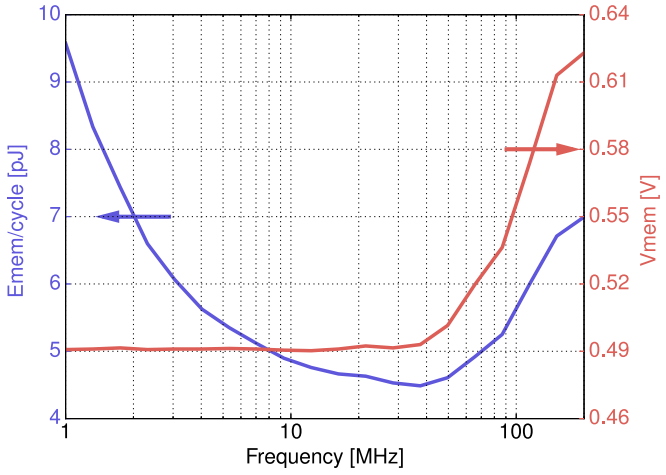
Figure 3a shows the design's timing-critical voltage and energy consumption in a performance window from 1 MHz to 200 MHz without body-biasing ($V_{bb} = 0$ V). The measured core energy consumption breaks the sub 10 pJ/cycle barrier over almost this entire range, resulting in an ultra-efficient and ultra-wide performance range. The MEP of the core is found at 21.5 MHz and 4.18 pJ/cycle. At 1 MHz the design reaches its minimal functional voltage of 250 mV. Even though the design suffers from leakage in deep sub-threshold, it is clear that in near-threshold the core works optimal over a wide range that provides MHz clock frequencies.

Figure 3b shows the same information for the $2 \times 32$ kB SRAM blocks. The energy-efficient range of the SRAM blocks coincides with the core's optimal range, resulting in a complete processor that exhibits the same wide energy-efficient performance range. The inherent leakage optimizations of the SRAM blocks only start to make a difference at low speeds. Note that $V_{MEM}$ scaling is stopped at 0.5 V to guarantee correct retention under all conditions.

Figure 4 shows the aforementioned iso-performance lines that result from body-biasing. These lines are almost perfectly linear, confirming the expectations based on the high-level formulas. The energy consumption along these iso-performance




(a)




(b)

Fig. 3. Energy at critical supply voltage over the entire frequency range without body-biasing. (a) the Zscale core. (b) the Zscale SRAM blocks.

lines exhibits a clear minimum at higher performances. In this point the trade-off between dynamic and leakage energy is optimal. However, due to the design's relatively low activity, the dynamic energy scales rather slowly with increasing supply voltage as visualized by the dynamic energy curves in figure 4. As a consequence, the resulting minimums are but slight improvements over the original no body-biasing points. Further, at lower performances, the impact of the dynamic energy is too low and the leakage impact too high to create similar minimums. To enlarge the body-biasing gains, future work could explore several options that increase the dynamic energy scaling or reduce the leakage energy while minimizing the impact on the circuit's speed.

Finally, the comparison in table II shows that this work outperforms the current state of the art in terms of performance scalability below 10 pJ/cycle energy consumption. This also results in a extremely low minimal energy delay product (EDP). Note that some other works use the Cortex-M0+ architecture which is inherently more energy-efficient.

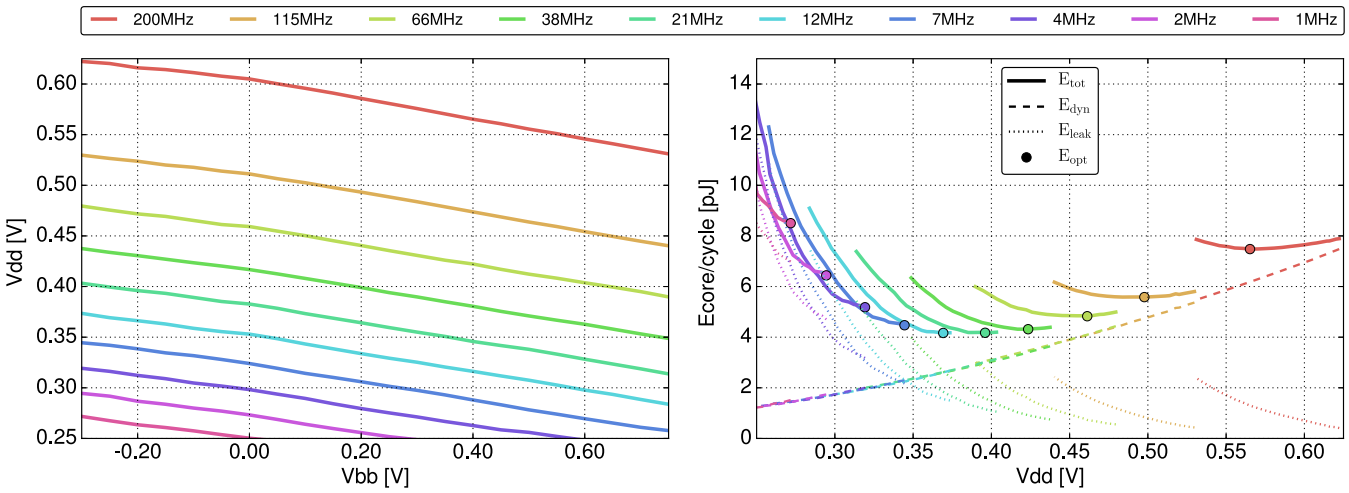



Fig. 4. Left: Iso-performance lines ranging from 1 MHz to 200 MHz. Right: Zscale core energy/cycle along these different iso-performance lines.

## IV. Conclusion

This paper presents a near-threshold 32-bit RISC-V microprocessor in 28 nm FDSOI. The processor core breaks the 10 pJ/cycle energy barrier over a 100x (2-200 MHz) performance range. This range makes the design ideal for edge computing oriented applications that exhibit high performance requirements combined with scalability and energy-efficiency. Further, it allows the design to fit an extensive set of applications that would normally require more dedicated devices, enabling significant economical savings through reusability. Besides, this work proves that conventional standard-cell libraries can be enabled for near-threshold designs through recharacterization at lower voltage. This effectively avoids the need for a custom near-threshold library, facilitating the design of low voltage circuits.

## Acknowledgment

The authors would like to thank Andreia Cathelin and Sylvain Clerc from ST-Microelectronics and Jean-François Paillotin from CMP for chip fabrication. Further, the authors would like to thank the Research-Foundation-Flanders (FWO) for the SB-fellowship that supports Roel Uytterhoeven's work.

## References

[1] T. Adegbija *et al.*, "Microprocessor Optimizations for the Internet of Things: A Survey," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 1, pp. 7–20, jan 2017.

[2] J. Myers *et al.*, "An 80nW retention 11.7pJ/cycle active subthreshold ARM Cortex-M0+ subsystem in 65nm CMOS for WSN applications," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, feb 2015, pp. 144–145.

[3] G. Lallement *et al.*, "A 2.7pJ/cycle 16MHz SoC with 4.3nW power-off ARM Cortex-M0+ core in 28nm FD-SOI," in *IEEE European Solid State Circuits Conf. (ESSCIRC)*, sep 2017, pp. 159–162.

[4] H. Reyserhove and W. Dehaene, "A Differential Transmission Gate Design Flow for Minimum Energy Sub-10-pJ/Cycle ARM Cortex-M0 MCUs," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 7, pp. 1904–1914, 2017.

[5] ——, "Design margin elimination in a near-threshold timing error masking-aware 32-bit ARM Cortex M0 in 40nm CMOS," in *IEEE European Solid State Circuits Conf. (ESSCIRC)*, sep 2017, pp. 155–158.

[6] Y. Lee, "Z-zscale: Tiny 32-bit RISC-V systems," 2015. [Online]. Available: https://riscv.org/wp-content/uploads/2015/06/riscv-zscale-workshop-june2015.pdf

[7] R. Rithe *et al.*, "The effect of random dopant fluctuations on logic timing at low voltage," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 5, pp. 911–924, 2012.

[8] U. Banerjee *et al.*, "An energy-efficient reconfigurable DTLS cryptographic engine for End-to-End security in iot applications," in *IEEE International Solid - State Circuits Conf. (ISSCC)*, feb 2018, pp. 42–44.

[9] C. Duran *et al.*, "A system-on-chip platform for the internet of things featuring a 32-bit RISC-V based microcontroller," in *Latin American Symposium on Circuits & Systems (LASCAS)*, feb 2017, pp. 1–4.

TABLE II
Summary and State of the Art Comparison

| | This work | ISSCC'18 [8] | LASCAS'17 [9] | JSSC'17 [4] | ESSCIRC'17 [3] | ISSCC'15 [2] |
|---|---|---|---|---|---|---|
| Technology | 28 nm FDSOI | 65 nm CMOS | 130 nm CMOS | 40 nm CMOS | 28 nm FDSOI | 65nm CMOS |
| Architecture | *RV32IM* | *RV32I* | *RV32IM* | Cortex-M0 | Cortex-M0+ | Cortex-M0+ |
| Memory | 2 × 32 kB | 16 kB + 64 kB | 4 kB | 64 kB | 2 × 4 kB | 8 kB + 16 kB |
| F @$V_{MEP}$ | 21.5 MHz @0.38 V | 20 MHz @0.8 V* | 160 MHz @1.2 V* | 13.7 MHz @0.37 V | 16 MHz @0.5 V | 0.75 MHz @0.35 V |
| F @$V_{MIN}$ | 1 MHz @0.25 V | - | - | 0.8 MHz @0.2 V | - | 29 kHz @0.19 V |
| $E_{core}$ @ $V_{MEP}$ | **4.18 pJ/cycle** | - | - | 8.80 pJ/cycle | 0.94 pJ/cycle | - |
| $E_{tot}$ @ $V_{MEP}$ | **8.81 pJ/cycle** | 40.4 pJ/cycle* | 167 pJ/cycle* | 43.22 pJ/cycle | 2.67 pJ/cycle | 11.7 pJ/cycle |
| Minimal $EDP_{tot}$ | **0.075 pJ/cycle/MHz** | 2.02 pJ/cycle/MHz* | 1.04 pJ/cycle/MHz* | 3.2 pJ/cycle/MHz | 0.17 pJ/cycle/MHz | 1 pJ/cycle/MHz** |
| $F_{range}$ @ $E_{core}$<10 pJ/cycle | **2-200 MHz** | - | - | 6-35 MHz | 16 MHz | - |

* Based on lowest reported µW/MHz. ** Estimated from figure.