Convolutional Neural Network Accelerator for TensorFlow Lite on Embedded FPGA

1st Yarib Nevarez
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

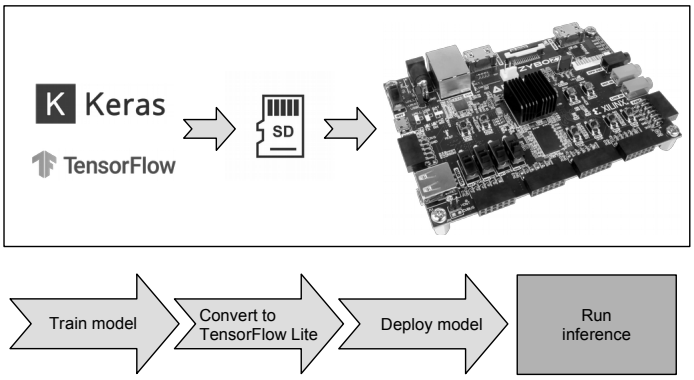

Fig. 1. Workflow.

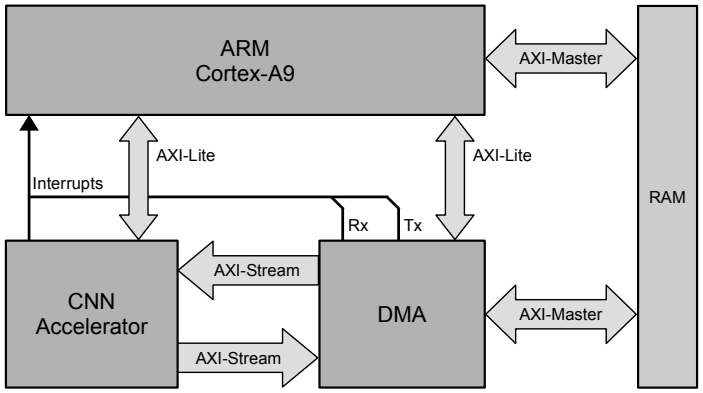

Fig. 2. System-level architecture of the proposed embedded platform.

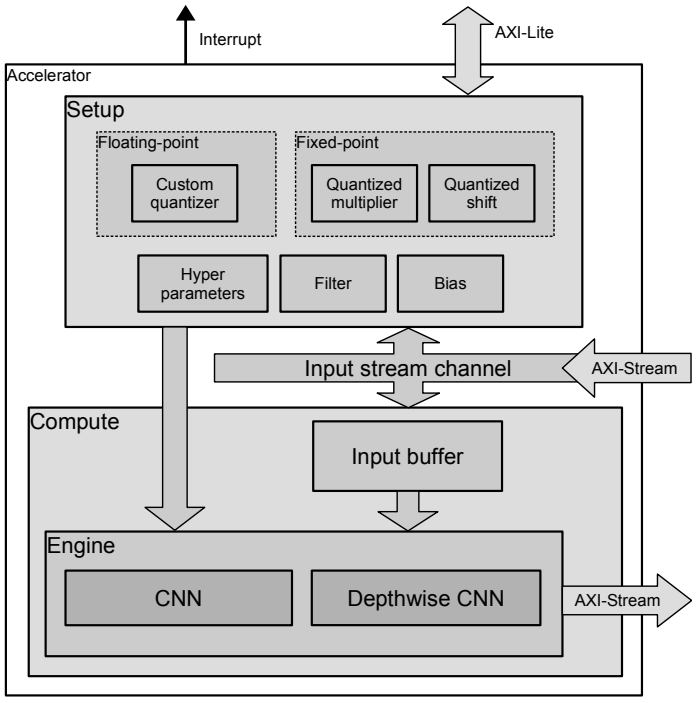

Fig. 3. Hardware architecture of the proposed accelerator.

***Abstract*—Spiking neural networks (SNNs) represent a promising alternative to conventional neural networks. In particular, the so-called Spike-by-Spike (SbS) neural networks provide exceptional noise robustness and reduced complexity. However, deep SbS networks require a memory footprint and a computational cost unsuitable for embedded applications. To address this problem, this work exploits the intrinsic error resilience of neural networks to improve performance and to reduce hardware complexity. More precisely, we design a vector dot-product hardware unit based on approximate computing with configurable quality using hybrid custom floating-point and logarithmic number representation. This approach reduces computational latency, memory footprint, and power dissipation while preserving inference accuracy. To demonstrate our approach, we address a design exploration flow using high-level synthesis and a Xilinx SoC-FPGA. The proposed design reduces $20.5\times$ computational latency and $8\times$ weight memory footprint, with less than 0.5% of accuracy degradation on a handwritten digit recognition task.**

***Index Terms*—Artificial intelligence, spiking neural networks, approximate computing, logarithmic, parameterisable floating-point, optimization, hardware accelerator, embedded systems, FPGA**

Acknowledgments

This work is funded by the *Consejo Nacional de Ciencia y Tecnologia – CONACYT* (the Mexican National Council for

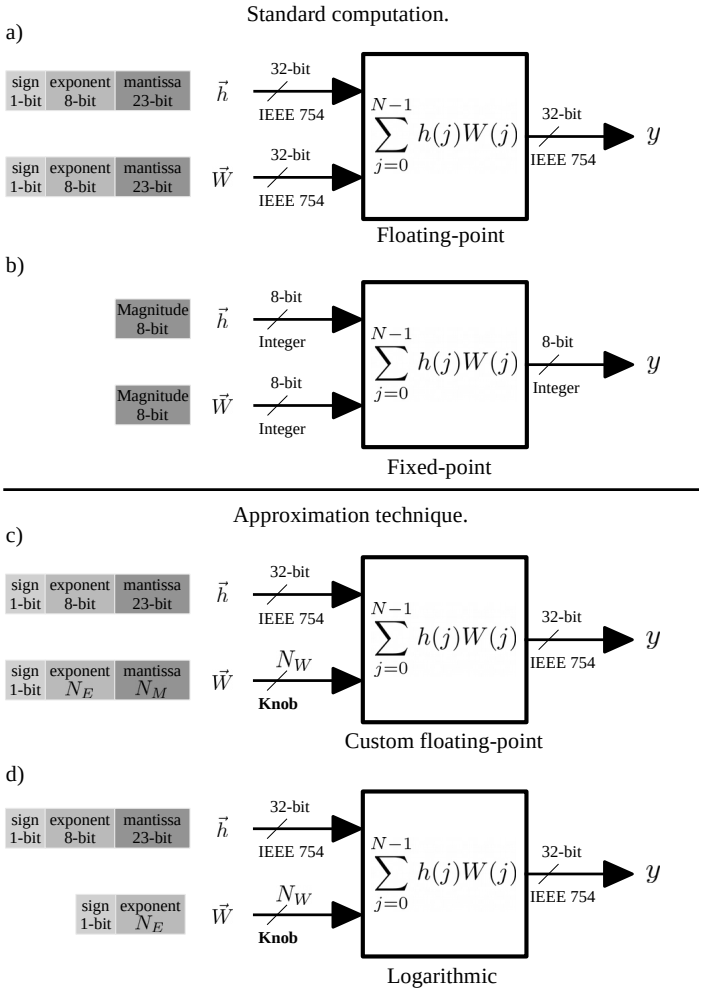

Fig. 4. Proposed hardware modules for vector dot-product.

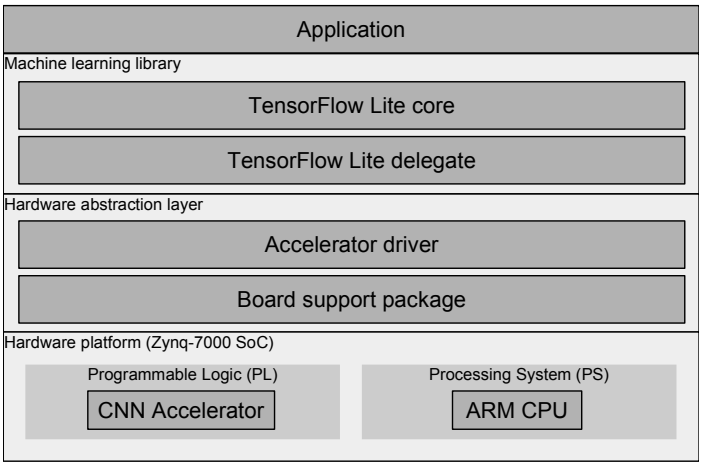

Fig. 5. System-level overview of the embedded software architecture.

Science and Technology).