POLITECNICO DI TORINO
Repository ISTITUZIONALE

PolHEEPo: A 65 nm 6mm2 RISC-V SoC Filling the Gap Between Cloud and Edge Computing

Original
PolHEEPo: A 65 nm 6mm2 RISC-V SoC Filling the Gap Between Cloud and Edge Computing / Giuffrida, L., Dolmeta, A., Naclerio, A., Piscopo, V., Mirigaldi, M., Manfredi, E., Varaldi, A., Petrolo, V., Guella, F., Vacca, M., Masera, G., Martina, M.. - In: IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS. II, EXPRESS BRIEFS. - ISSN 1549-7747. - (2026), pp. 1-1. [10.1109/TCSII.2026.3735193]

Availability:
This version is available at: 11583/3015727 since: 2026-09-20T11:48:07Z

Publisher:
IEEE

Published
DOI:10.1109/TCSII.2026.3735193

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

12 November 2026

PolHEEPo: A 65 nm 6 mm^2 RISC-V SoC Filling the Gap Between Cloud and Edge Computing

Luigi Giuffrida, Alessandra Dolmeta, Alessio Naclerio, Valeria Piscopo, Mattia Mirigaldi, Enrico Manfredi, Alessandro Varaldi, Vincenzo Petrolo, Flavia Guella, Marco Vacca, Guido Masera, and Maurizio Martina

***Abstract*—This paper presents PolHEEPo, a heterogeneous RISC-V System-on-Chip (SoC) jointly supporting machine-learning (ML) inference and on-node protection of telemetry payloads in battery-powered edge devices. Fabricated in 65 nm TSMC LP technology, the 6 mm^2 die extends X-HEEP with memory-mapped accelerators for ML, authenticated encryption, SHA-3 hashing, and PUF-based hardware identity. These include Rogue, a streaming 4×4 Coarse-Grain Reconfigurable Array (CGRA) for GEMM and nonlinear activations; Ascon and Keccak engines; and PUFfo, a ring-oscillator Physical Unclonable Function providing a hardware identity primitive. Operating from 0.8 V to 1.2 V at up to 120 MHz, PolHEEPo is evaluated on industrial predictive-maintenance and wearable cardiac-monitoring workloads combining on-device inference with a hardware-accelerated hashing and authenticated encryption pipeline. Silicon measurements show that execution using accelerators lowers the energy of the offloaded stages by 69% and 86%, while complete-window energy decreases from 2.81 mJ to 2.53 mJ and from 5.51 mJ to 2.80 mJ, respectively. For the cardiac monitor, accelerator offloading approximately halves both application-level energy and latency, yielding a projected battery autonomy of 6.5 days. To our knowledge, among the published silicon platforms considered, PolHEEPo is distinguished by jointly integrating on-device ML acceleration, lightweight authenticated encryption, a PQC-relevant Keccak/SHA-3 primitive, and a hardware identity primitive within a RISC-V SoC.**

***Index Terms*—RISC-V, Edge Computing, Coarse-Grain Reconfigurable Array, Cryptography, Keccak/SHA-3 primitive, Physical Unclonable Function, ML Inference, Secure IoT Node.**

I. Introduction

The proliferation of IoT devices, wearables, and industrial sensors is driving a shift toward a cloud-edge continuum, in which latency-sensitive and privacy-critical workloads are processed close to the sensor, and only protected summaries are forwarded to the cloud. This model imposes two simultaneous requirements on IoT nodes. First, sufficient compute capability for on-device inference is needed to prevent raw sensor streams from being exposed over the air. Second, cryptographic protection must secure telemetry over device lifetimes that, in industrial and medical deployments, routinely span 15–20 years. Such longevity requires dedicated, ultra-low-power accelerators providing continuous security without exhausting constrained energy budgets. Moreover, to counter forward-looking threats such as "harvest now, decrypt later", these accelerators must combine lightweight cryptography with PQC-relevant Keccak/SHA-3 primitive.

All authors are with the Department of Electronics and Telecommunications (DET), Politecnico di Torino, 10129 Torino, Italy (e-mail: name.surname@polito.it).

Existing edge SoCs typically address only one of these demands. Platforms such as Vega [1], TinyVers [2], and the broader PULP family deliver multi-GOPS ML inference within milliwatt power envelopes but lack a cryptographic subsystem. On the other hand, recent RISC-V security processors, such as the chip of Zhu *et al.* [3] and the standalone Ascon ASIC [4], provide cryptographic primitives but lack the heterogeneous compute required for local ML. To our knowledge, no published SoC silicon jointly integrates on-device ML acceleration, lightweight authenticated encryption, Keccak/SHA-3 hashing, also used in standardized Post-Quantum Cryptography (PQC) schemes, and a hardware identity primitive.

Figure 1: PolHEEPo 6 mm^2 prototype shot.

This paper presents PolHEEPo, a RISC-V SoC designed to address this gap. Building on the X-HEEP platform [5], it integrates *Rogue*, a streaming 4×4 Coarse-Grain Reconfigurable Array (CGRA) tightly coupled to the DMA subsystem for GEMM and non-linear activation approximation; *Ascon* and *Keccak* engines for lightweight authenticated encryption and hashing; and *PUFfo*, a ring-oscillator PUF integrated as a hardware identity primitive. We motivate this combination through representative secure edge scenarios spanning industrial maintenance and medical monitoring, namely a predictive-maintenance acoustic monitor and a wearable ECG monitor, where on-device inference and hardware-accelerated payload protection must coexist within battery-powered energy budgets. We report silicon power measurements for both use cases and compare PolHEEPo against state-of-the-art edge SoCs. Our evaluation focuses on the device-side portion of the sensing-to-cloud pipeline. The cloud-side processing, key-establishment and provisioning protocols, and wireless transport are therefore not part of the measured system. For the

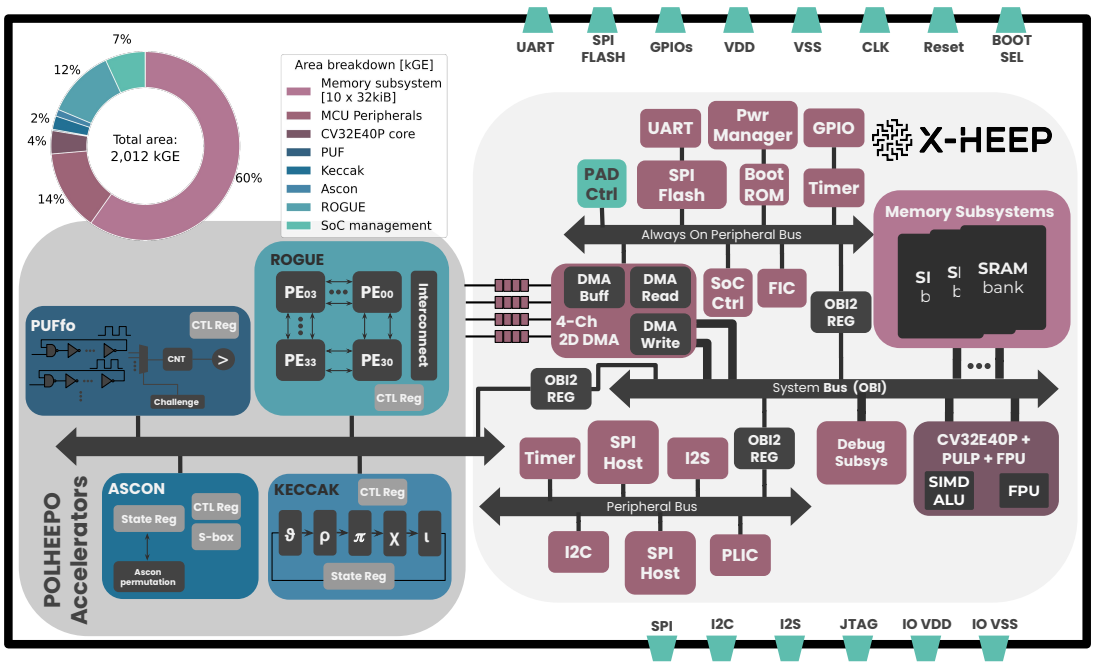

Figure 2: PolHEEPo system overview and post-synthesis area breakdown. Standard X-HEEP components are highlighted in **pink**, while custom-integrated accelerators are shown in **blue**, using a consistent color code across both views.

autonomy analysis, the transmission energy of a representative Bluetooth Low Energy (BLE) link is included as a first-order contribution to the overall energy budget.

II. System Overview

The PolHEEPo chip is a $6\,\mathrm{mm}^2$ silicon implementation of the X-HEEP platform [5] extended with several heterogeneous peripherals, namely *Rogue*, *Keccak* [6], *Ascon* and *PUFfo*, all designed in house. PolHEEPo was taped out in $65\,\mathrm{nm}$ low-power TSMC technology (Figure 1). Figure 2 summarizes the implemented system. All accelerators are available as loosely coupled memory-mapped peripherals on the system bus. PolHEEPo is designed as a heterogeneous platform targeting full-stack applications, integrating sensing, on-device inference, and payload protection within a single node, before data leave the device through a standard communication interface. Figure 2 reports the system area breakdown: the $320\,\mathrm{KiB}$ on-chip SRAM, organized as ten $32\,\mathrm{KiB}$ sub banks, dominates at 60% of the die, while the four accelerators together occupy 15% and the CV32E40P RISC-V core [7] 4%.

***Rogue**:* *Rogue* is an open-source streaming CGRA integrated within PolHEEPo's DMA subsystem to accelerate GEMM workloads and integer approximations of non-linear activation functions. It leverages a compact set of reconfigurable operators capturing recurring computational patterns across integer approximations and GEMM operations. The *Rogue* instance in PolHEEPo features a 4×4 heterogeneous processing element array fed through FIFO-based input and output interfaces by the four-channel 2D DMA of the microcontroller unit. By directly interfacing with the system DMA, *Rogue* processes data fetched from PolHEEPo's main memory or SPI interface on-the-fly, without dedicated load-store units or bus master ports, keeping bus pressure minimal.

***Keccak**:* The *Keccak* accelerator implements the Keccak-f[1600] permutation underlying SHA-3 and extendable-output functions [8]. The round-based datapath completes one permutation round per clock cycle, providing a compact high-throughput implementation suitable for PolHEEPo [6]. PolHEEPo does not implement a complete PQC scheme; the Keccak-f[1600] engine instead accelerates SHA-3/SHAKE primitives also used in standardized schemes such as ML-KEM and ML-DSA.

***Ascon**:* The ASCON cipher suite [9] provides lightweight authenticated encryption and hashing based on a 320-bit permutation. In PolHEEPo, ASCON is implemented as a loosely coupled, single-cycle permutation accelerator. This accelerator supports authenticated encryption and decryption, although only uplink encryption is exercised in the evaluated use cases. Moreover, the Sbox is software-configurable to enable on-silicon evaluation of alternative mappings with different side-channel properties [10], [11], [12]. In the current prototype, no hardware write protection is provided and configuration integrity relies on trusted software.

***PUFfo**:* PUFfo is a ring-oscillator-based Physical Unclonable Function (PUF) [15] designed to extract hardware-specific entropy from process variations across replicated oscillator stages. Its architecture comprises 32 13-stage ring oscillators and uses an 8-bit challenge to select and compare their outputs through lightweight multiplexing and counting logic. Over a sampling window, the selected signals are accumulated in two counters, and their relative behavior is folded into a 128-bit response stream by comparison, producing a device-specific fingerprint. In practical deployment the PUF response is not used directly as a cryptographic key: an enrollment phase records a reference response and the auxiliary information needed for error-tolerant reconstruction, so

that at boot PUFfo is re-evaluated, its response reconstructed against the enrolled reference, and the result passed through a Key-Derivation Function (KDF) to obtain the device-bound key. This prototype implements only the response-generation hardware; enrollment, reconstruction, and KDF are not characterized, although the on-chip Keccak engine already provides the primitive required for a hardware-assisted KDF. A full characterization of PUF reliability, inter-device uniqueness, and entropy under environmental variations lies outside this silicon evaluation and is left for future work.

III. Use cases

PolHEEPo is evaluated through two representative applications in the industrial and biomedical domains. Both share a common structure: the host acquires data, offloads selected processing stages to *Rogue*, and secures the result through a hash-then-authenticated-encrypt pipeline in which SHA3-256 computes the payload digest and ASCON-128 provides authenticated encryption before transmission. The two workloads reproduce the computational, memory, and data-volume characteristics of their target domains: network topology, arithmetic precision, memory footprint, and per-window data volumes are representative, while the evaluation targets platform-level execution rather than model accuracy. The weights are untrained and the anomaly thresholds serve only to exercise the conditional transmission path. For a given input, the CPU-only and accelerator-offloaded configurations produce the same anomaly score, transmission decision, and protected payload; offloading changes only the execution mapping, isolating its latency and energy impact.

Secure Machine Monitoring (Case 1). A predictive-maintenance edge pipeline for autonomous, low-power deployment. A $1\,\mathrm{s}$ audio burst from an I2S MEMS microphone at $16\,\mathrm{kHz}$ is reduced to a 64×32 INT8 feature map by a Mel-Frequency Energy Coefficient (MFEC) front-end and scored by a Convolutional Autoencoder (CAE). Scores above threshold trigger secure transmission of the entire 2048-byte raw feature map together with the score, preserving complete telemetry without exposing raw audio.

GRU-based ECG Processing (Case 2). A wearable cardiac monitor detecting arrhythmia over eight ECG leads sampled at $256\,\mathrm{Hz}$. Each $4\,\mathrm{s}$ window is processed by a diagonal Gated Recurrent Unit (GRU) over 1024 timesteps, evaluating two sigmoid gates and one tanh candidate per hidden unit at each step; the final hidden state is reduced to a scalar anomaly score. Every window is packaged into a compact report — hidden-state vector plus metadata, with a header flagging threshold crossings — and secured on-device before transmission.

IV. Results and discussion

PolHEEPo was characterized using a custom carrier board and an open-source test board [16]. Figure 3 reports the voltage-frequency operating space for supply voltages from $0.8\,\mathrm{V}$ to $1.2\,\mathrm{V}$ under a representative arithmetic workload, confirming the monotonic relationship between supply voltage and achievable frequency: the chip operates reliably up to $120\,\mathrm{MHz}$ at $1.2\,\mathrm{V}$, while lower voltages trade frequency for power, enabling the performance-energy tradeoffs required by the battery-powered scenarios of Section III. The $120\,\mathrm{MHz}$ ceiling is a measurement-setup limitation rather than a silicon one; post-layout timing closure supports higher frequencies.

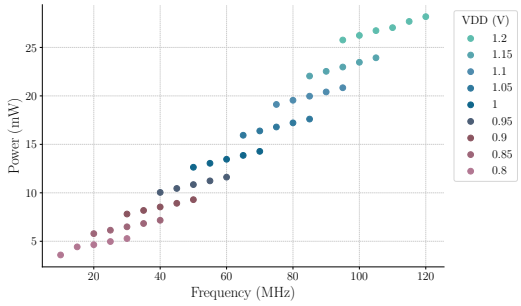

Figure 3: Total chip power consumption vs. operating frequency at supply voltages from $0.8\,\mathrm{V}$ to $1.2\,\mathrm{V}$ under a representative arithmetic workload.

Both applications were evaluated at $100\,\mathrm{MHz}$ and $V_{\mathrm{DD}} = 1.2\,\mathrm{V}$, comparing CPU-only execution against accelerator offloading. Table I provides the cross-platform comparison, while Table II reports the absolute per-stage latency and energy of both applications in the two configurations, together with the share of the CPU-only baseline attributable to each stage. Two reference bases are used in the following and are deliberately kept distinct. The CPU-only configuration is a same-die ablation rather than a comparison against prior art: it isolates the accelerator contribution while holding technology

Table I: Architectural and quantitative comparison of representative fabricated ML-capable edge SoCs and cryptographic implementations."✗" denotes the absence of the corresponding capability.

Parameter	TinyVers[2]	Vega[1]	SamurAI[13]	[4]	Yun[14]	PolHEEPo
Architecture and implementation						
Tech. (nm)	22	22	28	180	65	65
Area (mm^2)	6.25	12	4.5	1.13	6	6
Supply (V)	0.4–0.9	0.5–0.8	0.45–0.9	1.0–2.0	0.85–1.5	0.8–1.2
Freq. (MHz)	≤150	≤450	≤350	≤40	≤400	≤120
Memory (KiB)	132 L1 + 576 L2 + 512 MRAM	128 L1 + 1600 L2 + 4096 MRAM	464	4	16 VRF + 64 L2	320
CPU	RI5CY RV32	10×RI5CY RV32	RI5CY RV32	SERV RV32	CVA6 RV64GC	RI5CY RV32
ML IP	FlexML 8×8 Spatial Array	HW Conv. Engine	PNeuro 64 PEs	✗	✗	Rogue CGRA 16 PEs
Crypto IPs	✗	✗	AES, PRESENT TRIVIUM	ASCON	✗	Keccak, ASCON PUF

node, memory hierarchy, compiler, and measurement setup constant. Both configurations were compiled with the same toolchain and optimisation level, operate on INT8 data resident in on-chip SRAM, and include DMA configuration overhead in the reported stage latency. The input windows are preloaded in on-chip SRAM for these measurements; in deployment they would be streamed from the sensor interfaces through the DMA, an acquisition phase that lies outside the measurement boundary of Table II. Energy reductions are reported for the offloaded stages, whereas speedups refer to the complete application. This distinction accounts for the MFEC front-end in Case 1 and GRU linear operations in Case 2, which remain on the CPU and represent 86% and 42–44% of the respective baseline costs.

Using the on-chip accelerators, PolHEEPo achieves a $1.1\times$ speedup for the execution of the Secure Machine Monitor (Case 1) application compared to a CPU-only execution. This is expected because the dominant stage, the MFEC front-end, combines an FFT with bit-reversed strides, which is a pattern outside *Rogue*'s GEMM and activation-approximation operator set and streaming capabilities. SHA3-256 and ASCON-128 execute $7\times$ faster using the dedicated accelerator instead of running the same algorithms on the CPU. Restricted to the offloaded stages, accelerator use lowers energy from $400\,\mu\mathrm{J}$ to $126\,\mu\mathrm{J}$, a 69% reduction; over the complete application, where the non-accelerable MFEC front-end dominates, the same offload yields a 9.8% energy reduction.

At the system level, executing Case 2 with accelerator offloading rather than CPU-only execution yields an approximately $2\times$ speedup, together with an 86% energy reduction on the offloaded stages, from $3.18\,\mathrm{mJ}$ to $446\,\mu\mathrm{J}$; over the complete window, which also comprises the GRU linear operations executed on the core in both configurations, energy decreases from $5.51\,\mathrm{mJ}$ to $2.80\,\mathrm{mJ}$, as *Rogue*'s reconfigurable datapath efficiently executes the recurrent sigmoid/tanh activations that dominate the GRU. The linear operations, by contrast, are not a dense GEMM: the diagonal formulation gathers two input elements per hidden unit through an indirection table, a data-dependent access that the 2D DMA cannot generate, so they remain on the core. For each configuration, we estimate the energy of one processing window as $E_{\mathrm{win}} = \sum_i I_i V_{\mathrm{core}} t_i$, where I_i is the measured current of task i at the $1.2\,\mathrm{V}$ core rail, $V_{\mathrm{core}} = 1.2\,\mathrm{V}$, and t_i is the corresponding task latency obtained from the `mcycle` counter. The complete-window energy and latency of Case 2, reported in the *Complete application* row of Table II and equal to $2.80\,\mathrm{mJ}$ and $113\,\mathrm{ms}$ in the accelerated configuration, are the only measured inputs to the autonomy projection; every other term is an assumption based on literature. Over one period the node is modelled as drawing that measured compute energy, a BLE transmission term of $50\,\mu\mathrm{J}$ per window, and the deep-sleep energy accumulated over the remainder of the period at an assumed $25\,\mu\mathrm{W}$; the resulting average power is then compared against the usable energy of the cell after power-management losses, for which a conversion efficiency of 0.80 is assumed. Using the cell parameters of a representative implantable system [17], namely a $40\,\mathrm{mA\,h}$, $3.6\,\mathrm{V}$ cell, and one processing window every $4\,\mathrm{s}$ as dictated by the application, the projected autonomy increases from $3.4\,\mathrm{d}$ for CPU-only execution to $6.5\,\mathrm{d}$ with accelerator offloading.

This is a derived system-level figure and not a measurement, and two limitations follow from that. First, the projection presumes that the SoC is power-gated between windows. Second, no power-management circuitry is integrated in the prototype, so the conversion efficiency is assumed rather than measured. We therefore swept each assumed quantity one at a time over a deliberately wide range, holding the others at their nominal value: the processing period over $1\,\mathrm{s}$ to $60\,\mathrm{s}$ yields $1.7\,\mathrm{d}$ to $66.2\,\mathrm{d}$, the deep-sleep power over $5\,\mu\mathrm{W}$ to $500\,\mu\mathrm{W}$ yields $6.7\,\mathrm{d}$ to $4.0\,\mathrm{d}$, the power-management efficiency over 0.60–0.95 yields $4.9\,\mathrm{d}$ to $7.7\,\mathrm{d}$, and the BLE energy per window over $10\,\mu\mathrm{J}$–$1\,\mathrm{mJ}$ yields $6.6\,\mathrm{d}$ to $4.9\,\mathrm{d}$. Under the stated assumptions the compute term accounts for 95% of the per-window energy, so the projection is governed by measured rather than assumed quantities; more importantly, the ratio between the two configurations never falls below $1.55\times$ across every sweep and remains at $1.9\times$ wherever the compute term dominates, so the benefit attributed to accelerator offloading

Table II: Per-stage latency and energy of the two applications, measured on silicon at 100 MHz and $V_{\mathrm{DD}} = 1.2\,\mathrm{V}$. Energy is obtained as $E_i = I_i V_{\mathrm{core}} t_i$, with I_i the measured current on the core rail and t_i the stage latency read from the `mcycle` counter. The measurement boundary is the core supply rail and excludes the one-time PUF boot, the DMA-driven acquisition window, and the transmission stage. "Share" is the stage contribution relative to the CPU-only baseline of the corresponding application. Stages marked † execute on the CV32E40P core in both configurations and are therefore excluded from the offloaded-stage subtotal; the residual differences between their columns reflect measurement repeatability and not an effect of offloading.

		Energy				**Latency**			
	Stage	CPU-only (μJ)	Accel. (μJ)	Reduct. (%)	Share (%)	CPU-only (ms)	Accel. (ms)	Speedup ($\times$)	Share (%)
Case 1	MFEC extraction†	2408.6	2406.3	0.1	85.8	98.18	98.09	1.00	86.4
	CAE inference	122.8	88.7	27.8	4.4	4.78	3.73	1.28	4.2
	SHA3-256	170.8	23.3	86.4	6.1	6.55	0.91	7.16	5.8
	ASCON-128	106.2	13.5	87.3	3.8	4.17	0.57	7.30	3.7
	Offloaded stages	399.8	125.5	**68.6**	14.2	15.49	5.22	2.97	13.6
	Complete application	2808.4	2531.8	9.8	100.0	113.67	103.31	**1.10**	100.0
Case 2	GRU linear operations†	2330.7	2355.0	−1.0	42.3	96.00	97.00	0.99	44.1
	GRU activations	3158.7	442.6	86.0	57.3	121.00	16.00	7.56	55.5
	SHA3-256	13.0	2.6	80.0	0.24	0.45	0.09	4.84	0.21
	ASCON-128	11.0	1.2	89.3	0.20	0.39	0.04	8.80	0.18
	Offloaded stages	3182.7	446.4	**86.0**	57.7	121.84	16.14	7.55	55.9
	Complete application	5513.4	2801.4	49.2	100.0	217.84	113.14	**1.93**	100.0

does not rest on any of these assumptions. This result indicates that PolHEEPo is a viable platform for untethered wearable operation, subject to the availability of the assumed deep-sleep state. Table I positions PolHEEPo against representative edge System-on-Chip (SoC) platforms and standalone cryptographic implementations, restricted to silicon-characterized designs targeting edge deployment. Among the ML-capable platforms, TinyVers [2], Vega [1], and SamurAI [13] exploit $22\,\mathrm{nm}$ and $28\,\mathrm{nm}$ processes for higher peak frequencies and larger memory hierarchies. TinyVers and Vega integrate larger ML engines than *Rogue*'s 4×4 array—an 8×8 spatial array and a cluster coupled to a convolution engine, respectively—but neither provides a cryptographic subsystem. Conversely, the standalone ASCON chip [4] is special-purpose encryption acceleration without ML compute, while Yun [14] provides a general-purpose vector unit. SamurAI is therefore the closest in scope, the only one pairing a 64-PE ML engine with cryptographic acceleration, and additionally integrating an asynchronous event-driven wake-up subsystem that PolHEEPo does not implement. The two differ mainly in cryptographic scope: SamurAI's AES, PRESENT, and TRIVIUM engines offer symmetric confidentiality but no authentication, hashing, or hardware identity, leaving payload integrity to software. PolHEEPo instead targets the scope required over multi-decade deployments, with *Ascon* for single-pass authenticated encryption, *Keccak* for the SHA-3/SHAKE primitives underlying the payload digest and standardized PQC schemes, and *PUFfo* for a device-bound identity derived from on-die entropy rather than provisioned secrets. Within the platforms of Table I, PolHEEPo is thus distinguished by jointly integrating an ML accelerator (*Rogue*), lightweight authenticated encryption (*Ascon*), a PQC-relevant Keccak/SHA-3 primitive (*Keccak*), and a PUF-based hardware identity primitive (*PUFfo*) on a single die, at a competitive area cost: $6\,\mathrm{mm}^2$ matches Yun and is smaller than Vega despite the broader accelerator suite. The lower peak frequency relative to $22\,\mathrm{nm}$ chips is expected at $65\,\mathrm{nm}$. A silicon-level, iso-workload comparison is not attainable, as none of these platforms is available to us as a measurable device; we therefore confine our measured claims to PolHEEPo itself.

V. Conclusions

This work presented PolHEEPo, a $6\,\mathrm{mm}^2$ 65-nm RISC-V SoC combining on-device ML acceleration with Ascon, Keccak/SHA-3, and a PUF-based hardware identity primitive. Silicon measurements on two application pipelines show that inference and payload protection can coexist within a battery-powered energy budget, with accelerator offloading halving the energy and latency of the cardiac monitor. Future work will extend end-to-end, PUF, and side-channel characterization.

Acknowledgments

This work was funded by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU. The authors acknowledge the support they received from the Embedded Systems Lab (ESL) at EPFL, the CHIPS-IT foundation and the Italian Institute of Nuclear Physics (INFN).

References

[1] D. Rossi, F. Conti, M. Eggiman, A. D. Mauro, G. Tagliavini, S. Mach, M. Guermandi, A. Pullini, I. Loi, J. Chen, E. Flamand, and L. Benini, "Vega: A ten-core soc for iot endnodes with dnn acceleration and cognitive wake-up from mram-based state-retentive sleep mode," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 1, pp. 127–139, 2022.

[2] V. Jain and e. A. Giraldo, Sebastian, "Tinyvers: A tiny versatile system-on-chip with state-retentive emram for ml inference at the extreme edge," *IEEE Journal of Solid-State Circuits*, vol. 58, no. 8, pp. 2360–2371, 2023.

[3] Y. Zhu, W. Zhu, Y. Ouyang, J. Sun, M. Zhu, Q. Zhao, J. Yang, C. Chen, Q. Tao, G. Yang, A. Zhang, S. Wei, and L. Liu, "16.2 a 28nm 69.4kops 4.4µj/op versatile post-quantum crypto-processor across multiple mathematical problems," in *2024 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 67, 2024, pp. 298–300.

[4] K.-D. Nguyen, T.-K. Dang, B. Kieu-Do-Nguyen, D.-H. Le, C.-K. Pham, and T.-T. Hoang, "Asic implementation of ascon lightweight cryptography for iot applications," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 72, no. 1, pp. 278–282, 2025.

[5] S. Machetti, P. D. Schiavone, T. C. Müller, M. Peón-Quirós, and D. Atienza, "X-HEEP: An Open-Source, Configurable and Extendible RISC-V Microcontroller for the Exploration of Ultra-Low-Power Edge Accelerators," Jan. 2024. [Online]. Available: https://arxiv.org/abs/2401.05548v2

[6] V. Piscopo, A. Dolmeta, M. Mirigaldi, M. Martina, and G. Masera, "A Deep Dive into Integration Methodologies in RISC-V," in *Proceedings of the 22nd ACM International Conference on Computing Frontiers: Workshops and Special Sessions*, ser. CF '25 Companion. New York, NY, USA: Association for Computing Machinery, 2025, p. 30–33. [Online]. Available: https://doi.org/10.1145/3706594.3726969

[7] M. Gautschi, P. D. Schiavone, A. Traber, I. Loi, A. Pullini, D. Rossi, E. Flamand, F. Gurkaynak, and L. Benini, "Near-Threshold RISC-V Core With DSP Extensions for Scalable IoT Endpoint Devices," Feb. 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7864441

[8] G. Bertoni, J. Daemen, M. Peeters, G. Van Assche, and R. Van Keer, "FIPS PUB 202 - SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions," 2015. [Online]. Available: https://keccak.team/hardware.html

[9] C. Dobraunig, M. Eichlseder, F. Mendel, and M. Schläffer, "Ascon v1. 2: Lightweight authenticated encryption and hashing," *Journal of Cryptology*, vol. 34, no. 3, p. 33, 2021.

[10] E. Oswald, S. Mangard, C. Herbst, and S. Tillich, "Practical second-order dpa attacks for masked smart card implementations of block ciphers," in *Cryptographers' Track at the RSA Conference*. Springer, 2006, pp. 192–207.

[11] C. Carlet, "On highly nonlinear s-boxes and their inability to thwart dpa attacks," in *International Conference on Cryptology in India*. Springer, 2005, pp. 49–62.

[12] C. Carlet, A. Heuser, and S. Picek, "Trade-offs for s-boxes: Cryptographic properties and side-channel resilience," in *International conference on applied cryptography and network security*. Springer, 2017, pp. 393–414.

[13] I. Miro-Panades and et Al., "Samurai: A 1.7mops-36gops adaptive versatile iot node with 15,000× peak-to-idle power reduction, 207ns wake-up time and 1.3tops/w ml efficiency," in *2020 IEEE Symposium on VLSI Circuits*, 2020, pp. 1–2.

[14] M. Perotti, M. Cavalcante, A. Ottaviano, J. Liu, and L. Benini, "Yun: An open-source, 64-bit risc-v-based vector processor with multi-precision integer and floating-point support in 65-nm cmos," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 70, no. 10, pp. 3732–3736, 2023.

[15] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *2007 44th ACM/IEEE Design Automation Conference*, 2007, pp. 9–14.

[16] "esl-epfl/CHEEP-boards," May 2026, original-date: 2025-12-22T12:07:19Z. [Online]. Available: https://github.com/esl-epfl/CHEEP-boards

[17] F. D. Bono, N. D. Trani, A. Joubert, C. Caffey, A. Dentis, D. Demarchi, A. Grattoni, and P. Motto Ros, "In Vivo Wireless Powering of a Long-Acting Nanofluidic Drug Delivery Implant," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 20, no. 3, pp. 493–508, Jun. 2026. [Online]. Available: https://ieeexplore.ieee.org/document/11360751/