Chapter 10
Implementation and Evaluation of Skip–Links:
A Dynamically Reconfiguring Topology for Energy–Efficient NoCs

Simon J. Hollis
University of Bristol, UK

Chris Jackson
University of Bristol, UK

ABSTRACT
The Skip-link architecture dynamically reconfigures Network-on-Chip (NoC) topologies in order to reduce the overall switching activity in many-core systems. The proposed architecture allows the creation of long-range Skip-links at runtime to reduce the logical distance between frequently communicating nodes. This offers a number of advantages over existing methods of creating optimised topologies already present in research, such as the Reconfigurable NoC (ReNoC) architecture and static Long-Range Link (LRL) insertion. This architecture monitors traffic behaviour and optimises the mesh topology without prior analysis of communications behaviour, and is thus applicable to all applications. The technique described here does not utilise a master node, and each router acts independently. The architecture is thus scalable to future many-core networks. The authors evaluate the performance using a cycle-accurate simulator with synthetic traffic patterns and compare the results to a mesh architecture, demonstrating logical hop count reductions of 12-17%. Coupled with this, up to a doubling in critical load is observed, and the potential for 10% energy reductions on a 16x16 node network.

INTRODUCTION
Network-on-Chips (NoCs) provide a general purpose communications fabric for multi-core architectures that allow functional blocks to communicate with each other. These networks are intended to provide a scalable, power and space efficient means to support growth in component density and many-core designs (Dally & Towles, 2001). Modern Chip Multi-Processors (CMPs) exploit NoCs to efficiently connect many processing elements (PEs) and other devices.
System performance is dependent on low-latency, high-bandwidth communications between cores. As the number of cores integrated in a single chip increases, greater demands are placed on the communication network. Thus, the design of the inter-connection fabric and not the PE is now the bottleneck in computation. Power and area are central considerations in the design of the interconnect. This has led to a wide range of literature, with topics including power analysis (Banerjee, Mullins, & Moore, 2007), router design (Michelogiannakis & Dally, 2009) and routing algorithms such as Opt-y (Schwiebert & Jayasimha, 1993) and IVAL (Towles, Dally, & Boyd, 2003). However, the main issue today is the delivery of energy-efficient and scalable interconnects: the focus of this paper.

**NoC Design**

An ideal NoC provides an area-efficient low-power network that exhibits good latency characteristics across a range of traffic patterns. A NoC can be designed for a specific application, affecting the choice of topology, routing algorithm and router design. These specialised networks can offer a performance advantage over a regular architecture, but are inflexible and costly to design. They require static analysis of the target application’s inter-process communications behaviour and offer poor performance for general purpose computation.

In Asanovic et al. (2006) Dwarfs were introduced, where applications are classified in terms of their computation and communication patterns. For example, some applications have fixed, predictable communication between specific cores, while others have sporadic communication between all cores. In Asanovic et al. (2006) the shortcomings of fixed-topology on-chip interconnects are discussed and it is suggested that many applications have communication patterns for which an optimal topology can be configured. Therefore, the design of a flexible, common NoC topology is of paramount importance in the design of efficient next-generation systems. For example, Shalf et al. (2005) consider reconfigurable hybrid interconnects for Ultra-scale applications. They conclude that many Ultra-scale applications will benefit from a reconfigurable interconnection fabric. In this paper, we demonstrate how this can be achieved in a general sense, using the mesh topology as a basis. We will introduce dynamically-reconfigurable Skip-links to enable on-the-fly adaptation of the network topology to traffic behavior.

**Energy in NoCs**

As shown by Moore and Greenfield (2008), the energy cost of communicating a single 32-bit word off-chip is now 1300 times greater than the cost of a single ALU operation. With the increase in the use of parallel algorithms and multiprocessing, communication and not computation is becoming the power and performance bottleneck. NoC routers dissipate both dynamic and static power, in particular when switching and reading/writing flits to buffers. Inter-router links also consume a significant amount of energy. The energy consumption of routers is discussed by Hu and Marculescu (2005) in the context of a model proposed by Ye et al. (2002), and can be expressed in terms of the router energy $E_R$ (buffering and switching) and link energy $E_L$ requirements. In this model, the total energy required to send a single bit from node $n_i$ to $n_j$ is

$$E_{n_i\rightarrow n_j} = N_{\text{hops}} \times E_R + (N_{\text{hops}} - 1) \times E_L \quad (1.1)$$

where $N_{\text{hops}}$ is the number of routers the bit traverses between nodes $n_i$ and $n_j$. Equation (1) clearly shows that the energy efficiency of a NoC can be increased by reducing the value of $E_R$ (the energy expenditure when switching); or $E_L$, the