Complex Network Inspired Fault-Tolerant NoC Architectures with Wireless Links

Amlan Ganguly
Department of Computer Engineering
Rochester Institute of Technology
Rochester, NY, USA
+1-585-475-4082

amlan.ganguly@rit.edu

Paul Wettin, Kevin Chang
School of Electrical Engineering and
Computer Science
Washington State University
Pullman, WA, USA
+1-509-335-6249

{pwettin,jchang}@eecs.wsu.edu

Partha Pande
School of Electrical Engineering and
Computer Science
Washington State University
Pullman, WA, USA
+1-509-335-5223

pande@eecs.wsu.edu

ABSTRACT

The Network-on-Chip (NoC) paradigm has emerged as a scalable interconnection infrastructure for modern multi-core chips. However, with growing levels of integration, the traditional NoCs suffer from high latency and energy dissipation in on-chip data transfer due to conventional metal/dielectric based interconnects. Three-dimensional integration, on-chip photonic, RF and wireless links have been proposed as radical low-power and low-latency alternatives to the conventional planar wire-based designs. Wireless NoCs with Carbon Nanotube (CNT) antennas are shown to outperform traditional wire based NoCs by several orders of magnitude in power dissipation and latency. However such transformative technologies will be prone to high levels of faults and failures due to various issues related to manufacturing and integration. On the other hand, several naturally occurring complex networks such as colonies of microbes and the internet are known to be inherently fault-tolerant against high rates of failures and harsh environments. This paper proposes to adopt such complex network based architectures to minimize the effect of wireless link failures on the performance of the NoC. Through cycle accurate simulations it is shown that the wireless NoC architectures inspired by natural complex networks perform better than their conventional wired counterparts even in the presence of a high degree of faults.

Categories and Subject Descriptors

B.8.1 [**Performance and Reliability**]: Reliability, Testing and Fault-Tolerance

General Terms

Performance, Reliability.

Keywords

Wireless Links, Small-World, Fault-Tolerance.

1. INTRODUCTION

Massive levels of integration are making modern multi-core chips all pervasive in several domains ranging from scientific applications like weather forecasting, astronomical data analysis, and bioinformatics to even consumer electronics like graphics and animation. Design of multi-core integrated systems beyond the current CMOS era will present unprecedented advantages and challenges, the former being related to very high device densities

and the latter to soaring power dissipation issues. According to the International Technology Roadmap for Semiconductors (ITRS) in 2007, *the contribution of interconnects to chip power dissipation is expected to increase from 51% in the 0.13μm technology generation to up to 80% in the next five-year period*. This clearly indicates the challenges facing future chip designers associated with traditional scaling of metal interconnects and material innovation. Conventionally, the Network-on-Chip (NoC) paradigm is used as a scalable interconnection infrastructure for multi-core chips [1]. Advances in NoC research along several dimensions spanning architectural explorations, development of routing protocols, and reliability have made it the choice for the communication backbone of complex multi-core chips. However, despite its widespread adoption, traditional NoCs suffer from the limitations arising out of planar metal interconnect-based multi-hop communication, which in turn gives rise to high power and latency. Several attempts have been made to improve the limitations of conventional NoC architectures by innovations in routing and/or supplementary interconnect deployments [2]. However, due to the basic interconnect technology being the metal/dielectric combination the performance improvements were only incremental. Hence, advantages of radically new interconnect technologies need to be harnessed in NoCs. To enhance the performance of conventional metal interconnect-based multi-core chips a few radically different interconnect technologies are being currently explored such as, 3D integration, Photonic interconnects and multi-band RF or wireless interconnects [3-5]. All these new technologies have been predicted to be capable of enabling multi-core designs that improve the speed and power dissipation in inter-core data transfer significantly. However, these alternative interconnect paradigms are in their formative stage and need to overcome significant challenges pertaining to integration and reliability. So far, all these interconnect technologies have been used in existing multi-core platforms without significant architectural innovations, which undermines their adoptability in the face of potentially high levels of malfunctions.

Many naturally occurring large irregular networks like the internet, microbial colonies, cortical interconnections or social groups are found to display a striking resilience to faults either in the form of random failures or targeted attacks. Theoretical studies in complex networks reveal that certain types of network connectivity are inherently more resilient to failures [6]. A particular class of complex networks called small-world networks has a fairly homogeneous structure where all nodes have almost the same number of links. Both attacks and random failures of such links marginally affect the overall connectivity of this network. The reason for this behavior is the fact that it has a low average distance between cores and that no particular link contributes more to the interconnectedness of the system than the others due to the

homogeneity of the network. Adopting novel architectures inspired by such natural complex networks in conjunction with the emerging interconnection technologies will enable design of high-performance and robust multi-core chips for the future.

Wireless NoC is one of the promising alternatives to alleviate the performance limitations of traditional multi-hop counterparts. However, wireless interconnects with either Carbon Nanotube (CNT) based antennas or mm-wave metal antennas may encounter significant failure rates pertaining to issues of integration and transceiver design respectively. NoCs using these emerging interconnects demand high performance from inherently unreliable technology. Small-world networks have very short average distance between any pair of nodes and are known for their robustness and efficient communication mechanisms [7-8]. Hence, adopting a small-world connectivity with on-chip wireless links will not only significantly improve NoC performance; it will also enhance the robustness of the system. In this paper, we demonstrate how small-world NoC architectures are capable of handling a high-degree of possible faults in the wireless communication channels without significant impact on performance.

2. RELATED WORK

Conventional NoCs use multi-hop packet switched communication. To improve performance, the concept of express virtual channels is introduced in [9]. It is shown that by using virtual express lanes to connect distant cores in the network, it is possible to avoid the router overhead at intermediate nodes, and thereby greatly improve NoC performance. NoCs have been shown to perform better by inserting long range wired links following principles of small world graphs [2]. The design principles of three-dimensional NoCs and photonic NoCs are elaborated in various recent publications [3-4]. It is estimated that 3D and photonic NoCs will dissipate significantly less power than their electronic counterpart. Another alternative for low power is NoCs with multi-band RF interconnects [5]. Recently, the design of a wireless NoC based on CMOS Ultra Wideband (UWB) technology was proposed [10]. In [11] the feasibility of designing on-chip wireless communication network with miniature antennas and simple transceivers that operate at the sub-THz range of 100-500 GHz has been demonstrated. If the transmission frequencies can be increased to THz range then the corresponding antenna sizes decrease, occupying much less chip real estate. One possibility is to use nanoscale antennas based on CNTs operating in the THz frequency range [12]. Consequently building an on-chip wireless interconnection network using THz frequencies for inter-core communications becomes feasible. Design of a wireless NoC operating in the THz frequency range using CNT antennas is elaborated in [13]. All these emerging interconnect technologies are shown to improve the performance and power dissipation of NoCs. However due to the fact that these technologies are still in their formative stages their fabrication and integration with standard CMOS processes are unreliable. Moreover, system level techniques for sustaining gains in performance in presence of these inherently unreliable technologies have not received much attention from researchers. In this paper we propose an architecture that will sustain the advantages of one such technology even in the presence of inherent unreliability. In particular we propose the design of a wireless NoC using THz wireless links with CNT based antennas which will enable high gains in performance and power dissipation despite high rates of failures of the wireless links.

3. COMPLEX NETWORK INSPIRED ARCHITECTURE

Naturally occurring complex networks with a power-law based connectivity model [6] are shown to exhibit robustness against high rates of failures of nodes or links. One such type of network is commonly known as small-world network. In this work we propose the design of fault-tolerant small-world wireless NoCs (SWNoCs) with a power-law based interconnection architecture and investigate its performance in presence of failures of the wireless links. In this section we discuss the various aspects of the design methodology for the SWNoC architecture like topology, routing and physical layer.

3.1 SWNoC Topology

In the proposed SWNoC topology, each core is connected to a NoC switch and the switches are interconnected using wireline and wireless links. The topology of the SWNoC is a small-world network where the links between switches are established following a power law distribution as shown in (1).

$$P(i, j) = \frac{l_{ij}^{-\alpha} f_{ij}^{\beta}}{\sum_{\forall i} \sum_{\forall j} l_{ij}^{-\alpha} f_{ij}^{\beta}} \qquad (1)$$

Where, the probability of establishing a link, between two switches, i and j, $P(i,j)$, separated by an Euclidean distance of l_{ij} is proportional to the distance raised to a finite power [7]. The frequency of traffic interaction between the cores, f_{ij}, is also factored into (1) so that more frequently communicating cores have a higher probability of having a direct link. This frequency is expressed as the percentage of traffic generated from i that is addressed to j. This frequency distribution is based on the particular application mapped to the overall SWNoC and is hence set prior to wireless link insertion. Therefore, the apriori knowledge of the traffic pattern is used to establish the topology with a correlation between traffic distribution across the NoC and network configuration as in [14]. This optimizes the network architecture for non-uniform traffic scenarios. The parameters, α and β govern the nature of connectivity and the significance of the traffic pattern on the topology respectively. In particular, a bigger α would mean a very locally connected network with few or no long range links similar to that of a cellular automata based topology. Whereas, a zero value of α would generate an ideal small-world network following the Watts-Strogatz model [8] with long range shortcuts virtually independent of the distance between the cores. A higher value of β implies a higher probability of establishing direct links between switches with higher traffic exchange. Both of these parameters can be considered as design knobs that maybe tuned for a particular application to generate optimized network architecture depending on floorplan and traffic. In this work we look at the performance of a wireless NoC with small-world topology by taking into consideration physical parameters like interconnect characteristics. As long wired interconnects are extremely costly both in terms of power and latency we use wireless links to connect switches that are separated by a long distance. Depending upon the available wireless resources as explained in section IIID, we have a constraint on the maximum number of possible wireless links in the SWNoC, say, n. Hence, we make the n longest links in the SWNoC wireless while the others still remain wireline. Figure 1 represents such a SWNoC with 16 cores where each core is associated with a NoC switch (not shown for clarity). It has many short range local links as well as few long range shortcuts schematically represented by the arching
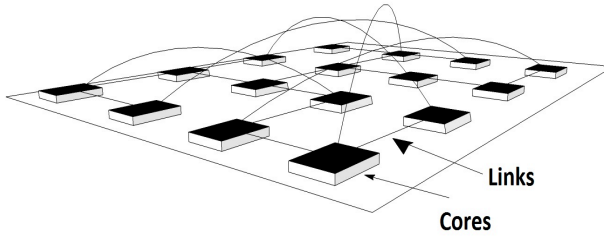
Figure 1. SWNoC architecture with short and long range links.

interconnects. The longest of these links would be wireless whereas the rest will be wireline.

Most naturally occurring small-world networks are created by self-assembly and does not guarantee a fully connected topology as it is fundamentally stochastic in nature. So, we adopted a simple iterative algorithm to ensure that the SWNoC is fully connected. To establish the network connectivity each pair of switch in the NoC is selected and a link is established between them with the probability given in (1). The network setup is repeated until a fully connected system is formed. We have assumed an average number of connections from each switch to other switches, $<k>$. Also, an upper bound, k_{max} is imposed on the number of links attached to a particular switch so that no particular switch becomes unrealistically large in the SWNoC. Also, it reduces the skew in the distribution of the links among the switches. The parameters α and β can also be chosen to aid in ensuring a fully connected network. For instance, a lower value of α or β means a more even distribution of the links and hence less chance of some switches left being isolated.

Thus, we propose a hybrid small-world wireless NoC architecture where the longest shortcuts are wireless and the rest of the interconnections are standard metal wires.

3.2 Hierarchical Wireless NoC
There is another recently proposed wireless NoC architecture with CNT-based wireless links called WiNoC [13]. This architecture is a hierarchical hybrid wireless NoC. The whole NoC is subdivided into smaller networks called subnets. Cores within a subnet are connected with each other in a regular wireline network like a mesh, ring or star. The cores are connected to a central hub in each subnet through wireline links. These hubs are then connected among themselves using wireline links in a ring forming the upper level of the network. The wireless links are optimally distributed as long range shortcuts between a few hubs. We will evaluate the performance of the SWNoC architecture with respect to WiNoC.

3.3 On-chip Antennas for the Wireless Links
Suitable on-chip antennas are necessary to establish the wireless links. In [15] the authors demonstrated the performance of silicon integrated on-chip antennas for intra- and inter-chip communication. They have primarily used metal zig-zag antennas operating in the range of tens of GHz. An ultra-wideband (UWB) antenna was used in the design of a wireless NoC [10] mentioned earlier in section II. The above mentioned antennas principally operate in the millimeter wave (tens of GHz) range and consequently their sizes are on the order of a few millimeters.

If the transmission frequencies and bandwidths are to be increased to THz/optical range then the corresponding antenna sizes decrease, occupying much less chip real estate. Characteristics of metal antennas operating in the optical and near-infrared region of

the spectrum of up to 750 THz have been studied [16]. Antenna characteristics of carbon nanotubes (CNTs) in the THz frequency range have also been investigated both theoretically and experimentally [17, 12]. Bundles of CNTs are predicted to enhance performance of antenna modules by up to 40dB in radiation efficiency and provide excellent directional properties in far-field patterns [18]. Moreover these antennas can achieve a bandwidth of around 500 GHz, whereas the antennas operating in the millimeter wave range achieve bandwidths of tens of GHz. Thus, antennas operating in the THz/optical frequency range can support much higher data rates. CNTs have numerous characteristics that make them suitable as on-chip antenna elements for THz frequencies. Given wavelengths of hundreds of nanometers to several micrometers, there is a need for virtually one-dimensional antenna structures for efficient transmission and reception. With diameters of a few nanometers and any length up to a few millimeters possible, CNTs are the perfect candidates. Such thin structures are almost impossible to achieve with traditional microfabrication techniques for metals. Radiation characteristics of multi-walled carbon nanotube (MWCNT) antennas are observed to be in excellent quantitative agreement with traditional radio antenna theory [12], although at much higher frequencies of hundreds of THz. Using various lengths of the antenna elements corresponding to different multiples of the wavelengths of the external lasers, scattering and radiation patterns are shown to be improved. Such nanotube antennas are good candidates for establishing on-chip wireless communications links and are henceforth considered in this work.

As mentioned above in section IIIA, a subset of all the switches in the SWNoC has wireless links. These wireless links are connected to the switches using a wireless port, *WP*. Consequently, each WP will require wireless transceivers to be able to establish the bi-directional wireless links. The THz antennas can be excited using external laser sources [13]. As mentioned in [4], the laser sources can be located off-chip or bonded to the silicon die. Hence their power dissipation does not contribute to the chip power density.

To achieve line of sight communication between WPs using CNT antennas working at THz frequencies, the chip packaging material has to be elevated from the substrate surface to create a vacuum for transmission of the high frequency electromagnetic (EM) waves. Techniques for creating such vacuum packaging are already utilized for MEMS applications [19], and can be adopted to make creation of line of sight communication between CNT antennas viable. In classical antenna theory it is known that the received power degrades inversely with the 4th power of the separation between source and destination due to ground reflections beyond a certain distance. This threshold separation, r_0 between source and destination antennas assuming a perfectly reflecting surface, is given by (2).

$$r_0 = \frac{2\pi H^2}{\lambda}$$

(2)

Here H is the height of the antenna above the reflecting surface and λ is the wavelength of the carrier. Thus, if the antenna elements are at a distance of H from the reflective surfaces like the packaging walls and the top of the die substrate, the received power degrades inversely with the square of the distance until it is r_0. Thus H can be adjusted to make the maximum possible separation smaller than the threshold separation r_0 for a particular frequency of radiation used. Considering the hundreds of THz frequency ranges of CNT antennas and depending on the separation between the source and destination pairs in a single chip, the required elevation is only a few tens of microns.

Chemical vapor deposition (CVD) is the traditional method for growing nanotubes in specific locations by using lithographically patterned catalyst islands. The application of an electric field during growth or the direction of gas flow during CVD can help align nanotubes. However, the high-temperature CVD could potentially damage some of the pre-existing CMOS layers. To alleviate this, localized heaters in the CMOS fabrication process to enable localized CVD of nanotubes without exposing the entire chip to high temperatures are used [20]. Many of these processes lead to stochastically varying CNT characteristics with high standard deviations from the intended properties. These can cause high rates of defects in the systems designed using these structures. In particular, unreliable manufacturing techniques can result in high degrees of faults in the WPs used in the SWNoC. Hence, we propose to adopt a complex network based NoC architecture using these antennas where even extremely high rates of failure of the antennas have marginal effect on the performance of the system.

3.4 Adopted Routing and Communication Protocols

In the proposed SWNoC, data is transferred via a flit-based wormhole routing [1]. Due to the irregular topology of the SWNoC, the adopted routing algorithm needs to be aware of the established topology. We adopted a decentralized, shortest-path routing strategy where at each switch the shortest path to the destination for a particular packet is determined and the flits are advanced to the next switch. This is done by an exhaustive search of all possible paths between the source and the destination for that packet. This shortest path computation is done only for the header flit, the body and tail flits simply follow the path established by the header. It is worth mentioning that many wired interconnects in the SWNoC are sufficiently long such that they cannot be traversed in a single clock cycle. While determining the shortest path for each packet, the actual number of cycles required for the flits to traverse those links has been considered. The routing algorithm is shown in Figure 2.

By using multiband laser sources to excite CNT antennas, different frequency channels can be assigned to pairs of communicating source and destination nodes. This will require using antenna elements tuned to different frequencies for each pair, thus creating a form of frequency division multiplexing (FDM) creating dedicated channels between a source and destination pair. This is possible by using CNTs of different lengths, which are multiples of the wavelengths of the respective carrier frequencies. High directional gains of these antennas, demonstrated in [18], aid in creating directed channels between source and destination pairs. In [21], 24 continuous wave laser sources of different frequencies are used. Thus, these 24 different frequencies can be assigned to multiple wireless links in the SWNoC in such a way that a single frequency channel is used only once to avoid signal interference on the same frequencies. This enables concurrent use of multi-band channels over the chip. The number of wireless links in the network can therefore vary from 24 links, each with a single frequency channel, to a single link with all 24 channels. It is shown in [13] that 24 links each with a single channel provides the best overall network performance as all the channels are distributed over the NoC resulting in a better connectivity. Hence, in this work we assume a maximum of 24 wireless links each with a single channel for both the SWNoC and the WiNoC architectures. Currently, high-speed silicon integrated Mach-Zehnder optical modulators and demodulators, which convert electrical signals to optical signals and vice versa are commercially available [22]. The optical modulators can provide 10Gbps data rate per channel on these links. This rate is also expected to increase in the near future. At the receiver a low noise amplifier (LNA) can be used to boost
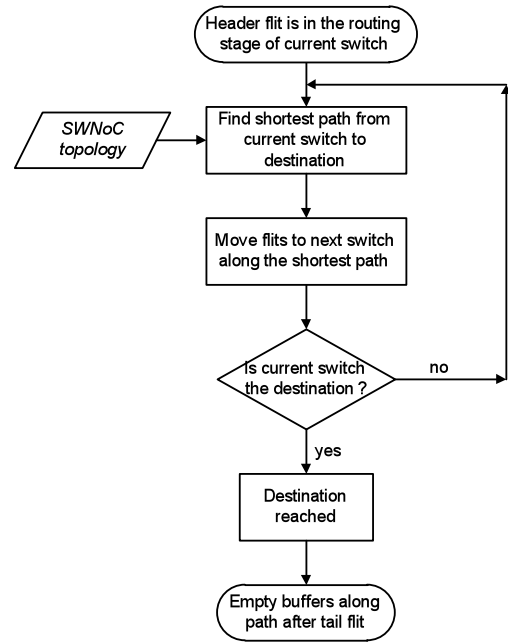

Figure 2. Distributed routing algorithm for the SWNoC.

the power of the received electrical signal, which will then be routed to the destination core. The modulation scheme adopted is non-coherent on-off keying (OOK), and therefore does not require complex clock recovery and synchronization circuits. Due to limitations in the number of distinct frequency channels that can be created through the CNT antennas, the flit width in NoCs is generally higher than the number of possible channels per link. Thus, to send a whole flit through the wireless link using a single channel a proper channelization scheme needs to be adopted. In this work we assume a flit width of 32 bits. Hence, to send the whole flit, time division multiplexing (TDM) is adopted. The various components of the THz wireless transceiver viz., the electro-optic modulators, the TDM modulator/demodulator, the LNA and the router for routing data on the network of hubs are implemented as a part of the WP.

4. EXPERIMENTAL RESULTS

In this section we characterize the performance of the proposed SWNoC in presence of wireless link failures through detailed system level simulations. We show that the performance of the proposed architecture is better than that of a conventional wireline mesh-based NoC even in the presence of very high rates of failure of the wireless links. The faults are injected into the system by randomly disabling a certain percentage of the wireless links. We also compare the performance of the SWNoC in presence of faults with the already proposed WiNoC architecture [13]. The effectiveness of the SWNoC in presence of non-uniform and application based traffic pattern is also evaluated. Lastly, we discuss the area overhead required for designing the SWNoC.

We evaluate the architectures considered in this work using a cycle accurate simulator which models the progress of data flits accurately per clock cycle accounting for flits that reach destination as well as those that are dropped. One hundred thousand iterations were performed to reach stable results in each experiment, eliminating the effect of transients in the first few thousand cycles. Two different system sizes with 64 and 256 cores arranged in equal-area tiles over a 20mmx20mm die are considered in this work. For the 64 core SWNoC $<k>$ and k_{max} is considered to be 4 and 8 respectively. For the 256 core system the value of those

parameters are chosen to be 6 and 10 respectively. The width of all wired links is considered to be same as the flit size, which is 32 in this paper. The particular NoC switch architecture has three functional stages, namely, input arbitration, routing/switch traversal, and output arbitration. The input and output ports including the WPs have four virtual channels per port, each having a buffer depth of 2 flits. Each packet consists of 64 flits. Similar to the wired links, we have adopted wormhole routing in the wireless links too. A simple flow control mechanism is adopted uniformly for wireless links in which, the sender WP stops transmitting flits only when a *full* signal is asserted from the receiver WP. This full signal is embedded in a control flit sent from the receiver to the sender only when the receiver buffer is filled above a predefined threshold. The network switches are synthesized from a RTL level design using 65nm standard cell libraries from CMP (http://cmp.imag.fr), using Synopsys Design Vision. The NoC switches are driven with a clock of frequency 2.5 GHz. The delays and energy dissipation on the wired links were obtained through HSPICE simulations taking into account the specific lengths of each link based on the established connections in the 20mm×20mm die following the logical connectivity of the small-world topology. The energy dissipation on the wireless links were obtained through analytical and experimental findings in [12] as outlined later in section IVC.

The next subsection describes the performance metrics that we use to evaluate the proposed SWNoC architecture. This is followed by the detailed analysis of its performance compared to conventional counterparts as well as architectures existing in literature using the same technology.

4.1 Performance Metrics

To characterize the performance of the proposed SWNoC architectures, we consider two network parameters: throughput and energy dissipation. Throughput is defined as the average number of flits successfully received per embedded core per clock cycle. The throughput, t is calculated according to (3).

$$t = \frac{\mu_C \phi}{N T_{sim}} \tag{3}$$

Where, μ_C is the total number of messages successfully routed, ϕ is the size of a single message in number of flits, N is the total number of cores in the NoC and T_{sim} is the simulation duration in number of cycles.

As a measurement of energy, we use the average energy dissipation per packet. Energy dissipation per packet is the average energy dissipated by a single packet when routed from the source to destination node through multiple switches, wired, and wireless links. For the wireless links, the main contribution to energy dissipation comes from the WPs, which include antennas, transceiver circuits and other communication modules like the TDM block and the LNA. Energy dissipation per packet, E_{pkt}, can be calculated according to (4) below.

$$E_{pkt} = \frac{\sum_{i=1}^{\phi} \sum_{j=1}^{N} (\eta_{wire,ij} E_{wire} + \eta_{wireless,ij} E_{wireless} + \eta_{buf,ij} E_{buf})}{N} \tag{4}$$

In (4), $\eta_{wire,ij}$, $\eta_{wireless,ij}$ and $\eta_{buf,ij}$ are the numbers of hops the ith flit of the jth message makes on wireline links, wireless links and the number of cycles it waits in the switch buffers due to congestions respectively. E_{wire}, $E_{wireless}$, E_{buf} are the energy dissipated by a flit traversing a single hop on the wired link including the switch, wireless link including the switch and WP and while waiting in the buffers in case of a congestion respectively.

4.2 Throughput of SWNoC with Wireless Link Failures

In this subsection we evaluate the performance of the proposed SWNoC architecture in terms of throughput with special emphasis on the behavior in presence of high rates of failure of the wireless links. First, we consider a uniform random spatial distribution of traffic.

Figure 3 shows the variation in throughput with injection load for the SWNoC for two different system sizes of 64 and 256 cores. The throughput is demonstrated for four situations, namely with no wireless link failure and with 25%, 50% and 75% of the links failing. For comparison we present the throughput of a conventional wireline mesh and the SWNoC architecture with all wireline links (SWNoC-wireline).

From Figure 3 we find that the throughput of the SWNoC is much higher compared to that of a simple wireline mesh of the same size. This is because of two reasons: the SWNoC, being connected following the principles of a small-world network has a much lower number of hops separating the cores along the shortest paths and the wireless links connecting the cores separated by long
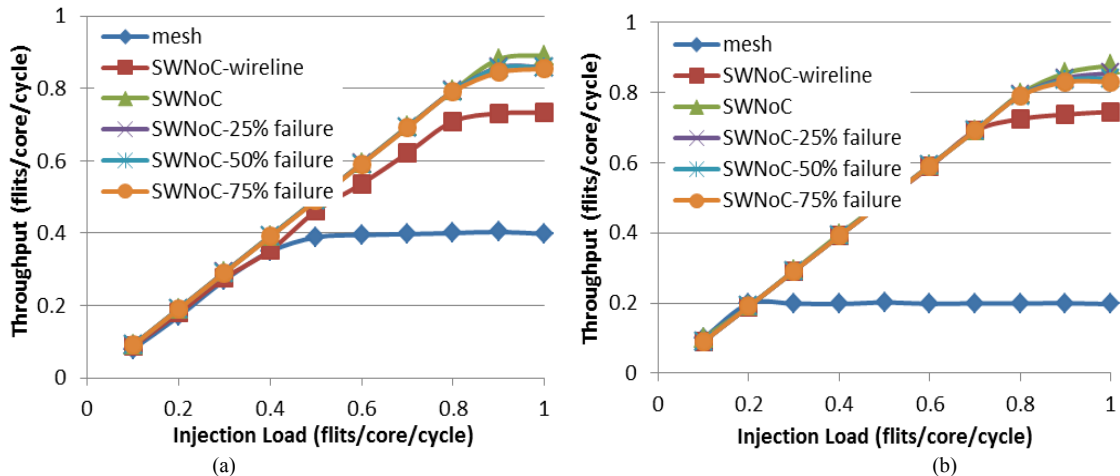

Figure 3. Throughput characteristic of SWNoC for (a) 64 core and (b) 256 core systems
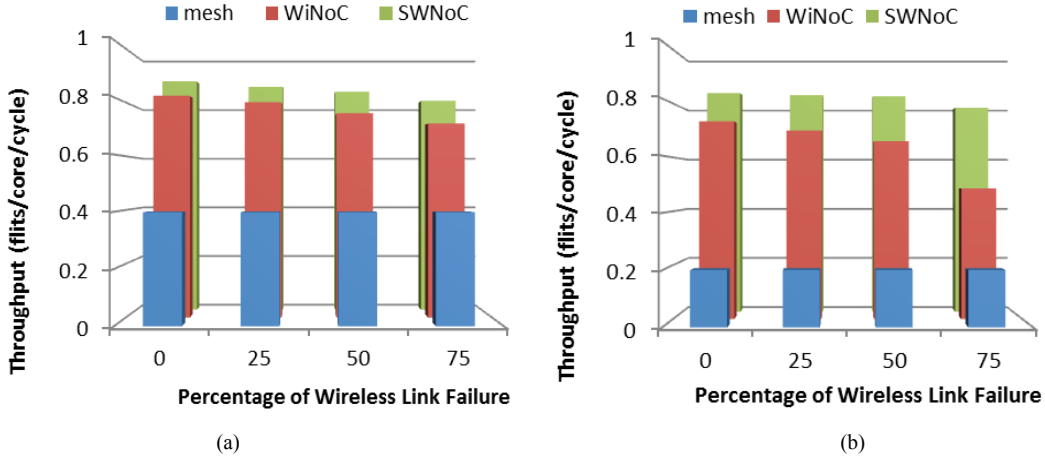
Figure 4. Saturation throughput of mesh, WiNoC and SWNoC for (a) 64 core and (b) 256 core systems in presence of faults

physical distances through single-hop low-latency links. The more interesting observation is that though we inject a high number of wireless link failures by randomly disabling a certain percentage of the wireless links, the throughput does not degrade considerably, especially for the larger system size. This is because the network connectivity is established following a small-world topology. It has been observed in [6] that small-world networks display amazing resilience to high rates of link failures. It has been shown that the average distance between nodes in small-world network increases very marginally with even extreme high rates of faults. Hence, the effect on the connectivity of the network is minimal. This is why the throughput does not degrade appreciably due to the failure of the wireless links. The slight degradation occurs due to the reduction in the total available wireless bandwidth for the SWNoC.

Next we compare the performance in terms of throughput at network saturation of the SWNoC and the hierarchical WiNoC in presence of the same degree of faults in Figure 4. The wireless link failures do not impact the performance of the flat wireline mesh architecture. Hence, its performance remains unchanged. Following the guidelines for best performance in [13] we considered the WiNoC to have 8 subnets each with 8 cores for a system size of 64 and 16 subnets each with 16 cores for the system size of 256. All the subnets were considered to have mesh-based connectivity among the cores within the subnets. We can see that the maximum achievable throughput of the SWNoC is better than the WiNoC in the absence of faults. This is because the whole SWNoC is a small-world network with low average distance between cores whereas; the subnets of the WiNoC are regular wireline networks, only the upper level of it has the small-world property. It is interesting to note that the performance of the WiNoC degrades comparatively more rapidly than that of the SWNoC with increasing degree of failure. The percentage reduction in throughput between the fault-free and 75% fault cases for the SWNoC is 8.6% for the 64 core system and 6.7% for the 256 core system. For the WiNoC the reduction in throughput is 11.1% and 33.9% for the 64 and 256 core systems respectively. This correlates to a very insignificant rise in the average distance between cores in the SWNoC as expected. The average distance between cores in the WiNoC however increases more with failure of the wireless links. This is because the WiNoC does not have a perfect small-world topology. The lower level of the hierarchy consists of wireline subnets which are established following a mesh topology. The degradation in performance for the WiNoC becomes more apparent for larger system size as the larger wireline subnets become the limiting factor in performance. The increase in average distance between cores is less with faults for larger system

size of the SWNoC and hence the percentage degradation in performance is even less for larger system sizes.

4.3 Energy Dissipation

In this subsection we evaluate the energy dissipation characteristics of the SWNoC and compare that with the conventional mesh. To determine the energy dissipation characteristics of the wireless architecture, we first estimated the energy dissipated by the antenna elements. As noted in [12], the directional gain of MWCNT antennas that we propose to use is very high. The ratio of emitted power to incident power is around -5dB along the direction of maximum gain. Assuming an ideal line-of-sight channel over a few millimeters, transmitted power degrades with distance following the inverse square law. Therefore the received power P_R can be related to the transmitted power P_T as

$$ P_R = \frac{G_T A_R}{4\pi R^2} P_T . \tag{5} $$

In (5), G_T is the transmitter antenna gain, which can be assumed to be -5dB [12]. A_R is the area of the receiving antenna and R is the distance between the transmitter and receiver. The energy dissipation of the transmitting antennas therefore depends on the range of communication. The area of the receiving antenna can be found by using the antenna configuration used in [12]. It uses a MWCNT of diameter 200nm and length 7λ, where λ is the optical wavelength. The length 7λ was chosen as it was shown to produce the highest directional gain, G_T, at the transmitter. In one of the setups in [12], the wavelength of the laser used was 543.5nm, and hence the length of the antenna is around 3.8μm. The area of the receiving antenna, A_T is calculated using these dimensions.

The noise floor of the LNA [23] is -101dBm. Considering the MZM demodulators cause an additional loss of up to 3dB over the operational bandwidth, the receiver sensitivity turns out to be -98dBm in the worst case. The length of the longest possible wireless link considered among all SWNoC and WiNoC configurations is 23mm. For this length and receiver sensitivity, a transmitted power of 1.3mW is required. Considering the energy dissipation at the transmitting and receiving antennas, and the components of the transmitter and receiver circuitry such as the MZM, TDM block and the LNA, the energy dissipation of the longest wireless link on the chip is 0.33 pJ/bit.

As mentioned earlier the values of energy dissipation on the wired links are obtained through HSPICE simulations taking into account the specific lengths of each link based on the established topology. The energy dissipation on the NoC switches are obtained by feeding a large set of data patterns into the gate-level netlists and

by running SynopsysTM Prime Power. The energy dissipations of the silicon-photonic components like the MZM modulators and demodulators are obtained from [22].

In Figure 5 we show the packet energy dissipation of the proposed SWNoC architecture in presence of various rates of wireless link failure for the two different system sizes considered in this work. For the sake of comparison we present the packet energy dissipation of the flat wireline mesh, a small-world NoC where all the links are wireline (SWNoC-wireline) as well the SWNoC with the long wires replaced by the wireless links in presence of various rates of failures. It can be seen that the SWNoC has significantly lower energy dissipation per packet compared to that of the mesh. This is because of better connectivity of the small-world based architecture compared to the regular mesh topology. In addition, the low power long range wireless links also contribute to a huge savings in energy dissipation signified by the difference in packet energy between the completely wireline SWNoC and the SWNoC with the wireless links. Due to the characteristics of the small-world topology of the SWNoC the energy dissipation increases slightly even in the presence of large number of wireless link failures. This demonstrates the resilience of the SWNoC towards random failures of the wireless interconnections without dissipating significant additional energy.

4.4 Performance of SWNoC in the Presence of Non-Uniform Traffic

So far we have considered only uniform random spatial distribution of traffic on the NoC to characterize its performance. In reality however, there could be various different types of traffic distributions on the chip depending on the applications considered. In this section we present the performance of the SWNoC in presence of several non-uniform traffic patterns. We consider both synthetic as well as application specific traffic scenarios on a SWNoC with 64 cores in this part. We considered two types of synthetic traffic to evaluate the performance of the proposed SWNoC architecture. First, a *transpose* traffic pattern [2] was considered where a certain number of cores were considered to communicate more frequently with each other. We considered 3 such pairs and targeted 50% of the traffic originating from those
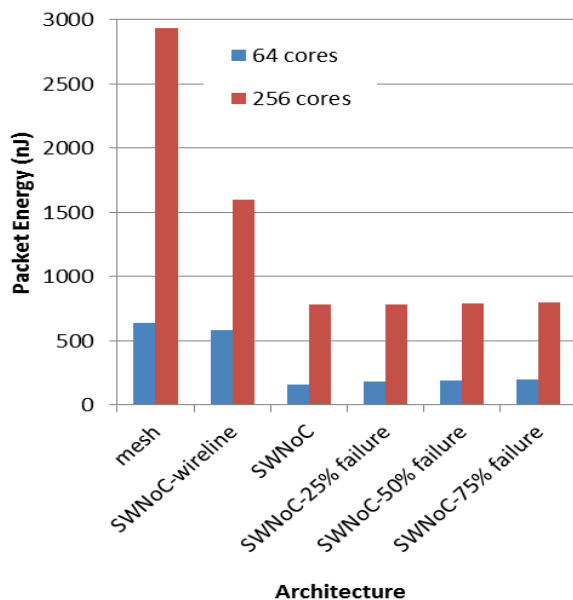

Figure 5. Packet Energy dissipation of the SWNoC architecture with different rates of wireless link failure
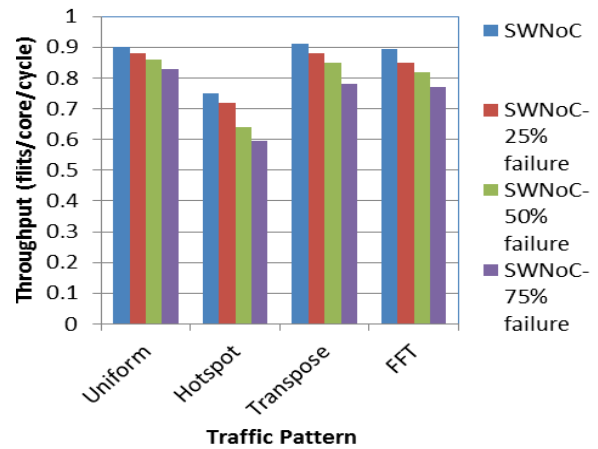

Figure 6. Saturation throughput of mesh and SWNoC for a 64 core system with different traffic patterns and wireless link failure rates

cores to the other core of the pair. The other synthetic traffic pattern considered was the *hotspot* [2], where all the cores in the SWNoC communicate with a particular core more frequently than with the others. We have assumed that all cores send 20% of the traffic originating in them to a particular core. To model application based traffic, a 256-point fast Fourier transform (FFT) application was considered, wherein each core performs a 4-point radix-2 FFT computation.

Figure 6 shows the maximum throughput at network saturation of the SWNoC in presence of the same rates of wireless link with various types of traffic. We see that for non-uniform traffic patterns there is relatively higher degradation in performance with increasing rates of failure than that in the case of uniform traffic. This is because, due to the adopted shortest-path routing strategy where packets are routed following only the shortest path from source to destination, failure of wireless links cause extra load on a particular alternative link, leading to congestion. However, it should be noted that the performance degradation with faults is still not significant.

4.5 Area and Wiring Overhead

Here we compare the area and wiring overhead of the SWNoC compared to a wireline mesh in order to deliver the high gains in performance. For the sake of comparison we also present the overheads required by the hierarchical wireless NoC, WiNoC. The silicon area overhead in the SWNoC arises due to the transceivers and the CNT antennas. However, the CNT antennas have very small diameters and hence, have negligible area overheads. Figure 7(a) quantifies total area overheads of the SWNoC compared to a flat wireline mesh for a 64 core system. It may be noted that WiNoC has much higher overheads as the hubs in each subnet of the WiNoC consume a lot of area as they have a large number of ports to all the cores in the subnet. Figure 7(b) shows the additional wiring requirements of the SWNoC with 64 cores. Due to the power-law distribution of the interconnects there are a few long wires in the SWNoC architecture compared to the mesh however, most of the interconnects are very short in length. Moreover, the total number of wires is the same as the average number of links attached to a switch was assumed to be the same as that of a mesh for the 64 core system. The WiNoC has more wires because each core in the subnets has direct connections to the hubs.

Figure 7 shows that due to the small-world network architecture adopted in the SWNoC, it achieves considerable improvements in performance even with very minimal silicon real estate and wiring
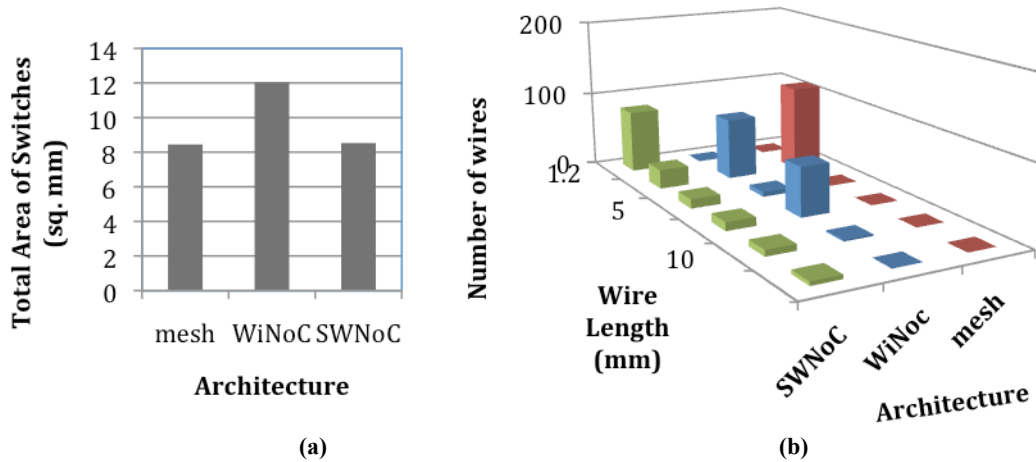
Figure 7. (a) Area and (b) wiring requirements of 64 core wireline mesh, WiNoC and SWNoC

overheads in excess of that already required by a flat wireline mesh. This indicates that the small-world network is a very efficient network topology with very small average distance between cores.

5. CONCLUSION

A natural complex network inspired small-world architecture where the long-range links are implemented through single-hop wireless channels improves the performance and energy efficiency of multi-core chips. Due to the inherent robustness of the small world based connectivity the proposed wireless NoC architecture outperforms its wireline counterparts even in presence of very high rates of wireless link failures without a significant penalty in energy dissipation. The principal contribution of this work is the demonstration of how architectural innovations can make NoC architectures robust and energy efficient even when designed with an emerging technology like on-chip wireless communication.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] L. Benini and G. D. Micheli, "Networks on Chips: A New SoC Paradigm," IEEE Computer, Vol. 35, Issue 1, January 2002, pp. 70-78.

[2] U. Ogras and R. Marculescu, "It's a Small World After All: NoC Performance Optimization Via Long-Range Link Insertion", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 14, No. 7, July 2006, pp. 693-706.

[3] V. F. Pavlidis and E. G. Friedman, "3-D Topologies for Networks-on-Chip," IEEE Transactions on Very Large Scale Integration (VLSI), Vol. 15, Issue 10, October 2007, pp. 1081-1090.

[4] A. Shacham et al., "Photonic Network-on-Chip for Future Generations of Chip Multi-Processors," IEEE Transactions on Computers, Vol. 57, no. 9, 2008, pp. 1246-1260.

[5] M. F. Chang et al., "CMP Network-on-Chip Overlaid With Multi-Band RF-Interconnect," Proc. of IEEE International Symposium on High-Performance Computer Architecture (HPCA), 16-20 February, 2008, pp. 191-202.

[6] R. Albert, et al., "Error and Attack Tolerance of Complex Networks", Nature, Vol. 406, July 2000, pp. 378-382.

[7] T. Peterman and P. D. L. Rios, "Spatial small-world networks: a wiring cost perspective", e-print arXiv:cond-mat/0501420.

[8] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature 393, pp. 440–442, 1998.

[9] A. Kumar et al., "Toward Ideal On-Chip Communication Using Express Virtual Channels," IEEE Micro, Vol. 28, Issue 1, January-February 2008, pp. 80-90

[10] D. Zhao and Y. Wang, "SD-MAC: Design and Synthesis of A Hardware-Efficient Collision-Free QoS-Aware MAC Protocol for Wireless Network-on-Chip," IEEE Transactions on Computers, vol. 57, no. 9, September 2008, pp. 1230-1245.

[11] S. B. Lee et al., "A Scalable Micro Wireless Interconnect Structure for CMPs", Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MobiCom), 20-25 September, 2009.

[12] K. Kempa, et al., "Carbon Nanotubes as Optical Antennae," Advanced Materials, vol. 19, 2007, pp. 421-426.

[13] A. Ganguly, et al., "Scalable Hybrid Wireless Network-on-Chip Architectures for Multi-Core Systems", IEEE Transactions on Computers (TC), August, 2010, DOI:
http://www.computer.org/portal/web/csdl/doi/10.1109/TC.2010.176.

[14] P. Bogdan and Radu Marculescu, "Quantum-Like Effects in Network-on-Chip Buffers Behavior," Proc. of IEEE Design Automation Conference, DAC, 4-8 June, 2007, pp. 266-267.

[15] J. Lin et al., "Communication Using Antennas Fabricated in Silicon Integrated Circuits," IEEE Journal of Solid-State Circuits, vol. 42, no. 8, August 2007, pp. 1678-1687.

[16] G. W. Hanson, "On the Applicability of the Surface Impedance Integral Equation for Optical and Near Infrared Copper Dipole Antennas," IEEE Transactions on Antennas and Propagation, vol. 54, no. 12, December 2006, pp. 3677-3685.

[17] P. J. Burke et al., "Quantitative Theory of Nanowire and Nanotube Antenna Performance," IEEE Transactions on Nanotechnology, Vol. 5, No. 4, July 2006, pp. 314-334.

[18] Y. Huang et al., "Performance Prediction of Carbon Nanotube Bundle Dipole Antennas," IEEE Transactions on Nanotechnology, Vol. 7, No. 3, May 2008, pp. 331-337.

[19] T.S. Marinis, et. al., "Wafer level vacuum packaging of MEMS sensors," Proc. of Electronic Components and Technology Conference, 2005. 31 May-3 June 2005, Vol. 2, pp.1081 - 1088.

[20] Y. Zhou et al., "Design and Fabrication of Microheaters for Localized Carbon Nanotube Growth", Proc. of IEEE conference on Nanotechnology, 2008, pp. 452-455.

[21] B.G. Lee et al., "Ultrahigh-Bandwidth Silicon Photonic Nanowire Waveguides for On-Chip Networks," IEEE Photonics Technology Letters, vol. 20, no. 6, Mar. 2008, pp. 398-400.

[22] W. M. J. Green et. al., "Ultra-compact, low RF power, 10Gb/s silicon Mach-Zehnder modulator," Optics Express, Vol. 15, No.25, pp. 17106-17113.

[23] A. Ismail and A. Abidi, "A 3 to 10GHz LNA Using a Wideband LC-ladder Matching Network," Proc. of IEEE International Solid-State Circuits Conference, 15-19 February, 2004, pp. 384-534.