Towards a Communication-Centric Design Methodology

A. Banerjee, R. Francis, J. Lee, J. May, S. W. Moore and R. D. Mullins
Computer Laboratory, University of Cambridge
Simon.Moore@cl.cam.ac.uk

The design of high performance computing devices has reached a major turning point. The next 20 years will require a step change in computer architectures and the way in which we approach many VLSI design tasks. In contrast to the past two decades, the scaling of CMOS fabrication technologies will no longer provide a simple route to improved performance. Central to many of these challenges is the ability to communicate on-chip in a low-power and robust fashion. In fact, technology scaling trends suggest that communication, not computation, will dominate delay, area and power budgets. Communication is also central to supporting an increasingly distributed design style where many heterogeneous or homogeneous IP blocks are integrated on a single chip to create a complete system. In many cases it will be the scheduling and management of these compute resources and the provision of the necessary communication resources that will be the major design and implementation challenge.

Our work to date has focused on the design and implementation of high-performance on-chip networks [3, 4]. Our latest test chip, Lochside, implements a 4x4 mesh network of low-latency virtual-channel routers. Our novel speculative router architecture permits each network hop (router and link) to be traversed in a single clock cycle.

The characteristics of these new communication-centric VLSI architectures also forces the choice of system-timing regime to be carefully reevaluated. The interconnection of many different, physically distributed, IP blocks operating at different or even adaptive clock frequencies poses significant challenges for existing synchronous design techniques. Our work has suggested new techniques for generating and distributing clocks while minimising power and skew [1]. Work has also explored the use of local clocks and event-driven synchronous systems [2]. Efficient interconnects may also be constructed using purely asynchronous design techniques.

The shift to communication-centric architectures involves much more than simply the replacement of on-chip buses with a scalable packet-switched interconnect. Future work is set to examine how communication-centric techniques can serve as a basis for reducing system complexity and cost, managing local power and thermal budgets, improving reliability and manufacturability and boosting performance.