電機資訊學院 2025 腦作 BRAIN PLUS HAND 實作專題競賽

T18 Tapeout – Caravel SoC with Integrated Hardware Accelerator of Falcon

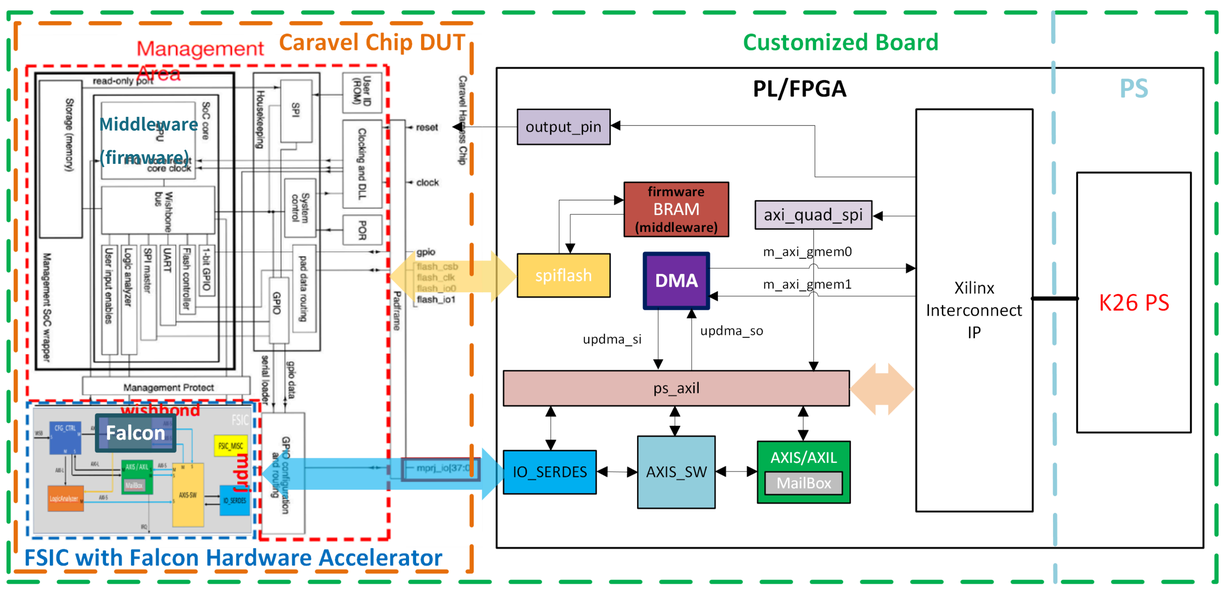
隊伍編號: EECS07 組長: 陳冠晰 組員: 劉祐瑋, 王彥智, 陳昇達, 陳柏翰

Abstract

The rise of quantum computing threatens traditional cryptographic protocols, leading to the development of post-quantum cryptography (PQC). Our project focuses on hardware acceleration of Falcon, a lattice-based digital signature scheme in PQC. To address the challenge of long computation times in software, we accelerate key operations such as Fast Fourier Transform (FFT), inverse FFT (iFFT), Number Theoretic Transform (NTT), and inverse NTT (iNTT). Using High-Level Synthesis (HLS), we streamline hardware design and optimize performance, enhancing efficiency and security.

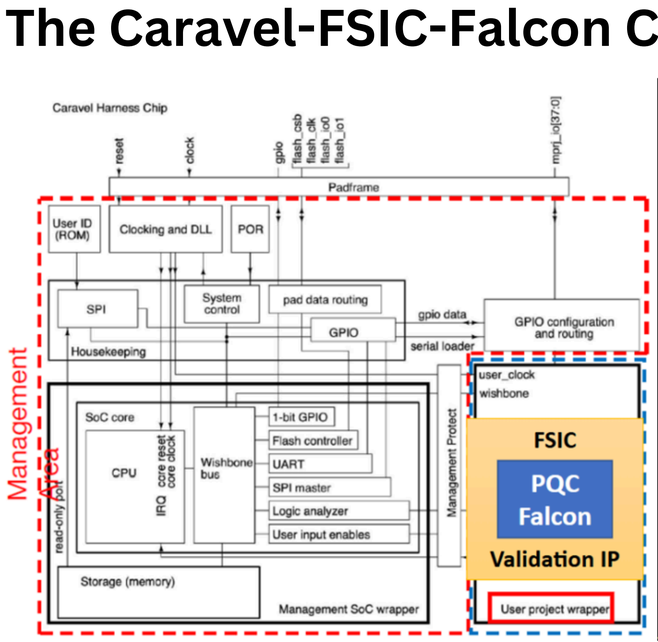
A particularly innovative aspect of our work is the hardware/software co-design approach, which significantly enhances the efficiency of executing the Falcon algorithm on hardware platforms. We collaborated with VIA Technologies to design a custom validation board—the Caravel FSIC platform—to validate the chip after tape out. This dedicated board enables efficient integration and reliable verification of our hardware/software co-design, marking a critical step in our project's development.

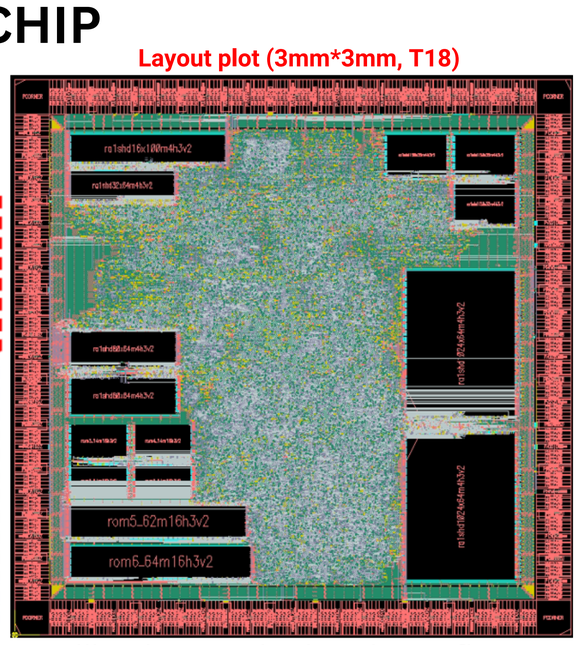
System View



The orange part represents the CHIP DUT, the block to be taped out. It includes the kernel, designed using High-Level Synthesis (HLS), the FSIC as a validation platform IP, and the Caravel SoC itself. The middleware, implemented as firmware, runs on the embedded RISC-V CPU within the Caravel SoC to manage computation scheduling. On the FPGA side, we have included multiple IPs and designed a DMA to handle the data transfer needed for kernel computation.

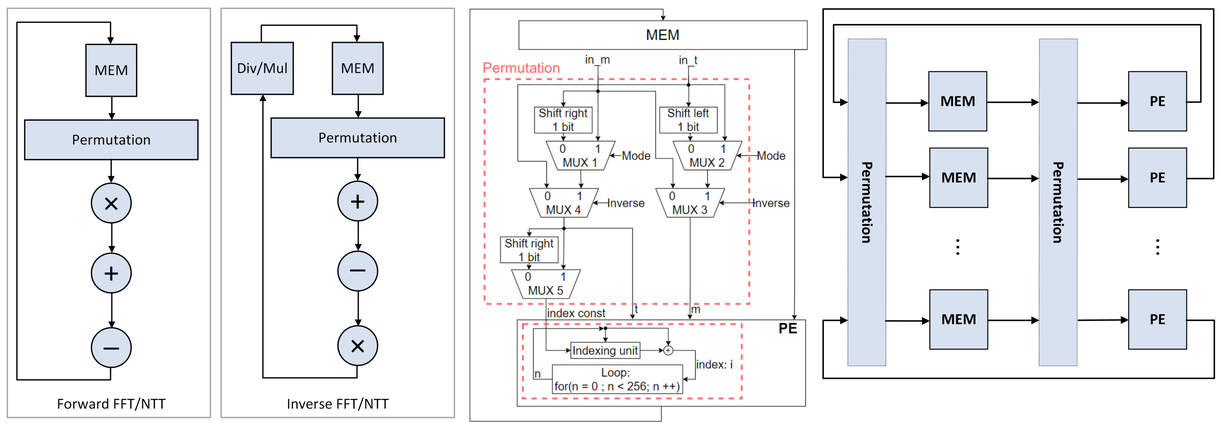
The Caravel-FSIC-Falcon CHIP





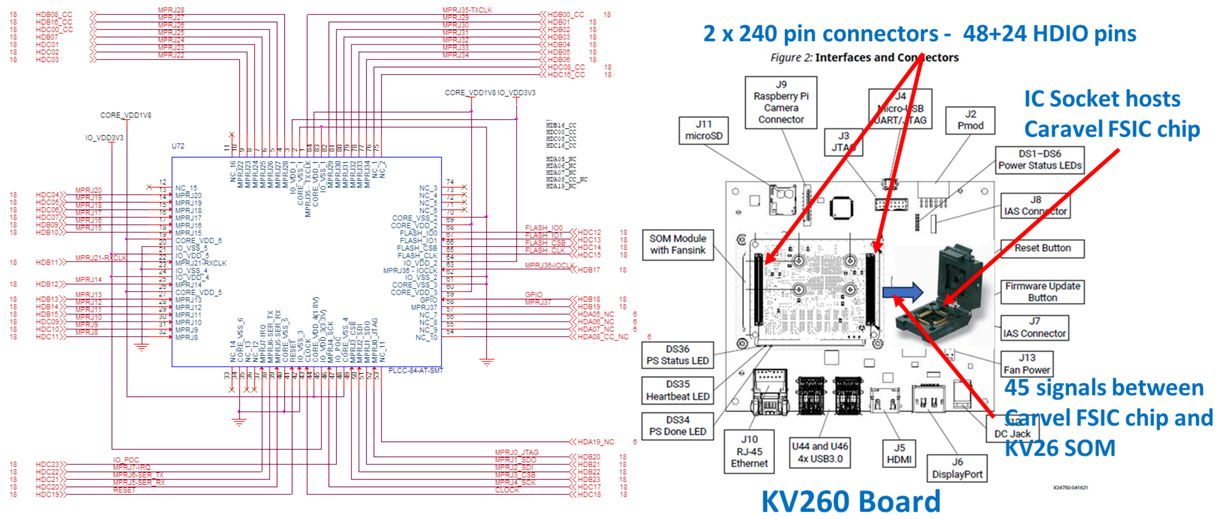
The image above shows the chip we taped out. Unlike the standard Cadence flow, we completed the APR process using Synopsys IC Compiler II on TSRI's EDA Cloud. As far as we know, this is the first endeavor to bring the Synopsys design flow in TSRI

Hardware Acceleration: Kernel Function



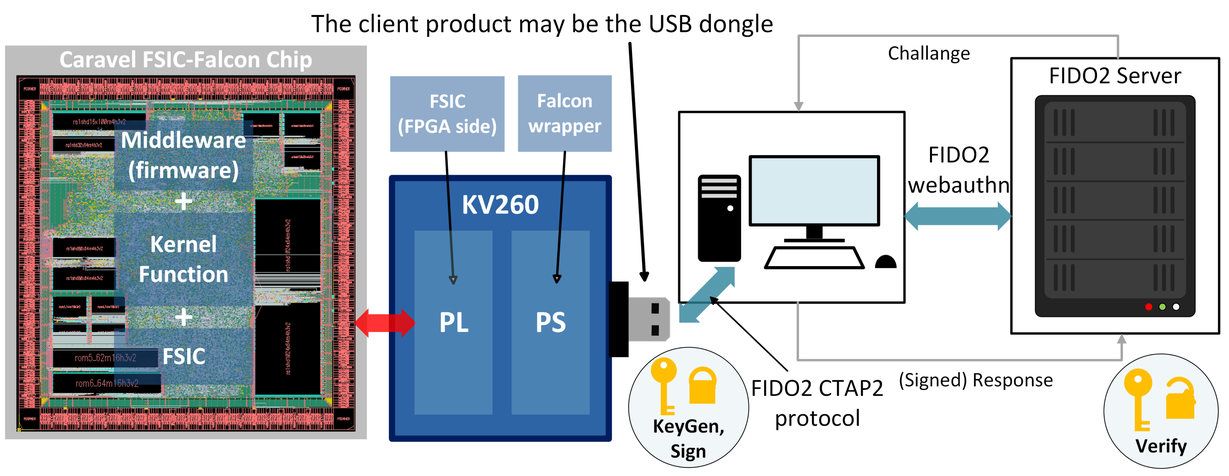
We combined FFT, iFFT, NTT, and iNTT into one hardware IP to minimize hardware resources. Our optimization started by implementing a one-port memory with a custom permutation algorithm. We used a processing element (PE) with an in-place memory buffer (MEM), a double shifter to reduce operator usage, memory sharing with different datatypes for FFT/NTT, and decomposing the complex and Montgomery multiplication with a shared multiplier in the PE to limit hardware resource usage.

System Validation Board



We collaborated with VIA Technologies to design the Caravel-FSIC validation board which added a PLCC socket on it, refer to the AMD KV260 schematic. We will replace the Chip DUT, originally simulated using an FPGA, with the actual chip that comes back after tapeout.

Possible Future Product



In the future, the product could be developed in the form of a USB dongle. By plugging the USB dongle into a PC, KeyGen and Sign operations are performed on the PC side via the FIDO2 CTAP2 protocol, while verification is carried out on the server side. This approach enhances security and simplifies authentication across different platforms.

Performance

Kernel execution time

Function	FFT (ms)	iFFT (ms)	NTT (ms)	iNTT (ms)
Python	28.3617	29.9833	32.8956	34.3557
Original HLS	1.7429	2.2315	3.3797	4.3449
Final optimization	0.1372	0.1631	0.1951	0.1773

Speed Up x200

Falcon execution time

Version	KeyGen (ms) (compute pk)	Sign (ms)	Verify (ms)
Original software	100.4	2053.3	139.9
HW/SW co-design with middleware	18.9	747.5	60.7