EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

EDITORIAL Open Access

Quantization of VLSI digital signal processing systems

Gabriel Caffarena[1*], Olivier Sentieys[2], Daniel Menard[2], Juan A López[3] and David Novo[4]

Digital systems have finite precision, which imposes a maximum bound on the accuracy of the results of the computed mathematical operations. The so-called quantization process, also *wordlength* optimization, aims at finding cost-efficient hardware architectures that comply with a given maximum accuracy loss. Floating-point arithmetic is commonly used to perform scientific computations because it provides high dynamic range and mathematical precision.

However, certain applications require the use of dedicated hardware to achieve high computation rates and low power. The computation speed is achieved by means of making use of highly parallel implementations, as well as custom data storage mechanisms (i.e. registers, local memories, etc.). The use of floating-point arithmetic in such systems is prohibitive and it is typically replaced by fixed-point arithmetic, which turns to be more cost-effective, or by restricted floating-point arithmetic. In any case, the designer must face an optimization problem *-quantization-* where the proper precision for each arithmetic operation is searched, resulting in a low-cost hardware implementation that complies with a minimum quality criterion.

Quantization is not an easy task, and in some cases it is oversimplified in order to meet the time-to-market constraints, leading to far from optimal results. However, in some other cases such a simplification is not possible without seriously compromising the viability of the system. As a result, an exhaustive quantization is carried out, implying the extensive use of time-consuming techniques such as computer simulations. Therefore, improvements in quantization error estimation techniques as well as novel methodologies able to handle industrial size systems within a reasonable design time are of crucial importance.

This special issue covers three major areas related to the quantization process: (i) the analysis of the selection of coefficient and signal precision on the design of linear systems, (ii) the efficient implementation and precision analysis of key IP cores for multimedia and communication systems and, (iii) the precision-wise high-level synthesis of DSP algorithms.

The first set of papers focuses on the analysis of the quantization effects of the filter structures.

In the paper "Sensitivity-based Pole and Input-Output Errors of Linear Filters as Indicators of the Implementation Deterioration in Fixed-Point Context", a classical sensitivity analysis for the finite precision implementation of linear filters is extended and improved to consider the exact fixed-point format of the coefficients. Thus, the proposed specialized framework and indicators evaluate and select with improved accuracy the most convenient realization among a wider scope of filter structures for non-uniform quantization of the coefficients.

In "Complexity-Aware Quantization and Lightweight VLSI Implementation of FIR Filters", a complexity-aware quantization framework for FIR filters is presented. It is based on the integration of three optimization techniques: signed-digit coefficient encoding, optimal scaling factor exploration, and common subexpression elimination. The proposed approach saves around 50% of additions, leading to silicon area reductions of up to 34%.

The next three papers deal with the efficient design of fixed-point IP cores.

The paper "Automatic IP Generation of FFT/IFFT Processors with Word-Length Optimization for MIMO-OFDM Systems" presents an accurate precision analysis and a core generator for FFT/IFFT fixed-point cores. The generator makes use of a specific wordlength search algorithm that leads to efficient implementations that comply with recent MIMO-OFDM standards.

In "Novel VLSI Algorithm and Architecture with Good Quantization Properties for a High Throughput

* Correspondence: gabriel.caffarenafernandez@ceu.es
[1]Department of Information and Telecommunication Systems, University San Pablo CEU, Madrid, Spain
Full list of author information is available at the end of the article

Springer

© 2012 Caffarena et al; licensee Springer. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Area Efficient Systolic Array Implementation of DCT", a novel DCT implementation is presented. The proposed approach achieves significant area reductions by means of a new algorithm and architecture that poses good numerical properties that can be efficiently exploited to optimize cost. A precision analysis is provided that highlights the benefit of the approach in comparison to traditional implementations.

A fixed-point model of the maximum a-posteriori (MAP) decoding algorithm of turbo and low-density parity-check (LDPC) codes is presented in "Fixed-Point MAP Decoding of Channel Codes". The analysis is performed considering the turbo and LDPC codes of WiMAX and 3GPP-LTE and it allows identifying the key parameters that affect both precision and implementation cost.

The last paper, "Latency-Sensitive High-Level Synthesis for Multiple Word-Length DSP Design", deals with the High-Level Synthesis of Fixed-Point implementations of DSP algorithms. The proposed technique combines scheduling and binding for multiple wordlength operators with variable area and latency costs. The results yield that the approach leads to latency reductions of up to 19% and area reductions of 9%.

Acknowledgements
The guest editors would like to thank the work of both authors and reviewers. They would also like to thank the Editor in Chief Phillip Regalia and the technical staff of the EURASIP Journal on Advances in Signal Processing.

Author details
[1]Department of Information and Telecommunication Systems, University San Pablo CEU, Madrid, Spain [2]University of Rennes, IRISA/INRIA, Lannion, France [3]Departamento de Ingeniería Electrónica, Universidad Politécnica de Madrid (U.P.M.), Madrid, Spain [4]École Polytechnique Fédérale de Lausanne, Laussane, Switzerland

Received: 14 April 2011 Accepted: 15 February 2012
Published: 15 February 2012

doi:10.1186/1687-6180-2012-32
Cite this article as: Caffarena *et al.*: Quantization of VLSI digital signal processing systems. *EURASIP Journal on Advances in Signal Processing* 2012 2012:32.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Immediate publication on acceptance
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com