Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Electrical, Electronics and Communications Engineering
(35^{th}cycle)

Energy-efficient Data-Processing Architectures

Coping with device and circuit-level nonidealities

By

Carmine Paolino

******

Supervisors:
Prof. Gianluca Setti, Supervisor
Prof. Fabio Pareschi, Co-Supervisor

Doctoral Examination Committee:
Prof. David J. Allstot, Referee, Oregon State University
Prof. Mario Lanza, Referee, KAUST
Prof. Fernando Corinto, Politecnico di Torino
Prof. Maurizio Martina, Politecnico di Torino

Politecnico di Torino
2023

Abstract

This thesis focuses on designing and evaluating low-energy systems for signal acquisition and processing. The main theoretical foundation for the work is Compressed Sensing (CS), i.e., a theory that guarantees correct signal recovery despite violating the overarching constraints imposed by Shannon's Sampling Theorem. When a signal can be represented sparsely in a given vector basis, CS enables low-energy encoding and significantly fewer measurements than traditional sampling. The sparsity constraint is commonly satisfied in various fields, ranging from biological signals, such as electrocardiograms and electroencephalograms, to radar pulses and multi-tone RF communication systems. The thesis presents the detailed analysis of two CS encoders with appropriate strategies to address hardware nonidealities, either by ad-hoc circuit design or by appropriate decoding techniques.

The first encoder is a Charge-Redistribution (CR), Successive-Approximation-Register (SAR) Analog-to-Digital converter (ADC). Its architecture redefines the sampling operations of a typical CR SAR ADC by allowing a more granular decomposition of the capacitive array. The independent control of its capacitive elements enables the simultaneous storage of individual samples and the computation of the encoder output through entirely passive means. Therefore, energy consumption is significantly reduced compared to previously proposed architectures. Parasitic capacitances, leakage currents, and matching are some of the issues being addressed, together with effective mitigating strategies. Preliminary experimental evaluations are carried out on a physical implementation designed in UMC's 180 nm technology.

A second encoding platform is built on top of a Phase-Change Memory (PCM) Analog Array through numerical models of its steady-state response, programming variability, and conductance drift. Two techniques to drive the array, namely, voltage-encoded inputs and constant-amplitude time-encoded pulses, result in significantly different concerns from an applicative and design standpoint. Consequently, the

resiliency of well-known CS reconstructors is tested against the observed device variability, and a new, iterative decoding strategy is proposed.

The numerical models are also used as the core computational kernel within PCM-based Neural Network Layers to verify the inherent redundancy and adaptability of Deep Neural Networks (DNNs) as an effective workaround to device nonidealities. Popular classification tasks (MNIST, Fashion-MNIST, CIFAR-10), and a spectral-content estimation task, are all addressed through PCM-based DNNs.

The work highlights the importance of the simultaneous optimization of system, circuit, and device-level parameters for the effectiveness of low-energy hardware implementation, especially for the success of emerging technologies such as Phase-Change Memories.