Exploring the Scalability and Performance of Networks-on-Chip with Deflection Routing in 3D Many-core Architecture

AWET YEMANE WELDEZION
Abstract

Three-Dimensional (3D) integration of circuits based on die and wafer stacking using through-silicon-via is a critical technology in enabling "more-than-Moore", i.e. functional integration of devices beyond pure scaling ("more Moore"). In particular, the scaling from multi-core to many-core architecture is an excellent candidate for such integration. 3D systems design follows is a challenging and a complex design process involving integration of heterogeneous technologies. It is also expensive to prototype because the 3D industrial ecosystem is not yet complete and ready for low-cost mass production. Networks-on-Chip (NoCs) efficiently facilitates the communication of massively integrated cores on 3D many-core architecture. In this thesis scalability and performance issues of NoCs are explored in terms of architecture, organization and functionality of many-core systems.

First, we evaluate on-chip network performance in massively integrated many-core architecture when network size grows. We propose link and channel models to analyze the network traffic and hence the performance. We develop a NoC simulation framework to evaluate the performance of a deflection routing network as the architecture scales up to 1000 cores. We propose and perform comparative analysis of 3D processor-memory model configurations in scalable many-core architectures.

Second, we investigate how the deflection routing NoCs can be designed to maximize the benefit of the fast TSVs through clock pumping techniques. We propose multi-rate models for inter-layer communication. We quantify the performance benefit through cycle-accurate simulations for various configurations of 3D architectures.

Finally, the complexity of massively integrated many-core architecture by itself brings a multitude of design challenges such as high-cost of prototyping, increasing complexity of the technology, irregularity of the communication network, and lack of reliable simulation models. We formulate a zero-load average distance model that accurately predicts the performance of deflection routing networks in the absence of data flow by capturing the average distance of a packet with spatial and temporal probability distributions of traffic.

The thesis research goals are to explore the design space of vertical integration for many-core applications, and to provide solutions to 3D technology challenges through architectural innovations. We believe the research findings presented in the thesis work contribute in addressing few of the many challenges to the field of combined research in many-core architectural design and 3D integration technology.

Keywords
Alpha-model, Average distance, B-Model, NoC, Zero-load predictive model, deflection routing, q-routing