Shaheen: An Open, Secure, and Scalable RV64 SoC for Autonomous Nano-UAVs

L. Valente*, A. Veeran¶, M. Sinigaglia*, Y. Tortorella*, A. Nadalini*, N. Wistoff†, B. Sá‡, A. Garofalo*, R. Psiakis**, M. Tolba¶, A. Kulmala**, N. Limaye, O. Sinanoglu§, S. Pinto‡, D. Palossi†‖, L. Benini*†, B. Mohammad¶, D. Rossi*
* University of Bologna, Italy, ¶ Khalifa University, UAE, † ETH Zürich, Switzerland, ‡ Universidade do Minho, Portugal,
** Technology Innovation Institute, UAE § New York University Abu Dhabi, UAE, ‖ IDSIA USI-SUPSI, Switzerland

I. Introduction

Palm-sized unmanned aerial vehicles (nano-UAVs) improve safety and reduce costs in critical applications such as search&rescue and human-robot interaction [1] [2]. As nano-UAVs evolve from research prototypes to products, their flight computers need to (i) run increasingly complex multi-tasking workloads with large memory footprints within a few hundred mW power budget; (ii) support virtualization and secure operation in uncontrolled or even hostile deployment scenarios.

We present a low-power heterogeneous System-on-Chip (SoC) architecture implemented in 22 nm FDX technology that tackles both challenges by coupling a 64-bit RISC-V *Host* processor featuring hardware virtualization support with an 8-core Programmable Multi-Core Accelerator (PMCA). The PMCA's cores extend the RISC-V Instruction Set Architecture (ISA) with Digital Signal Processing (DSP) and Deep Learning (DL) operations. The PMCA achieves up to 31 GOp/s and 125 GOp/s on 8-bit and 2-bit integer kernels, respectively. Its peak energy efficiency of 2.3 TOp/s/W enables the efficient execution of DL-based workloads such as human pose estimation [2].

The Host supports full virtualization: to the best of our knowledge, it is the first silicon implementation fully compliant with the ratified RISC-V ISA hypervisor extension. Furthermore, it features ISA and micro-architecture extensions for timing channel mitigation to isolate concurrent execution of multiple software stacks (trusted and untrusted), preventing security threats and ensuring multi-domain operations. The Shaheen SoC, thanks to its open, secure, and scalable architecture with Host/PMCA decoupling and advanced virtualization, enables the execution of the full vertical stack of applications needed for the current and future generation of autonomous nano-UAVs.

II. Heterogeneous Secure System

Architecture: The Host processor is the CVA6 [3] core, a 6-stage, single-issue, in-order, 64-bit RISC-V core, supporting the RV64GC ISA, 32 kB data and 16 kB instruction cache, virtual memory, and Physical Memory Protection (PMP). The core also features hardware support for virtualization compliant with the RISC-V hypervisor extension [4] guaranteeing temporal and spatial isolation when running heterogeneous software stacks while minimizing size, weight, power, and cost (SWaP-C) [5]. At the same time, a wide set of microarchitectural enhancements [4] reduces the virtualization's overhead, leading to higher performances. Additionally, it allows for the co-existence of full-fledged Operating Systems (OSes), such as Linux, with control-oriented Real-Time Operating Systems (RTOSs), easing the portability of legacy robotic software stacks, see Fig. 1.

To improve the security of concurrent software stacks and prevent micro-architectural timing channels (e.g., Spectre) that could enable unwanted information flow across security boundaries, the core implements the temporal fence instruction, namely *fence.t* [6]. This instruction provides the capability of clearing vulnerable microarchitectural states and enables a history-independent context-switch latency (Fig. 8) while offering a low implementation effort, i.e., less than 1% performance impact and negligible hardware costs [6].

Nano-UAVs' State-of-the-Art (SoA) SoCs rely on 32-bit processors (Tab. II), small on-chip ScratchPad Memories (SPM) (e.g., 1 MB), and off-chip memories accessible through peripheral interfaces (e.g., quad-SPI). Instead, the memory hierarchy of the proposed SoC features, in addition to a 1 MiB SPM, an ultra-low-power, low-area, low-pin-count, fully-digital HyperRAM memory controller that can drive up to 512 MB @ 1.6 Gbps directly on the AXI-4 bus. Furthermore, our SOC provides two 8-bit CPI interfaces and a complete set of peripherals (i.e., I2C, (Q)SPI, SDIO, UART, PWM) to enable full compatibility with open standard flight controllers, such as the Pixhawk FMUv5 [7].

The PMCA can execute computationally intensive tasks offloaded by the Host. The PMCA is built around 8 CV32E40P-based cores [8] sharing 16×16 kB SRAM banks, i.e., 256 kB L1SPM. The cores feature dedicated RV32 extension, including hardware loops, post-increment LD/ST, unified MAC&Load operation, narrow bit-width, and mixed-precision SIMD extensions (down to 2-bit) relevant for inference of quantized neural networks [8]. The PMCA's cores also feature FPUs supporting FP32, SIMD FP16, and bfloat extensions relevant for many UAV mission control tasks such as Model-Predictive Control (MPC) and refinement learning [2] (Fig. 7). The cluster can deliver up to 4.2 GFLOp/s when working on 32-bit FP data and up to 125 GOp/s on 2-bit integer data for DL applications (Fig. 3). Offloading a computational workload to the PMCA boosts its performance from 27× on 32-FP (Fig. 5) to 610× on heavily quantized DL workloads, thanks to the parallel execution and the dedicated ISA. A dedicated IOPMP unit filters the PMCA's accesses to the shared memory regions to guarantee memory isolation between coexisting applications. The PMCA is supported by a full software stack featuring a heterogeneous compilation toolchain and bare-metal and Linux-based offloading, leveraging the OpenMP programming model (Fig. 1).

To enhance security in the context of a globalized supply chain and growing threats like intellectual property (IP) piracy, reverse engineering, and hardware Trojans, Shaheen integrates logic locking [9] on two key hardware IPs: the memory controller, and the PMCA, avoiding unauthorized access to these system components. Logic locking, described in Fig 6, consists in modifying a hardware IP to add a new input, specifically a *logic locking key*, to be applied to unlock the original IP functionality. Without the proper *logic locking key* loaded, the chip is non-functional.

Results: Fig. 2 and Tab. I show the chip micrograph and main features, while Tab. II compares Shaheen with the SoA. Shaheen features an unprecedented level of micro-architectural RISC-V security support add-ons, full support for hypervisor-based multi-tenancy, concurrent support for RTOS, and full-blown OS. Our SoC integrates flight control as well as mission control capabilities through a high-performance PMCA, delivering 1.5× better performance than the SoA, all within a tiny form factor of $9mm^2$ and a power envelope of just 200 mW (Fig. 4), fully compatible with the ultra-constrained power envelope of nano-UAV [2].

References

[1] A. Suleiman *et al.*, "Navion: A 2-mw Fully Integrated Real-time Visual-Inertial Odometry Accelerator for Autonomous Navigation of Nano Drones," *IEEE JSSC*, 2019.
[2] E. Cereda *et al.*, "Deep Neural Network Architecture Search for Accurate Visual Pose Estimation aboard Nano-UAVs," *arXiv*, 2023.
[3] OpenHW, "CVA6," *Github repository*, 2023. [Online]. Available: https://github.com/openhwgroup/cva6
[4] B. Sá *et al.*, "CVA6 RISC-V Virtualization: Architecture, Microarchitecture, and Design Space Exploration," *arXiv*, 2023.
[5] B. Sá *et al.*, "A First Look at RISC-V Virtualization From an Embedded Systems Perspective," *IEEE Transactions on Computers*, 2022.
[6] N. Wistoff *et al.*, "Systematic Prevention of On-Core Timing Channels by Full Temporal Partitioning," *IEEE Transactions on Computers*, 2022.
[7] Pixhawk, "PX4," 2023. [Online]. Available: https://docs.px4.io/main/en/
[8] A. Garofalo *et al.*, "XpulpNN: Accelerating Quantized Neural Networks on RISC-V Processors Through ISA Extensions," in *DATE*, 2020.
[9] N. Limaye *et al.*, "Thwarting all logic locking attacks: Dishonest oracle with truly random logic locking," *IEEE TCAD*, 2020.
[10] GreenWavesTechnology, "Ai-Deck, GAP8," 2023. [Online]. Available: https://www.bitcraze.io/products/ai-deck/
[11] D. Rossi *et al.*, "Vega: A Ten-Core SoC for IoT Endnodes With DNN Acceleration and Cognitive Wake-Up From MRAM-Based State-Retentive Sleep Mode," *IEEE JSSC*, 2021.
[12] A. Di Mauro *et al.*, "Kraken: A Direct Event/Frame-Based Multi-sensor Fusion SoC for Ultra-Efficient Visual Processing in Nano-UAVs," in *IEEE Hot Chips 34 Symposium (HCS)*, 2022.

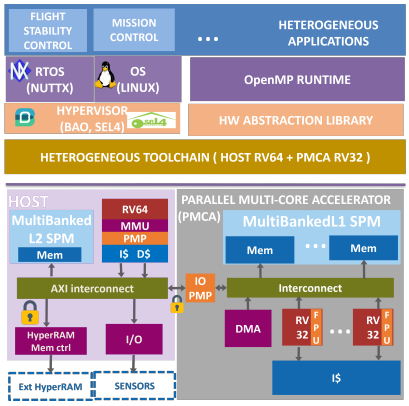


Fig. 1: Shaheen Architecture and Software Stack.

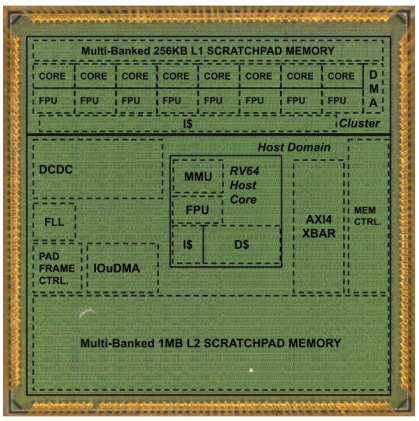


Fig. 2: Die micrograph.

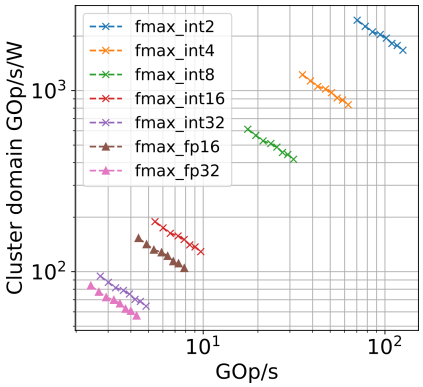


Fig. 3: PMCA energy efficiency on matrix multiplication (2 N-bit OP = 1 N-bit MAC).

L2 mem., L1 mem (SRAM)	1MiB, 256KiB
Off-chip CPU mem. (HyperRAM)	8MiB - 512MiB
VDD Range	0.625-0.8V
Cluster Max Freq. , CVA6 Max. Freq.	500MHz, 600MHz
Power Envelope	200mW

TABLE I: Physical implementation details.

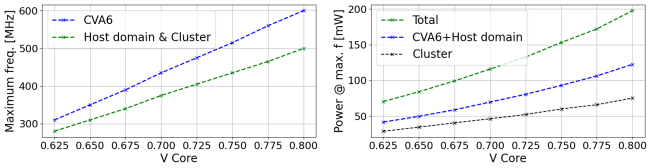


Fig. 4: Frequency and power envelope varying V_{DD}.

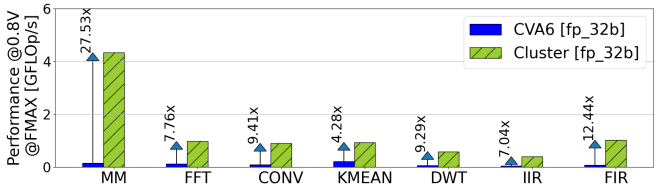


Fig. 5: Cluster speedup over the Host on FP benchmarks, 1MAC=2OPs.



Fig. 6: Logic locking high-level flow.

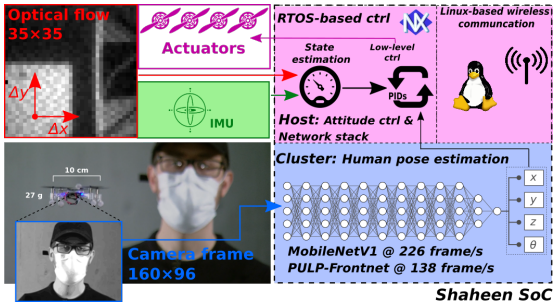


Fig. 7: Application scenario: the Host processor fuses sensor data to implement attitude control and runs Linux-based legacy software such as wireless network stack. The PMCA runs the CNN-based pose estimation task fed by a low-resolution front-looking camera [2].

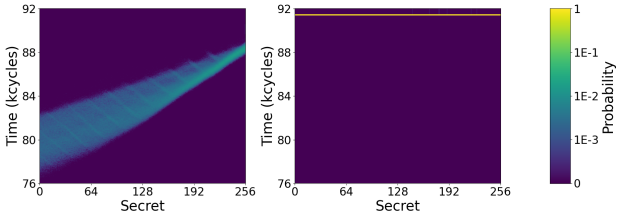


Fig. 8: On the x-axis, the secret number of L1 D$ lines accessed by the Trojan. On the y-axis, the corresponding probe time measured by the spy. On the left, in the *unmitigated* case, the correlation between the Trojan's secret and the spy's probe time indicates a covert channel. On the right, when *fence.t* with *MICRORESET* is used [6], there is no correlation.

	STM32-H7	STM32-F4	GAP8 [10]	Vega [11]	Kraken [12]	Shaheen (this work)
Target board	Pixhawk	Crazyflie	AIDeck	AIDeck	AIDeck	**AIDeck / Pixhawk**
Technology	40nm	90nm	55nm	22nm FDSOI	22nm FDSOI	**22nm FDSOI**
Die Size	-	-	$10mm^2$	$12mm^2$	$9mm^2$	$9mm^2$
CPU	Cortex M7	Cortex M4	9x RI5CY	10x RI5CY-NN	9x RI5CY-XNN	**CVA6 + 8x FLEX-V**
Supported OS	RTOS	RTOS	RTOS	RTOS	RTOS	**Linux/RTOS/ Hypervisor**
Host-compute FP support	SP-FPU, DP-FPU	-	-	SP-FPU	SP-FPU	**SP-FPU, DP-FPU**
Security Features	Crypto/hash accelerators	-	-	-	-	**Side-channel protection, PMP, Logic Locking, IOPMP**
Peak Performance	240MFLOp/s(FP32) 390MOp/s (8b)	72MOp/s (8b)	6 GOp/s (8b)	3.7GFLOp/s(FP32) 7GFLOp/s(FP16) 15.6GOp/s(8b)	3.1GFLOp/s(FP32) 21GOp/s(8b) 85GOp/s(2b)	**4.2GFLOp/s(FP32) 7.9GFLOp/s(FP16) 31GOp/s(8b) 125GOp/s(2b)**

TABLE II: Comparison against commercial and research SoC for nano-drones.