Novel Interconnect Infrastructures for Massive Multicore Chips – An Overview

Partha Pratim Pande, Amlan Ganguly, Benjamin Belzer
School Of Electrical Engineering and Computer Science
Washington State University
Pullman, USA
(pande, ganguly, belzer)@eecs.wsu.edu

Alireza Nojeh, Andre Ivanov
Department of Electrical and Computer Engineering
University of British Columbia
Vancouver, BC, Canada
(anojeh, ivanov)@ece.ubc.ca

Abstract—With the well-known trend of CMOS scaling as per Moore’s Law, traditional on-chip interconnect systems are reaching the point of having a very limited ability to meet the performance needs and specifications of Systems-on-Chip (SoCs). The conventional two-dimensional (2D) copper-based IC has inherent limitations due to the geometrical constraints of the planar structure. Innovative interconnect paradigms based on optical technologies, RF/wireless, carbon nanotubes, or 3D integration are promising alternatives that may indeed overcome the challenges encountered. In this paper we present an overview of different emerging non-traditional approaches to achieve massive degree of integration in a single chip. The advantages and underlying challenges of each method are highlighted.

I. INTRODUCTION

The current trend in System-on-Chip (SoC) design in the ultra deep sub-micron (UDSM) regime and beyond is to integrate a huge number of functional and storage blocks in a single die [1]. The possibility of this enormous degree of integration gives rise to new challenges in designing the interconnection infrastructure for these big SoCs. Extrapolating from the existing CMOS scaling trends, traditional on-chip interconnect systems have been projected to be limited in their ability to meet the performance needs of SoCs at the UDSM technology nodes and beyond. This limit stems primarily from global interconnect delay significantly exceeding that of gate delays. While copper and low-k dielectrics have been introduced to decrease the global interconnect delay, they only extend the lifetime of conventional interconnect systems a few technology generations. According to the International Technology Roadmap for Semiconductors (ITRS), [2] for the longer term, material innovation with traditional scaling will no longer satisfy the performance requirements and new interconnect paradigms are needed. Continued progress of interconnect performance will require approaches that introduce materials and structures beyond the conventional metal/dielectric system, and may require information carriers other than charge. According to the ITRS, multiple options have been envisioned to provide alternatives to the metal/dielectric system among which interconnect innovations such as optical, RF/wireless or 3D integration are promising alternatives. In addition to the design aspects, the other significant challenges are the problems of reliability and developing adequate CAD tools. With technology scaling the interconnect fabrics will be severely affected by different sources of transient and permanent failures. The big question that needs to be addressed is how can we produce reliable, predictable interconnect architectures out of inherently unreliable components. The new interconnect technologies will enable integration of heterogeneous components (electrical and non-electrical) in a single SoC. Design of these complex and heterogeneous SoCs will require new simulation and modeling tools.

In this paper our aim is to highlight the advantages and challenges associated with the emerging novel interconnect paradigms for designing SoCs in near future.

II. 3D INTEGRATION

Interconnect delays are increasingly dominating performance of SoCs due to increases in chip size and reduction in the minimum feature size. A simple way to reduce the burden of high frequency signal propagation across monolithic ICs is to reduce the line length needed by stacking active devices using 3D interconnects. The layers of active devices are separated by a few tens of micrometers. Consequently, the 3D interconnects allow communication among active devices with very small distance required for signal propagation. Topol et al., in [3] describe the advantages of manufacturing in a 3D IC process, including improvements in power, noise, logical span, density, performance, and functionality. One major advantage of the 3D IC paradigm is that it allows for the integration of “dissimilar technologies”, e.g. memory, analog, MEMS, etc. on a single die.

As shown in Fig. 1, three-dimensional (3D) integrated circuits (ICs), which contain multiple layers of active devices, have the potential for enhancing system performance [4]. According to [3], three-dimensional ICs allow for performance enhancements even in the absence of scaling. This is the result of each transistor being able to reduce interconnect length and access more nearest neighbors. Besides this clear benefit, package density is increased

Benjamin Belzer is supported by the National Science Foundation under grant (CCF-0635390) for a part of the work presented here

978-1-4244-1684-4/08/$25.00 ©2008 IEEE 2777

Authorized licensed use limited to: Washington State University. Downloaded on June 17,2010 at 16:25:23 UTC from IEEE Xplore. Restrictions apply.
which affect the net yield of 3D chips. Mismatches between yields of different layers, arising due to mismatches between yields of different layers, needs to be considered [6]. Additionally die yield issues may affect the yield of the entire IC. In 3D devices an important problem in 3D circuits arises from the interconnect and device reliability in 2D devices. Due to reduction in chip size of a 3D implementation, the performance of microprocessors by forming a processor-memory stack. The integration of processor and memory in a stack enables a large increase in performance. In particular, 3D integration enables the use of very wide buses (>1024 bits) for vertical communication. In addition to ultra wide buses, a stack provides a very short distance between processor and memory, decreasing memory access times considerably.

3D ICs will have significant impact on the design of multi-processor SoC (MP-SoC) platforms. The Network on Chip (NoC) has emerged as an effective methodology for designing big multi-core SoCs. However, the conventional two dimensional (2D) IC has limited floor-planning choices and, consequently, limits the performance enhancements arising out of NoC architectures. The performance improvement arising from the architectural advantages of NoCs will be significantly enhanced if 3D ICs are adopted as the basic fabrication methodology. The amalgamation of two emerging paradigms, namely NoCs in a 3D IC environment, allows for the creation of new structures that enable significant performance enhancements over more traditional solutions. With freedom in the third dimension, architectures that were impossible or prohibitive due to wiring constraints in planar ICs are now possible, and many 3D implementations can outperform their 2D counterparts [5].

In spite of all the advantages discussed above, 3D integration has its own shortcomings. The most important issue in 3D ICs is the heat dissipation. Thermal effects are already impacting interconnect and device reliability in 2D circuits. Due to reduction in chip size of a 3D implementation, 3D circuits exhibit a sharp increase in power density. Another important problem in 3D circuits arises from the interconnect coupling capacitance and crosstalk. In 3D devices an additional electrical coupling between the top layer metal of the first active layer and the devices on the second active layer needs to be considered [6]. Additionally die yield issues may arise due to mismatches between yields of different layers, which affect the net yield of 3D chips.

III. OPTICAL INTERCONNECTS

Optical communication is another promising approach to replace the traditional conductor/dielectric system for signal transmission in a SoC. The design of an optical clock distribution network is demonstrated in [7]. It requires a single photonic source coupled to a symmetrical waveguide structure routing to a number of optical receivers. At the receivers the high speed optical signal is converted to an electrical one and provided to local electrical networks. It is not feasible to route the optical signal all the way down to the individual gates as each drop point will require a receiver circuit, which consumes area and power. The clock signal is initially routed optically to a number of drop points, which will cover a zone and in each zone the final part of the clock network will be electrical. Through detailed analysis it is shown that the power dissipation in optical clock distribution will be lower than that of an electrical clock distribution.

In addition to the clock network the optical communication medium can be used as the interconnection backbone for MP-SoC platforms. In this regard the design of a photonic NoC has been demonstrated in [8]. Using wavelength division multiplexing (WDM) techniques, combined with photonics and optoelectronics optical networks on chip (ONoC) can be designed. As mentioned in [8], the design of optical NoCs is based on the advances made over the past several years in silicon photonics. Fabrication capabilities and integration with commercial CMOS chip manufacturing open up new avenues in designing on-chip optical networks. Integration of a fully functional photonic system on a VLSI die is envisioned. The photonic elements necessary to build a complete on-chip photonic network, like dense waveguides, switches, modulators and detectors, are now viable for integration on single silicon chip [8]. The photonic on-chip network is demonstrated to be more power efficient than the electrical network for high-bandwidth communications [8].

Although the optical interconnect option has many advantages, still many aspects of this new paradigm need more extensive investigation. Though very high bandwidth is achievable as the signal travels at the speed of light, the delays associated with the rise and fall times of optical emitters and detectors need to be considered. The speed of light in the transmitting medium, losses in the optical waveguides and the signal noise due to coupling between waveguides are the other important issues that need more careful investigation.

IV. WIRELESS INTERCONNECTS

A relatively radical alternative to metal/dielectric interconnects is to use transmission of signals via RF/microwave. Transmission in this case can be wireless communication through the package and IC structures or through waveguides. Chang et al. [9] discussed various implementation issues regarding building RF/Wireless interconnects for inter-and intra-chip communications. They advocated the use of microwave transmission in guided media, such as the micro strip transmission line (MTL) or coplanar waveguide (CPW), instead of wireless communication. According to them, for transmission of RF/microwave signals in free space the antenna aperture size will be too large for on-chip implementation. But we believe nanoscale antennas based on carbon nanotubes (CNTs), such as shown in Fig. 2, can be used to create an on-chip wireless interconnect network using optical frequencies. CNTs, hollow cylindrical carbon structures with nanoscale diameters, have attracted significant interest.

![Figure 1. 3D IC from SOI process](image_url)
attention over the past decade and a half due to their outstanding electronic and mechanical properties. With extremely high aspect ratios, they are virtually ideal one-dimensional wires with potential for antenna applications. For instance, a single-walled carbon nanotube (SWNT) with a diameter of approximately 1 nanometer can have a length of up to a millimeter or even centimeter. Nanotube devices with lengths of a few hundreds of nanometers or micrometers are routinely fabricated presently, and would be suitable for visible or infrared antennas. In addition, their nearly defect free structure and high conductance (ballistic transport) would lead to low loss antennas. Another important factor is their high current carrying capacity of $10^9$ A/cm², which is orders of magnitude higher than silver and copper: one individual SWNT can carry several micro amperes of current, and this is crucial in order to ensure enough transmission power. Moreover, antenna arrays could be formed by placing a number of SWNTs next to each other in order to increase the total transmitted power, or potentially improve the directionality of the emission pattern, if desired.

In fact, antenna effects in CNTs have been studied experimentally [10-12] and theoretically [13, 14]. The authors in [15] even contemplate the usage of centimeter-long nanotube antennas operating at microwave frequencies to interface nanodevices to the macroscopic world.

**Figure 2.** A CNT device that we propose could be used as an antenna

Our group proposes to use short CNT antennas as the foundation of a wireless communication infrastructure (channels) to efficiently link multiple components of the SoC. Nanotubes can be metallic or have energy band gaps of up to approximately 1 eV. Given the direct band gap nature, these are appropriate for optical emission devices in the infrared part of the spectrum. Thus, channel operating frequencies (characterizing the channels linking individual cores) are envisaged to be in the range of $\sim 10^{14-10^{15}}$ Hz from the emission point of view. In terms of reception, nanotubes can be used in a much wider frequency range. These tiny devices will drastically reduce the area overhead compared to any other microfabricated antenna. The possibility of wireless interconnects for on-chip clock distribution has been explored already [16]. For clock signal distribution, a single tone (channel frequency) suffices. In the case of inter-core data transmissions considered here, multiple channels are required to achieve acceptable throughput. In wireless communication, channelization can be achieved by using either time-division (TD), code-division (CD), frequency-division (FD), or some combination of these. FD, also called multi-carrier communications, is the most suitable channelization scheme for on-chip wireless, as it can be realized with a number of practical filtering schemes. Orthogonal Frequency-Division Multiplexing (OFDM) is an attractive option for this purpose.

OFDM can easily adapt to severe channel conditions without complex equalization. It is robust against narrow-band co-channel interference and intersymbol interference (ISI), and there is no cross-channel interference if cyclic prefixes are used [17]. It can be efficiently implemented using the Fast Fourier Transform (FFT), and can also be implemented with fast transforms based on other orthogonal basis functions. In the on-chip wireless communication infrastructure, the principal problem may arise from the implementation of the multi-carrier modulation scheme. But in the on-chip environment we are not bound by existing OFDM standards. Instead, the multi-carrier scheme needs to be carefully tailored for such a miniaturized wireless application.

V. RELIABILITY INVESTIGATIONS

One of the principal challenges of all these emerging interconnect methodologies will be to achieve reliable communication under technology limitations. The major cause affecting the reliability of the global interconnects is the shrinking of the feature size, which exposes them to different faults of permanent, transient or intermittent nature. There are multiple ways to handle the reliability issues and according to [18], they can be classified as one of the following:

- **Hardware redundancy**
- **Information redundancy**
- **Time redundancy**
- **Hybrid of hardware and time redundancy**.

The fault tolerant design methods can operate at different abstraction levels. There are five key elements in a comprehensive approach to fault-tolerant design: avoidance, detection, containment, isolation, and recovery. Ideally, these are implemented in a modular, hierarchical design, encompassing an integrated combination of hardware and software techniques [19]. The implementation of fault tolerant methods at different levels of abstraction will be more cost effective. There has to be more extensive investigations regarding the exact implementation method to design robust interconnection infrastructures. Instead of building massive hardware-level redundancy it might be more economical to apply error control coding techniques. But depending on existing single error correcting codes it will not be possible to achieve the reliability requirements in the presence of high error rates in the future technology nodes. Low-complexity multiple error correcting codes need to be designed and their performance needs to be studied for on-chip communication environments.

Among different possibilities, it is difficult to predict which method will be best to design a reliable interconnect infrastructure. We may have to incorporate different techniques at different abstraction levels. Hardware redundancy may be adopted in the very low level and error control coding may be applied at the higher levels. To achieve optimum results multiple approaches need to be blended efficiently.
VI. DESIGN TOOLS

One of the desirable characteristics of these emerging interconnect paradigms is the ability to integrate heterogeneous components on a single die. CMOS needs to be integrated with MEMS, photonic components, etc. For design and simulation of such heterogeneous platforms we need CAD tools capable of handling multi-domain needs. The simulation for the heterogeneous systems can be done in a couple of ways [20]. First, the design tool can provide descriptions of different parts in a single simulation language. As an example, VHDL-AMS allows the extension of VHDL modeling to analog and mixed-signal domains, as well as to non-electrical media like optical. Similarly SystemC-AMS and Verilog-AMS can be utilized for this purpose. Second, the design tool can preserve specific descriptions of various parts and run various simulators in parallel, i.e., co-simulation. The Chatoyant [21] tool developed at the University of Pittsburg is an example of a co-simulation tool used for mixed technology microsystems.

We also need system-level models to integrate disparate components, so that the performance of the SoC can be explored rapidly. This will enable designers to create an overall execution model from the abstract interfaces and interconnections between the heterogeneous components [20]. As an example, Briere et al. have used Transaction Level Modeling (TLM) to design an optical on-chip network [22]. TLM facilitates system design at a higher abstraction level than the Register Transfer Level (RTL) and will speed up the simulation of a heterogeneous system.

The open SystemC initiative [23] is working on standardizing methodologies for design, modeling and verification for mixed technology systems.

VII. CONCLUSIONS

According to ITRS, due to the continuous increase in frequency and power of integrated circuits, we need to explore the feasibility of radically different interconnect systems beyond the traditional metal/dielectric concept. In this paper we have highlighted advantages of a few evolving interconnect paradigms. We have discussed the design challenges involving each new methodology and also the reliability issues. The need for new CAD tools to speed up the design process has also been discussed. All these emerging methodologies need extensive investigations to make them viable alternatives to replace the existing interconnect infrastructures.

REFERENCES