The Grand Challenges in Microelectronics Design

µGC3: Moore For Less – Performance-driven design for next-generation chip technology;

30th Sept, Computer Labs,
University of Cambridge

Supported by the IET

Objectives

The objective of this Grand Challenge is to develop advances in technology and design that continue to deliver improvements in component density, performance, and cost-per-function in-line with Moore's Law whilst reversing the trend of past advances for electronics to demand an ever increasing share of the world's energy resources. These advances will be hard-won because of the adverse pressures of device variability, the end of advances in instruction level parallelism (ILP), the difficulties of designing with high levels of concurrency, and the challenges of designer productivity. This GC is very much in line with the ITRS, (International Technology Roadmap for Semiconductors) except that we now demand that the performance improvements come at no increase in power.

This grand challenge focuses on the concept of delivering "more for less": steering Moore's law through (or round) the various technology barriers that are becoming increasingly formidable. This is vital in moving the basic premise of Moore's law away from raw transistor numbers into a more sophisticated metric of performance in the future.

A key aspect of achieving such progress is the design (productivity) gap, a vital component in making progress and an area where the UK design community has a lot to offer.

Agenda:

9:30: Registration

10:00 **Dr Simon Moore,** University of Cambridge, "Introduction to the Moore for Less theme"

10:15 **Dr Marcelo Cintra,** University of Edinburgh, "Beyond Auto-Parallelisation: Compilers for Many-Core Systems"

10:45 **Prof. Ian Watson,** University of Manchester, "Memory Architecture for Many-Core Systems"

11:15 Coffee Break

11:45 **Dr George Constantinides,** Imperial College, "FPGA-based computation - making the most of your hardware"

12:15 **Krisztian Flautner,** R&D Director ARM Ltd – "The Wall Ahead is Made of Rubber"

12:45 Networking Lunch

14:00 **Dr Simon Moore,** University of Cambridge, "Communication, the next resource war"

Register here: http://www.regonline.co.uk/moore_for_less

14:30 **Workshop Session** – facilitated by partner organisations and aimed at identifying research project that might be of use to industry

15:30 **Feedback Session** – nominated speakers report the outcomes of each workshop group

16:00 Coffee

16:15 **Daniel Dearing, eKTN** – Summary and follow-up

16:30 **Networking**

17:00 Close

Presentation Abstracts

Dr Marcelo Cintra, University of Edinburgh. 'Beyond Auto-Parallelisation: Compilers for Many-Core Systems'

Optimising compilers for parallel shared-memory systems have been mainly associated with auto-parallelisation. Despite many decades of work on auto-parallelisation techniques these have only had marginal success and then only within a small range of scientific and engineering applications.

With multi-core systems now ubiquitous on the desktop/laptop and with many-core coming soon, the role of optimising compilers for such parallel systems has to be revisited. In particular, a new approach is needed for parallelisation of more irregular applications. The traditional auto-parallelisation approach must be replaced with holistic approaches that require more active - while still sensible - programmer participation and that allow for speculative parallelisation using recent techniques such as thread-level speculation and transactional memory. In this talk I will present our efforts at the University of Edinburgh to develop such a holistic and speculation-enabled compiler framework.

Prof. Ian Watson, University of Manchester. 'Memory Architecture for Many-Core Systems'

With future advances in technology, it will soon be possible to integrate a large number of processor cores on a single chip. The central processing part of these cores is unlikely to change significantly from those available today and the task of placing these on a single chip will be relatively straightforward. The big challenge is in how to enable these cores to co-operate in order that their collective potential can be harnessed to execute general programs. This is both a hardware and a software problem. A major hardware challenge is in devising a memory and communication architecture which will scale. It is highly unlikely that a way will be found to extend the strict bus-based memory coherence schemes found on current systems to more than a few tens of cores. The key may be to realise that future high level parallel software will not require the sort of fine grain, cache line level coherence currently provided.

This talk will consider the problems and possibilities for scalable memory systems particularly in the light of high level programming models which are being considered for many-core systems.

Dr George Constantinides, University of Manchester. 'FPGA-based computation - making the most of your hardware'

Field-Programmable Gate Arrays (FPGAs) are 'programmable hardware', and may be configured at start-up or on-the-fly to implement a wide variety of digital electronic designs. These devices are commonly used for rapid prototyping of complex computational systems. However, I will discuss a different paradigm - the use of FPGAs as general purpose computers, or as part of a general purpose computer system. In this approach, conceived by multiple researchers over the last 10-15 years, but only now coming of age, the general purpose FPGA is configured to act as a special purpose computational machine for whatever computation is required at the present time; the architecture is designed around the algorithm.

We will briefly discuss the evolution of FPGA architectures and draw comparisons with the evolution of general purpose processors, before discussing two main advantages of modern FPGA architectures, the ability to design numerical representation around the algorithm, and the potential memory bandwidth available.

Some of the numerous optimization problems arising when efficiently mapping to these architectures will be mentioned and we will discuss some of our approaches to tackle these problems.

Krisztián Flautner, R&D Director ARM Ltd – "The Wall Ahead is Made of Rubber"

Silicon technology evolution over the last four decades has yielded an exponential increase in integration densities with steady improvements of performance and power consumption at each technology generation. This steady progress has created a sense of entitlement for the riches that future process generations would bring. Today, however, classical process scaling seems to be dead and living up to technology expectations requires continuous innovation at many levels, which comes at steadily progressing implementation and design costs. Solutions to problems need to cut across layers of abstractions and require coordination between software, architecture and circuit features.

Dr Simon Moore, University of Cambridge. 'Communication, the Next Resource War'

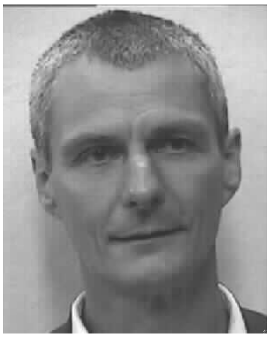
Scaling of electronics technology has brought us to a pivotal point in the design of computational devices. Technology scaling favours transistors over wires which has led us into an era where communication takes more time and consumes more power than the computation itself. We will discuss how this technology driver inevitably pushes us toward a communication-centric approach to computer systems design, from computer architectures through to algorithm design

Speaker Bios

Krisztián Flautner

Krisztián Flautner is the director of R&D at ARM Ltd. ARM designs the technology that lies at the heart of advanced digital products with more than ten billion processors deployed by early 2008. He leads a global team which is focused on the understanding and development of technologies relevant to the proliferation of the ARM architecture. The group's activities cover a wide breadth of areas ranging from circuits, through processor and system architectures to tools and software. Key activities are related to high-performance computing in energy-constrained environments. Flautner received a PhD in computer science and engineering from the University of Michigan, where he is currently appointed as a visiting scholar. He is a member of the ACM and the IEEE.

Professor Ian Watson



Ian Watson is Professor of Computer Science in the School of Computer Science, University of Manchester. He has worked for many years in the field of parallel computer architecture and is probably best known for his work on Dataflow and Reduction Machine architectures for the execution of declarative programming languages. His interests centre on the requirements of languages and computational models and their influence on the underlying machine architecture. His recent work has been concerned with the structure of multi-core and many-core machines with particular emphasis on the memory systems needed to support emerging new parallel programming approaches.

Dr Marcelo Cintra

Dr Cintra is currently a Lecturer of Computer Science at the University of Edinburgh. He is a member of the Institute for Computing Systems Architecture, within the School of Informatics. He received the Ph.D. degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 2001. He received both a M.Sc. degree in Computer Engineering and a B.Sc. degree in Electrical Engineering from the University of São Paulo in 1996 and 1992, respectively.

His research interests lie in the general areas of Computer Architecture, Parallel and High-Performance Computing, and Optimising Compilers. He has published in major journals and conferences in these areas. He was a guest editor of the Transactions on High-Performance Embedded Architectures and Compilers and has served in the program committee of

Euro-Par'08, SCOPES'08, IPDPS'07, ICPP'04, and IPDPS'03. He currently co-directs the Compiler and Architecture Design Group. His research is currently supported primarily by EPSRC and the European Commission. He is a member of the European Network of Excellence in High Performance Embedded Architectures and Compilers (HiPEAC).

Dr George A. Constantinides

George A. Constantinides received the M.Eng. (Hons) in Information Systems Engineering and the Ph.D. degree from Imperial College London, London, U.K., in 1998 and 2001, respectively. In 2002, he joined the faculty at Imperial College London and is currently a Senior Lecturer.

Dr. Constantinides is associate editor of the IEEE Transactions on Computers and the Journal of VLSI Signal Processing. He was programme co-chair of the IEEE International Conference on Field-Programmable Technology in 2006 and Field Programmable Logic and Applications in 2003, and serves on the technical program committees of several conferences, including FPL, FPT, and DATE, where he is chair of the Architectural Synthesis track in 2009.

Dr Simon Moore

Simon Moore is Reader in Computer Architecture at the Computer Laboratory, University of Cambridge where he leads the Computer Architecture research group (currently 11 PhD students, 2 RAs and two lecturing staff). His current interests are focused on a communication centred approach to the design of computer systems and algorithms.