Journal of Embedded Computing 2 (2006) 137–139
IOS Press

Guest-editorial

Issues in Embedded Single-Chip Multicore Architectures

Sandro Bartolini and Roberto Giorgi
Dipartimento di Ingegneria dell'Informazione, University of Siena, Italy
E-mail: bartolini@ing.unisi.it; giorgi@unisi.it

Nowadays and future embedded and special purpose systems need a *qualitative* step forward in the research efforts better than continue in *quantitatively* improve the designs: it's time for scaling-out architectures, instead of scaling-up frequency. As transistor count is still increasing as expected by Moore's law, recent challenges like wire-delay, design complexity, and power requirements are becoming more and more important. These problems are preventing the evolution of chip architecture in the directions followed in the previous decades, when clock frequency as well could scale-up with Moore's law. Many researchers and companies have started to look at building multiprocessors on a single chip, following both past and novel design solutions: no doubt that we are all expecting several cores on a single chip in the near future.

Such single-chip architectures are also expected to have full success in the embedded domain. In this case, application specific requirements would also demand both computational power provided by CPU cores, embedded memory, high-speed on-chip networks, but also different and novel ways to clusterize resources, and managing overall design complexity in a competitive time-to-market. In addition, low-power consumption and efficient power management are transversal constraints that should to be considered when proposing and developing such system architectures.

Clearly, this opens several new challenges for (embedded) computer architects as well as for the (embedded) experts in each of the software layers above the hardware. In particular, programmers may have to change the programming model, compilers may have to consider new ways to parallelize the application, and multiple levels of parallelism are likely to be taken into account. On the hardware side, an open question is: which approach would give major benefits?

Tiling small cores, thinking to new organizations, or recovering many lessons learned from the past experience of multiprocessors and supercomputers?

Research efforts in this field both from industry and from academy are focused on a variety of strategic topics that have to be addressed in order to develop future embedded single-chip multicore architectures:

- parallelism extraction and exploitation, as well as, techniques for modular design of complex multicore architectures;
- low power designs and techniques for energy-efficient architectures;
- Cache and memory subsystems able to deliver high performance and efficient support to the running applications.

This issue has attracted a good number of quality papers in the above-mentioned crucial topics. In particular, several contributions address the parallelism issues.

Ródenas et al. propose an analysis and evaluation of a multilevel parallelism approach for chip multiprocessor (CMP) on the experimental IBM BG/Cyclops architecture. The reference benchmarks are multi-zone parallel applications and their parallelism is exposed through OpenMP. Paper results highlight that accurate thread grouping, balancing and conscious allocation of

ISSN 1740-4460/06/$17.00

threads on the architectural resources can allow for high performance gains.

Oliver et al. present Synchroscalar, a single-chip, multi-core architecture specifically targeted for multimedia applications in the embedded domain. In particular, the authors address energy efficiency and low complexity of design through specific parallelization of activities. Inter-core interconnection, multiple frequency domains and resource tiling are analyzed for seeking power efficiency for the target applications.

Fatemi et al. propose a solution for overcoming the limitations in the interconnection network of traditional SIMD architectures. The proposal addresses a reconfigurable interconnection network that is able to deliver almost the same performance as a fully connected one (FC-SIMD) but without the area overhead of the latter. SIMD architectures are addressed because of their area-energy efficiency. The efficiency in the interconnection network is achieved through a specific delay line in the instruction bus that is able to implement a pipelining between instructions able to exploit the SIMD resources. Efficient and effective pipelining can be achieved only if an appropriate software scheduling is adopted. The paper shows the required features of such a scheduler.

The paper from Gu et al. studies the possibility to achieve more than a hundred-fold speedup with a parallel architecture over a serial one in case of a computing intensive application: a gate level logic simulator. The application is implemented through an Explicit- multi-threading (XMT) parallel programming model in order to exploit the on-chip parallelism: mainly fine-grained SIMD-like parallelism is addressed.

Abderazek et al. study a Queue based instruction-set architecture processor that can be particularly suitable for achieving high-performance in selected fields of the embedded domain. In particular, a general queue based instruction set and architecture is presented, along with the methodology to derive application-specific designs. In addition, the work shows how to address the issues of performance verification and functional correctness validation through different platforms like FPGA emulation as well as Verilog simulation.

Also research on cache and memory systems is crucial in the field of embedded architectures. Naz et al. study the possibility of employing a very small data cache, split for serving temporal and spatial streams respectively. The authors highlight that such approach can improve performance of embedded applications (MiBench benchmarks are considered) without much area and energy overhead.

Settle et al. study how to support the memory footprint of co-active application threads through specific hardware-based adaptable cache allocation systems. This is an interesting topic because chip multiprocessors are expected to run multithreaded workloads and the cache interference between different threads can reduce the system performance dramatically. The proposed dynamic management solution allows a 47% improvement over a fully shared two-level cache and a 10% improvement over the state-of-the-art cache partitioning model.

Low power topics are extremely important in the field of the upcoming complex embedded systems and La Fratta et al. address this issue from the system-level design perspective. In fact, the main problem in SoC design is the development of new methodologies and tools for keeping costs low while meeting the required levels of power consumption and performance. High-level design space exploration tools are expected to assist system designers to tune the partitioning and organization of on-chip resources to seek efficient high-performance architectures. The authors exemplify their proposal in case of a single-chip message passing parallel computer and highlight the memory configuration that delivers the best power-performance tradeoff.

Dybdahl et al. propose to increase the performance of embedded on-chip DRAM using a destructive approach that halves the access time on the reference architecture but requires a small write-back cache to restore memory content. The work demonstrates that the approach is able to increase performance (14%) with a reduced (3%) increase in power consumption.

Finally, Chu et al. study an efficient cache scheme that aims to reduce power consumption and conflict misses for single-core or multi-core embedded computing architecture. The proposed cache uses one gate stage before the access to the cache line, which allows buffering the last memory reference and thus implementing a tiny L0 cache. The idea is that if the data is in the buffer, the cache access can be aborted to reduce power consumption. Results show that the proposed scheme outperforms various conventional caches like direct-mapped and 2-way set-associative ones.

The guest editors would like to thank Paolo Bennati – very warmly – for his invaluable help and support in the management of the interaction with the authors and with the IOS-Press editors. His contribution has been extremely important for keeping the various activities within the established timetable.

Finally, a special thanks goes to the anonymous reviewers that made possible to achieve from two up

to four peer-reviews per paper, thus contributing with their suggestions to the final version of the accepted papers. Globally, we handled more than 50 reviews and we wish to explicitly thank: Bardine Alessandro – University of Pisa (Italy), Bechini Alessio – University of Pisa (Italy), Bennati Paolo – University of Siena (Italy), De Langen Pepijn – Delft University of Technology (The Netherlands), Foglia Pierfrancesco – University of Pisa (Italy), Kavi Krishna – UNT, The University of North Texas (USA), Kuzmanov Georgi – Delft University of Technology (The Netherlands), Mangano Daniele – University of Pisa (Italy), Moscu Panainte Elena – Delft University of Technology (The Netherlands), Popovic Ioannis Sourdis, Delft University of Technology (The Netherlands), Zdravko – University of Siena (Italy), Shahbahrami Asadollah – Delft University of Technology (The Netherlands).