

10-2012

Wireless Networks-on-Chips: Architecture, Wireless Channel, and Devices

David W. Matolak

University of South Carolina - Columbia, matolak@cec.sc.edu

Avinash Kodi

Savas Kaya

Dominic DiTomaso

Soumyasanta Laha

See next page for additional authors

Follow this and additional works at: https://scholarcommons.sc.edu/elct_facpub



Part of the [Digital Circuits Commons](#), [Digital Communications and Networking Commons](#), and the [Systems and Communications Commons](#)

Publication Info

Postprint version. Published in *IEEE Wireless Communications*, Volume 19, Issue 5, 2012, pages 58-65.

This Article is brought to you by the Electrical Engineering, Department of at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

Author(s)

David W. Matolak, Avinash Kodi, Savas Kaya, Dominic DiTomaso, Soumyasanta Laha, and William Rayess

WIRELESS NETWORKS-ON-CHIPS: ARCHITECTURE, WIRELESS CHANNEL, AND DEVICES

DAVID W. MATOLAK, AVINASH KODI, SAVAS KAYA, DOMINIC DITOMASO,
SOUMYASANTA LAHA, AND WILLIAM RAYESS, OHIO UNIVERSITY

ABSTRACT

Wireless networks-on-chips (WNoCs) hold substantial promise for enhancing multicore integrated circuit performance, by augmenting conventional wired interconnects. As the number of cores per IC grows, intercore communication requirements will also grow, and WNoCs can be used to both save power and reduce latency. In this article, we briefly describe some of the key challenges with WNoC implementation, and also describe our example design, iWISE, which is a scalable wireless interconnect design. We show that the integration of wireless interconnects with wired interconnects in NoCs can reduce overall network power by 34 percent while achieving a speedup of 2.54 on real applications.

INTRODUCTION

As the size and complexity of integrated circuits (ICs) continues to grow, new design and implementation challenges arise. These circuits (chips) are even now highly sophisticated systems, built from multiple subsystems and processors (cores) that are interconnected to form miniature networks. Depending on the architecture and the tasks the system is conducting, large amounts of data must be moved among various processors. With current technologies, this is implemented with traditional wired lines. As the number of processors and data exchanges grows, these wired lines are becoming insufficient to handle the required data rates, and may form “bottlenecks” to system performance. Wire lines also dissipate considerable power as their dimensions decrease to accommodate a larger number of wired interconnections among the many processors.

Thus, to alleviate the long wire delays and high power consumption of future multicore computer ICs, many prototypes and commercial designs are using network-on-chip (NoC) packet switching architectures. Wireless interconnects can improve NoCs by reducing the power dissipation of long “global” wires while providing high-bandwidth and low-latency communication [1, 2]. Wireless interconnects can provide some unique benefits including:

- Reduced power dissipation by avoiding multi-hop communication as in traditional metallic interconnects
- Reduced IC area overhead (fewer wires, waveguides) and lower parasitics
- Reuse of complementary metal oxide semiconductor (CMOS) wireless transceiver device designs

Wireless technologies have an advantage of being a mature form of communication with many well-known applications implemented in wireless local area networks, cell phones, and so on. This existing knowledge in the wireless/radio frequency (RF) field will facilitate the integration of wireless interconnects for NoCs, or WNoCs. Yet even with the relative maturity of wireless communication technologies, scaling these to very small sizes while concurrently scaling data rates to multiples of tens of Gb/s presents significant challenges in multiple areas, including network architecture, wireless propagation modeling and antennas, and low-power circuit and device design.

Some recent work has proposed wireless technologies to improve NoC performance. Lee *et al.* [3] proposed a two-tier (hierarchical) NoC design called WCube using a wired grid on one tier and centralized wireless hubs on the second tier. Another recent hybrid design proposed in [1] used several centralized wireless hubs connected in a ring. Although these designs provide low power and low latency solutions, numerous simplifying assumptions were made, and the design of a practical transceiver remains a significant challenge. The work in [4] proposed iWISE (inter-router Wireless Scalable Express Channel), a wireless-wired hybrid design for a large number of cores, by distributing wireless hubs across the network and using a token sharing scheme. Due to space limitations our literature citations must be brief.

In this article we briefly survey the challenges encountered in WNoCs and identify advantages and disadvantages of various options in their implementation. This entails choices in terms of frequency bands, network architecture and multiple access, and antenna and device design. After this survey, we describe our example WNoC design, a hybrid (i.e., wired and wire-

Technology/ Design Area	Frequency Band		
	50–150 GHz	150–500 GHz	500 GHz–3 THz
Circuits, Devices	Status: currently feasible Technology: RF-CMOS, substrate SOI	Status: encouraging Technology: SiGe-BiCMOS, substrate SOI	Status: immature Technology: III-V/Si hybrid, substrate alumina
Antennas, Propagation	Status: challenging Issues: electrically-small (inefficient) antennas, near field coupling	Status: challenging Issues: nearing conventional antennas, far-field conditions	Status: reasonable Issues: at highest f 's, propagation analysis conventional, antennas immature
System, Architecture	Issues: throughputs low to-moderate, SDM very difficult Area: low-Q inductors, antenna size Power: manageable	Issues: sufficient throughput, SDM challenging Area: very lossy substrates, ultra-low Q Power: challenging	Issues: ample throughput, SDM possible Area: limited by waveguides and-sources Power: very challenging

Table 1. WINOC system and technology challenges for three potential frequency bands.

less) iWISE NoC architecture for current chip multiprocessors (CMPs). The iWISE design is a wireless, low-power, and area-efficient NoC design that when scaled can provide enhanced throughput for future multicore architectures. Our wireless interconnects are designed for the 64-core version of iWISE, which represents current CMPs. Based on current RF CMOS technology, our design uses an ultra-low-power on-off keying (OOK) transceiver in the mmwave frequency range. Although not all technologies for our WINoC design currently exist, with future developments in areas such as carbon-based electronic materials and nanostructures, much of our transceiver design could be implemented; hence, our design is aimed to be illustrative rather than a present-day nanonetwork. The iWISE multiple access (MA) approach employs both time and frequency division to offer efficient and flexible sharing of the wireless links. Using current 32 nm CMOS technology, our network can reduce the dissipated network traffic power by up to 34 percent when compared to leading NoCs. The performance of our network using the real application benchmark suite, SPLASH-2 [5], shows an average speedup by a factor of 2.54 over popular alternative (wired) topologies such as the Flattened Butterfly (FB) and mesh.

The remainder of the article is structured as follows. We provide our survey of WINoC challenges. We describe the iWISE architecture and MA, and we summarize features of the wireless channel. We discuss transceiver device design and illustrate WINoC performance. Conclusions are then given.

WINOC CHALLENGES

In order to ensure that wireless links truly enhance NoC performance, they must:

- Provide high throughputs (e.g., tens of gigabits per second)
- Employ power- and area-efficient transceivers
- Employ efficient MA across the shared spatial channel

Providing tens of gigabits per second among multiple cores is nontrivial; this is particularly

true when frequency spectrum is limited. Although link distances are very short, wireless transceiver power dissipation must be minimized, and in the low mmwave frequency range, antennas will be inefficient due to their small electrical size. These large data rates also challenge circuit design, as most digital circuits cannot currently operate at these rates, and required serial-parallel conversions may introduce unacceptable overhead in power and complexity, so very simple modulation/demodulation schemes may be required. When spectrum is limited, time and frequency division must be used to allow sharing of the wireless medium. Spatial-division multiplexing (SDM) could provide welcome spatial reuse of time-frequency resources, but this is extraordinarily challenging at mmwave frequencies at present.

Thus, trades among various options in devices, modulations, and MA must be made, and for this it is of interest to look at frequency bands higher than the mmwave bands. Increasing the carrier frequencies of course introduces other challenges. In Table 1 we provide a summary of these considerations as a function of frequency band. We consider circuits/devices, antennas/propagation, and system/architecture. No clear optimum is evident, although selecting the “middle ground” frequency band of 150–500 GHz may allow satisfaction of the largest number of criteria. Note that there are several research groups focusing on developing wireless RF solutions for on-chip communication. Their solutions typically tackle a single design related to architecture, channel, or devices at a given frequency band, but not all three areas. Hence, other than the single-channel design in [6], there has been no convincing and complete implementable solution to date for WINoC architecture due to challenges highlighted in Table 1.

Finally, in this brief survey, we note that the actual wireless propagation channel will need to be characterized carefully for a given NoC “landscape.” This extremely complex 3D landscape constitutes the physical, multilayer, dielectric, and conductor environment through which the wireless signals will propagate, and will include various dielectric constants and loss tangents.

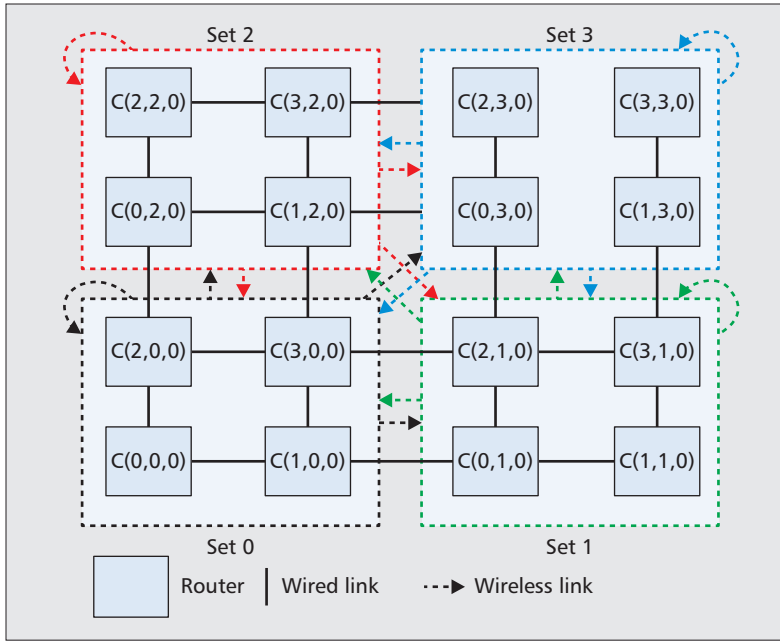


Figure 1. *iWISE* with $C = 4$ and $S = 4$ ($G = 1$ group), showing the wireless communication between sets.

Although there will be no motion and hence the channel will be largely time invariant,¹ there will likely be multipath propagation, which could limit performance.

IWIS ARCHITECTURE AND MULTIPLE ACCESS

The *iWISE* architecture is a scalable, wireless hybrid NoC. The architecture is separated into hierarchical subsections that define the communication protocol as shown in Fig. 1. Four cores ($N = 4$) are concentrated into (wired to) one cluster, and each cluster has its own data packet router. Routers are the nodes that are connected by the links (wired or wireless). The function of the router is to move packets from source to destination. Routers consist of buffers that store data, crossbars that switch or move data, and wireless transceivers. The four-core cluster has been shown [4] to be an effective design to reduce the router area overhead as well as serialization latency.

Our *iWISE* MA scheme uses both time and frequency division to enable wireless transmission from any core in any cluster to any other core in any other non-adjacent cluster (adjacent-cluster communication is wired). Clusters are grouped into sets. Set-to-set duplexing is via frequency division, and transmit multiplexing employs time division; this ensures single-carrier wireless transmission by any wireless modulator. The use of both wired and wireless communications provides efficiency and flexibility at the expense of a slight increase in MA complexity. The discussion here pertains to wireless communication among cores for the 64-core architecture. The hierarchical structure we use groups clusters by fours into sets, and a collection of four sets constitutes a group. Figure 1 illustrates the organization, where the solid lines without

arrows connecting adjacent clusters denote wired transmissions.

We identify clusters as $C(c, s, g)$, where c = cluster, s = set, and g = group, with $c \in \{0, 1, \dots, C-1\}$, $s \in \{0, 1, \dots, S-1\}$, and $g \in \{0, 1, \dots, G-1\}$. For the 64-core design, $C = S = 4$, and $G = 1$; that is, $C = 4$ clusters per set and $S = 4$ sets per group. Hence, the total number of cores in *iWISE* is the product $N \times C \times S \times G$. The “partial token” MA scheme allows each cluster within a set the opportunity to transmit to each of the S sets using S frequency channels — this is done over C consecutive time frames (single-carrier transmission). For a time frame duration T_f , C consecutive frames represent a time-division multiplexed cycle of duration $T_c = CT_f$. During a time frame each of the C clusters within a set can transmit to a unique set using a set-to-set-specific frequency channel. Duplexing is complicated by intra-set communication, which is wireless for non-adjacent (“diagonal”) clusters and wired for adjacent (horizontal or vertical) clusters. All transmissions are orthogonal.

Figure 2 illustrates the Set-0-transmission portion of the time-frequency (TF) plane for an example allocation (conceptual SDM is also indicated via two TF planes in two angular directions). Figure 2 applies to a uniform traffic demand, in which each cluster has data to transmit on each frame of each cycle. The other sets employ analogous blocks of four frequency channels of the same bandwidth, frame, and cycle times. Notation $C(c, s, g) \rightarrow S_i$ denotes transmission from the specified cluster to set S_i . Transmission control is regulated by tokens. Each cluster has a length- S token vector denoted \mathbf{y} , with elements that are either zero or one, with one denoting the index of the set to which the cluster may transmit during that frame; only one of the elements in each cluster’s \mathbf{y} is one. To clarify via example, for cluster 2 in set 0 transmitting to set 0 ($C(2,0,0) \rightarrow S_0$, on f_{00} , at $2T_f \leq t \leq 3T_f$) we have $\mathbf{y}_{2,0,0} = [1, 0, 0, 0]$; for this cluster transmitting to set 1 ($C(2,0,0) \rightarrow S_1$) $\mathbf{y}_{2,0,0} = [0, 1, 0, 0]$, cluster 2 transmitting to set 2 has $\mathbf{y}_{2,0,0} = [0, 0, 1, 0]$, and cluster 2 transmitting to set 3 has $\mathbf{y}_{2,0,0} = [0, 0, 0, 1]$. Table 2 defines the notations used in Fig. 2 for our orthogonal TF MA scheme. Practical research issues include adjacent-channel interference (ACI) from imperfect filtering, and co-channel interference when larger throughputs force TF reuse. To illustrate the potential of WiNOCs, our initial estimates of performance assume sharp filtering plus guard bands to limit ACI.

For the uniform TF allocation, with each frequency channel’s bandwidth B Hz, with binary modulation the channel’s data rate $R_b \cong B$ b/s; hence, with S^2 channels the aggregate throughput is BS^2 . In this orthogonal MA example scheme, any cluster’s data rate to any other cluster (intra- or intersets) has a maximum value of $R_b/S = R_b/4$. Our initial approach has each frame carry 20 bits.

For non-uniform traffic requirements, the MA scheme must be modified. For example, if $C(2,0,0)$ in Fig. 2 has no data to transmit to set S_1 on f_{01} during frame time $T_f < t < 2T_f$, it may transmit via wire to either adjacent cluster $C(0,0,0)$ or $C(3,0,0)$. During this same time

¹ Note that some time variation due to thermal effects as the chip heats could be present — this also complicates channel modeling.

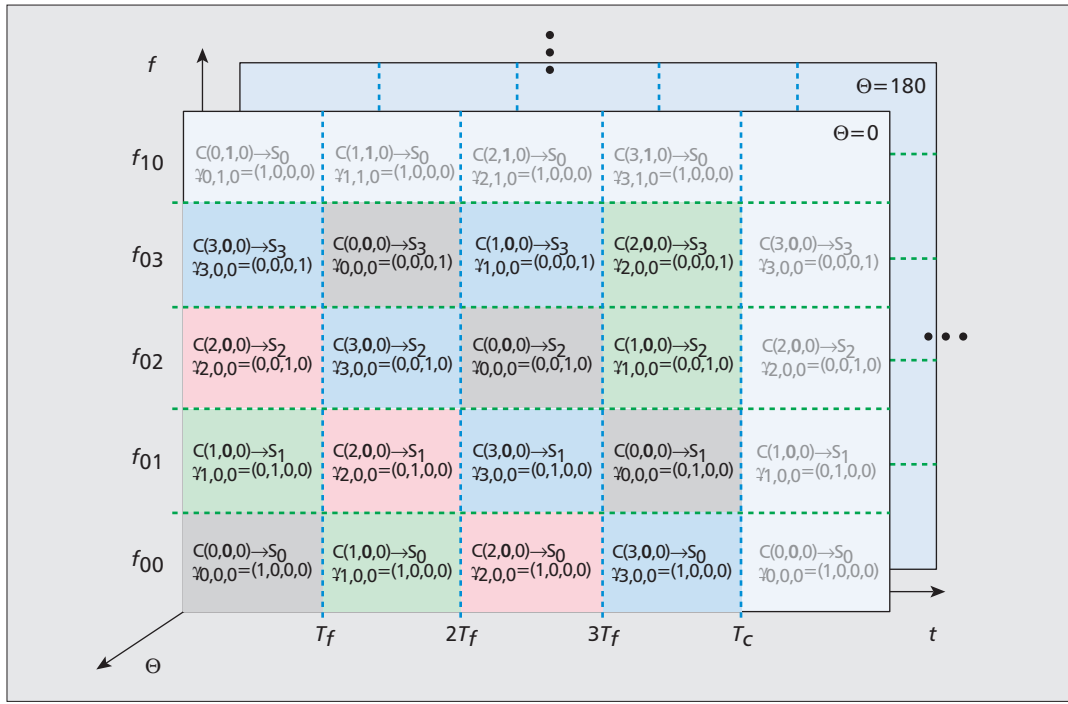


Figure 2. Time-frequency allocation for set 0 transmissions, equal data rate per cluster case.

frame, if any of the other clusters in set 0 is not transmitting wirelessly on another frequency channel, it may use f_{01} to transmit to set 1. This requires frequency-adjustable modulators. Additional adjustments to MA assignments can also be made; for brevity these are not detailed here, but note that if the network controller knows sufficiently in advance which specific wired transmissions are scheduled to take place, the f_{ii} time-frequency slots can be freed for use by any other cluster, enhancing network efficiency. If each cluster has *multiple* wireless modems, multi-frequency transmission by any cluster is possible, at the expense of replicating the modem hardware. The “network controller” is implemented using the tokens described earlier. Tokens are continuously circulating between the same set of clusters and captured when needed. The only control logic required is local to each router, which simply reads the token value, then passes the token to the next cluster. This control logic does not need to communicate with any other cluster. The tokens use their own separate side channels and do not affect the transmission of packets. Since only four tokens circulate between clusters in each set, each token can be mapped with 2 bits. Since this is one-tenth our packet size and tokens only circulate between neighboring clusters, area and power overhead is minimal.

CHANNEL MODELING

First, we note that it is not possible to obtain any truly accurate channel models without precise specification of the physical “landscape” of the WiNoC. The landscape is defined by the dimensions and electrical properties (conductivity σ , permittivity ϵ , and permeability μ) of all objects/surfaces in the environment through which the electromagnetic waves from transmit-

Parameter	Definition
f_{ij}	frequency channel for transmissions from set i to set j ; $i, j \in \{0, 1, \dots, S-1\}$; thus S^2 total frequency channels, each of B Hz
$\gamma_{c,s,g}$	$[\gamma_{csg,0} : \gamma_{csg,1} : \gamma_{csg,2} : \gamma_{csg,3}]$ = token vector for cluster c , set s , group g ; $\gamma_{csg,i} \in \{0, 1\}$; $\gamma_{csg,i} = 1$ is an indicator that transmission from cluster $C(c,s,g)$ is to set i

Table 2. MA parameter definitions.

ter to receiver propagate. This landscape could ultimately be quite complex, rendering accurate analysis of the channel impossible without sophisticated computer computations. At present we do not define the landscape in detail, and our initial channel model simply employs estimates. Channel modeling is ongoing work.

By “channel model,” we primarily mean the attenuation and delay dispersion characteristics, over our frequency band(s) of interest, over expected link distances. Antennas are often excluded from many channel models, but in the tens-to-hundreds of gigahertz or terahertz frequency ranges, we will ultimately need to incorporate antenna characteristics as well. Novel approaches such as carbon nanotubes, semiconductor nanowires, or even metamaterials may be required to keep antenna sizes small with acceptable radiation efficiencies. Electromagnetic compatibility is also an issue for future study.

Since neither the transmitter (Tx) or receiver (Rx) are moving, there will be no fading (i.e., each channel is time-invariant). Nonetheless, particularly with conducting surfaces present, reflections will yield multipath propagation, and this will yield spatial variation of field strength

If the performance degradation (in bit error ratio, BER) is significant enough, we may need to explore remedies via signal processing and/or multiple access re-design (e.g., reducing channel bandwidth) and/or physical link re-design (e.g., adding spatial suppression — directive antennas — to suppress MPCs).

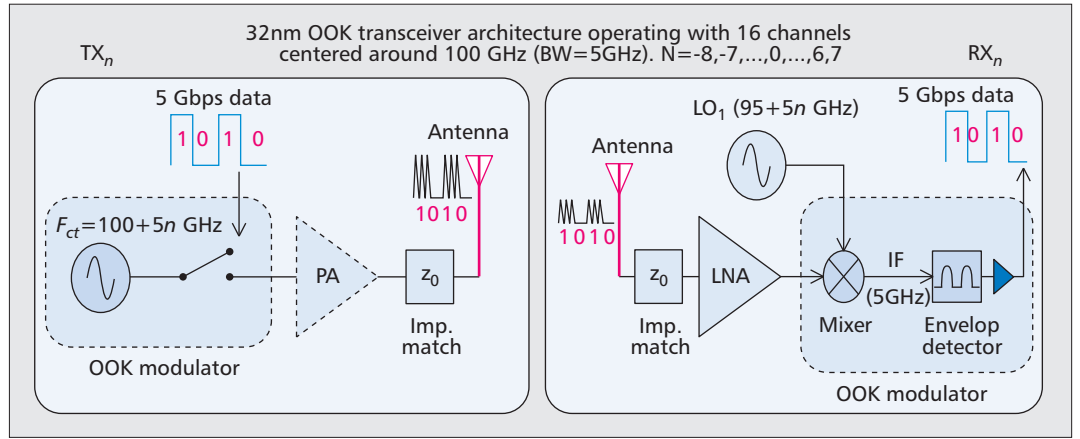


Figure 3. A generic low-power OOK transceiver design, operating at 5 Gb/s. Sixteen such channels are operated in adjacent bands.

unique to each Tx-Rx pair. For the high data rates we are targeting (≥ 10 Gb/s), even small amounts of dispersion can be performance limiting (e.g., delay spreads on the order of tens of picoseconds can cause distortion). A 10 ps delay difference corresponds to a path length difference of 3 mm in air.

For our initial estimates of attenuation (or propagation path loss), we employ simulation results in [7], where attenuation in dB can in effect be modeled using a log-distance formula (common in terrestrial communications), specifically, $A(d) = 19.8 \log_{10}(d/d_0) + A_0$, where $A(d)$ is the attenuation at distance d in units of centimeters, $d_0 = 1$ cm, and A_0 is a constant dependent upon material. This equation pertains to frequencies of several hundred gigahertz, and distances 0.5 mm to 5 cm. The path loss exponent here is $n = 1.98$ — essentially the free-space value of 2. For antennas within polyamide, $A_0 = 44.6$ dB, and within silicon, $A_0 = 79.6$ dB. For our architecture, the maximum distance is 21 mm (assuming a core to be 2.5 mm on a side and inter-router spacing of 5 mm), yielding attenuations of 11.4 and 36.4 dB. This model presupposes far-field conditions. Reference [1] assumes a vacuumed-out chamber environment for the WINoC, with a ground plane, and employs the well-known 2-ray model, which gives a path loss exponent $n = 4$ beyond a “break distance.” Vacuuming (a complication) may not be needed, and the 2-ray model may often be a gross simplification.

The actual channel will be three-dimensional, but the height h should be typically small compared to d , the length of the (square) IC, e.g., we might have $h < d/10$ or $h < d/100$. The non-uniformity of the landscape will complicate channel impulse response (CIR) estimation. For example, the landscape may include multiple dielectric (and/or conducting) layers, “steps” and “plateaus,” irregular geometric shapes due to fabrication imperfections, etc. Thus for the non-homogeneous and non-isotropic WINoC landscape, even an empirical model like that for $A(d)$ above will depend upon orientation and location within the WINoC. In other words, several models of this form (or at least a worst-case model) would likely be needed in practice to

characterize the spatial variation of path loss across the landscape.

A model for delay dispersion is even more difficult to estimate. The 2-ray model is the simplest model for a multipath channel. Determining the number of multipath components (MPCs) and their amplitudes may require full-wave electromagnetic field analysis; depending on the landscape and data rates, only the Tx-Rx CIR with the largest delay spread (worst case) may need to be estimated. For bandwidths beyond a few gigahertz, the MPCs themselves may be frequency-dependent (although this may be moderate). For the time-invariant WINoC, statistical measures such as the root-mean square delay spread (RMS-DS) — common for mobile channels — may not be optimal. The use of the “longest” CIR yields the maximum amount of delay dispersion imposed on any communication signal. This in turn allows us to estimate the degradation caused by this dispersion. If the performance degradation (in bit error ratio, BER) is significant enough, we may need to explore remedies via signal processing and/or multiple access redesign (e.g., reducing channel bandwidth) and/or physical link redesign (e.g., adding spatial suppression — directive antennas — to suppress MPCs).

We have also estimated upper bounds on delay spread via known results for a reverberation chamber, in which we model the WINoC landscape as a “micro-chamber” surrounded by boundaries. In this case, very large delay spreads are found: the largest delay spreads, up to a few nanoseconds, occur at the lowest frequency, and for the largest cavity sizes. Delay spreads of nanoseconds would be severely performance-limiting, and this points to the need for more research in this crucial area.

WIRELESS TRANSCIVER DESIGN

Figure 3 shows our proposed Tx and Rx design. The bandwidth of each wireless link is $B \approx 5$ GHz (data rate $R_{bc} = 5$ Gb/s); we use an equal distribution of bandwidth for each set. Each set is allocated four unique wireless links of width B_c , for transmission to each of the four possible sets. This yields a total of 16 distinct wireless

² Note that separate impedance matching networks may be required for the wired and wireless connections to routers.

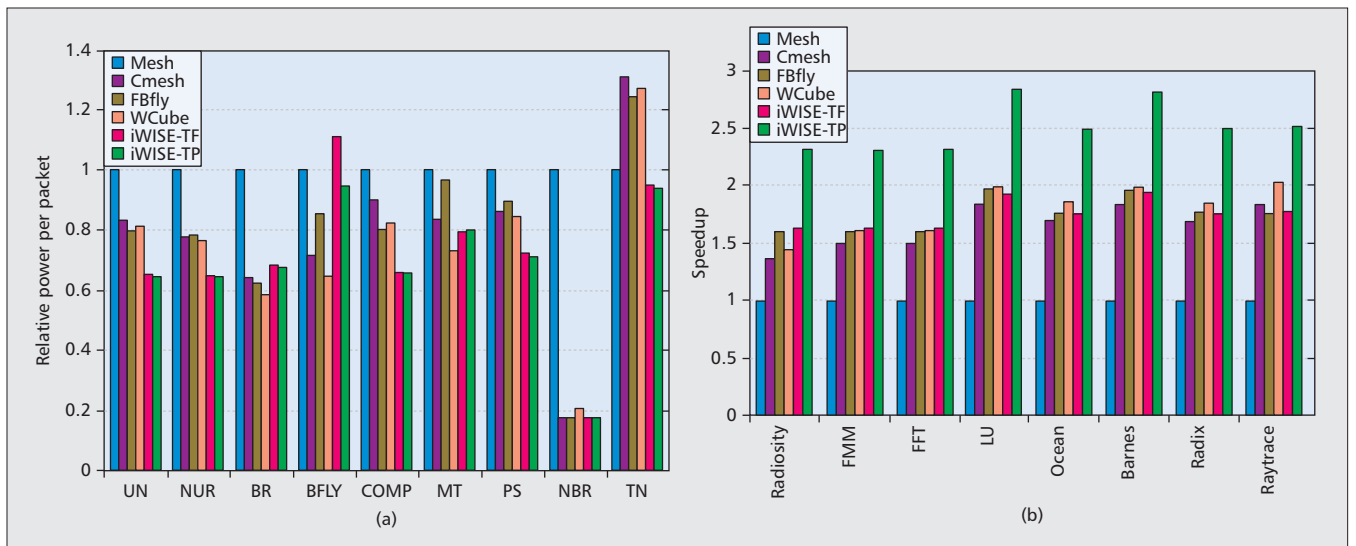


Figure 4. a) Power per packet of 64 core networks relative to mesh under different traffic loads; b) simulation speed-up relative to mesh for different Splash-2 applications.

links, comprising a total occupied bandwidth of approximately $W = 80$ GHz [3], plus 15 1-GHz guardbands. This is in general agreement with [8], where a bandwidth of 18 GHz at frequency 55 GHz was obtained. Local oscillators (LOs) generate the different carriers, and power amplifiers (PAs) and low-noise amplifiers (LNAs) amplify the outgoing and incoming signals. Impedance matching and filtering is used to transmit/receive the correct frequencies.² Each modem has 4 separate Tx and Rx antennas, tuned to a portion of the band: these are placed on a layer *above* the router substrate. These antennas are not required to cover the entire band of 95 GHz, but cover approximately 1/4 of that. Our current design is a scaling of that found in [9].

Modulation is on-off keying (OOK) and demodulation is non-coherent; the simplicity of this technique yields very low power consumption and ultra-compact architecture. Several design challenges exist for the iWISE OOK transceiver: common OOK circuits have been developed for low data rate ($\ll 1$ Gb/s) and low power sensor networks, where carrier frequency is low (< 2.4 GHz) and large on/off-chip antennas may be employed. OOK wireless links at 60 GHz, on the other hand, have been typically optimized for longer ($>> 10$ cm) range transmission, at much higher power consumption levels. We instead optimize an OOK architecture similar to [10] for ultra-low-power operation and ultra-short ~ 2 cm range, carrier frequency ~ 52 – 148 GHz, and on-chip antennas. Passive RF filtering is a nontrivial area of current research; guard bands mitigate ACI at the expense of system bandwidth. Space limitations here preclude a detailed link budget and description of filter design, but such filters will likely be based on compact resonators built with CMOS-compatible inductors on magnetic thin films, as well as novel designs of quarter-wavelength coplanar and/or micro-strip waveguides with through-silicon vias. At higher frequencies, resonant plasmonic structures and metamaterials (as

in the case of on-chip antenna designs) can be also used to resonate at the desired frequency or filter out unwanted bands. We assume 0 dB gain antennas based upon existing literature.

Combining optimistic technology scaling for a 32 nm RF silicon on insulator (SOI)-CMOS process, and the fact that signal levels can be further reduced due to shorter transmission distances (thereby conserving power and area), it will be possible to design OOK circuits with as low as 1 pJ/bit energy efficiency [1]. Recent on-chip transceiver designs have demonstrated energy efficiencies of 0.33 pJ/bit [1], 2 pJ/bit [11], and 4.5 pJ/bit [3], in the range of our predicted value of 1 pJ/bit. This energy can be further reduced with more efficient design techniques such as independently-driven double-gate (or FinFET) devices that allow very compact envelope detection and LO/voltage-controlled oscillator circuits, as well as efficient single-transistor mixers and tunable-gain PAs. Additionally, with the ultra-short WINoC communication ranges, it may even be possible to further reduce the power/size requirements of the LNA design and eliminate the PA and the impedance matching circuitry altogether using pulse driven compact antennas [12]. Moreover, an optimum use of ultra-fast SiGe BiCMOS devices could allow even higher data-rates in the baseband and LNA/PA circuitry, thus allowing a single OOK transceiver to deliver up to the near 32 Gb/s that will likely be required in future NoCs [13].

PERFORMANCE EVALUATION

Our wireless interconnects are implemented in the iWISE-64 topology and compared against electrical (wired) NoC designs including mesh, concentrated mesh (Cmesh), Flattened Butterfly (FBfly), and the wireless topology WCube [7]. We scale WCube to 64 cores by placing four wireless routers near the corners of a CMesh network. Commonly in NoC designs, a packet (frame) of data is divided into subsections called flits. Here we assume a packet contains four flits

³ Note, of course, that this complicates the transceiver design.

Since four links are shared in TP, a cluster may have to only wait four cycles for a token compared to 16 cycles in TF. WCube performs similar to Cmesh but worse than iWISE due to the combination of the few wireless links in WCube and the low network load of the benchmarks, which will leave the wireless link underutilized.

of size 5 bits. This is somewhat unrealistic since typical flit sizes are 64 or 128 bits; however, this is justified by the equally restricted wireless occupied bandwidth of 80 GHz. For a fair comparison, the data rates were adjusted such that the *bisectional bandwidth* (throughput) of each network was kept equal; the bisectional bandwidth in NoC design is defined as the bandwidth between two equal parts of a network. The MA scheme described earlier is called iWISE token-partial (TP). For an additional comparison, a less restrictive sharing scheme called iWISE token-full (TF) is also simulated. The iWISE-TF scheme ignores set organizations and simply shares the 16 links among the 16 clusters (a sort of “on-demand” MA).³ Token delays for iWISE-TP and iWISE-TF were accounted for by adding cycle delays. Since iWISE-64 is a one-hop architecture, there is no deadlock or livelock.

POWER DISSIPATION AND AREA OVERHEAD

For the 32 nm metal wired links, an energy of 0.18 pJ/bit for length 1 mm with a 1 GHz clock was used. For the wireless links, an energy of 0.8 pJ/bit at 5 mm was estimated for 32 nm CMOS technology using Synopsys HSPICE, and is within our predicted value of < 1 pJ/bit. The buffer and crossbar energy dissipations were estimated from Synopsys Design Compiler to be 0.011 pJ/bit and 0.108 pJ/bit, respectively. Circulation of the 2-bit tokens uses metal wired links, and power is calculated using the metal link energy above.

The power dissipation of the networks was calculated using a cycle accurate simulator running the following synthetic traffic loads on the 64-core NoC: uniform random (UN), non-uniform random (NUR), bit reversal (BR), butterfly (BFLY), complement (COMP), matrix transpose (MT), perfect shuffle (PS), neighbor (NBR), and Tornado (TN). During simulation, the number of link, buffer, and crossbar traversals was counted and the network power dissipation was calculated using their corresponding energy values. Gain tuning at the PA was used to linearly adjust the transmit power according to the distance between source and destination. The average power dissipation per packet is shown in Fig. 4a. Overall, iWISE-TP and TF save an average of 18 percent power. The power savings is largely dependent on how often the wireless links are used. In addition to the wireless links having a lower power than wired links, the long wireless transmissions allow packets to skip over intermediate routers, further lowering power. For UN traffic, 80 percent of the packets use wireless links, resulting in a power savings of approximately 35 percent. WCube consumes slightly less power than Cmesh but more power than iWISE. iWISE uses more wireless links than WCube, giving packets more opportunities to use these lower-energy links. With CMOS technology becoming smaller, the wireless interconnect circuitry will consume even less power.

The addition of the wireless links actually adds area overhead to the network. The wireless link area estimated from Synopsys was 0.094 mm² for 32 nm technology. This area is significantly larger than the wired link and router areas. However, this area overhead is the trade-

off encountered for the power savings cited previously and the speed-up addressed in the next section. However, the metal wire and router areas increase with the data packet size, but the wireless link area does not. Additionally, as CMOS technology continues to become smaller, so will the wireless transceiver. Furthermore, as previously noted, if transmission distances shrink enough, elimination of the LNA and/or PA and/or impedance matching circuitry will remove a major contribution to area. Finally, smaller transceivers might be designed using the latest developments in carbon-based electronic materials and nanostructures.

SPEED-UP

Speed-up is a measure of the reduction in time for the NoC to complete a task. The SPLASH-2 workloads are benchmarks that represent task applications for future CMPs. The workloads were run on the full execution-driven simulator SIMICS from Wind River. The communication traces of these workloads were extracted from the full system simulator and then executed on a cycle-accurate network simulator. The traces were run on the 64-core networks until execution was completed. The total number of clock cycles for each network to finish execution of each trace was recorded. The access time for the memory closest to the core (level 1 cache) was assumed to be 2 cycles, for the next closest memory (level 2 cache) 4 cycles, and for the main memory access time 160 cycles. On average iWISE-TP finished execution of the SPLASH-2 workloads 2.54 times faster than the other networks as shown in Fig. 4b. This speedup is due to the one-hop architecture of iWISE. Wireless links lower the latency of packets by transmitting long distances, allowing faster execution. iWISE-TP improves the execution time by 60 percent on average over the 16 hop mesh network and approximately 30 percent over the FBfly topology. The FBfly topology is only a two hop network, however, the long wired links cause high latency when compared to the faster wireless links. iWISE-TP outperforms iWISE-TF by approximately 30 percent due to the different token delays of each network. Since four links are shared in TP, a cluster may have to only wait four cycles for a token compared to 16 cycles in TF. WCube performs similar to Cmesh but worse than iWISE due to the combination of the few wireless links in WCube and the low network load of the benchmarks, which will leave the wireless link underutilized.

CONCLUSION

In this article, we reviewed some of the challenges faced in the design of WINoCs. We also provided a description of our proposed wireless interconnect for the hybrid iWISE NoC architecture that improves network performance and power. We use TDM and FDM to allocate wireless links to 4-core clusters to efficiently communicate between sets of clusters. Our transceiver for the wireless link uses an OOK technique to achieve an ultra-low power and compact design. Distributing these wireless links among routers and using our MA scheme creates a low power,

one-hop path for packets that reduces network power consumption compared to fully wired architectures. Although there are still significant challenges for WINoC implementation, in terms of channel modeling, transceiver devices, and architecture design, we believe that WINoCs show great promise for enhancing the performance of future multi-core ICs.

ACKNOWLEDGMENT

This research was supported by NSF awards, CCF-0915418, CCF-1054339 (CAREER) and ECCS-1129010

REFERENCES

- [1] A. Ganguly et al., "Scalable Hybrid Wireless Network-on-Chip Architectures for Multicore Systems," *IEEE Trans. Computers*, vol. 60, no. 10, Oct. 2011, pp. 1485–502.
- [2] R. Wu, Y. Wang, and D. Zhao, "A Low-Cost Deadlock-Free Design of Minimal-Table Rerouted xy-Routing for Irregular Wireless NoCs," *4th ACM/IEEE Int'l. Symp. Networks-on-Chip (NoCs)*, 2010, pp. 199–206.
- [3] S. B. Lee et al., "A Scalable Micro Wireless Interconnect Structure for CMPs," *Proc. 15th Ann. Int'l. Conf. Mobile Computing and Networking*, Beijing, China, 2009, pp. 217–28.
- [4] D. DiTomaso et al., "iWISE: Inter-Router Wireless Scalable Express Channels for Network-on-Chips (NoCs) Architecture," *19th Ann. IEEE Symp. High-Performance Interconnects*, Aug. 2011, pp. 11–18.
- [5] S. C. Woo et al., "The Splash-2 Programs: Characterization and Methodological Considerations," *ACM SIGARCH Computer Architecture News*, vol. 23, May 1995, pp. 24–36.
- [6] S. Deb et al., "Enhancing Performance of Network-on-Chip Architectures with Millimeter-Wave Wireless Interconnects," *Proc. IEEE Int'l. Conf. Application-Specific Systems, Architectures and Processors*, 7–9 July 2010, pp. 73–80.
- [7] S. Lee et al., "A Scalable Micro Wireless Interconnect Structure for CMPs," *Proc. MobiCom '09*, Beijing, China, 20–25 Sept. 2009.
- [8] X. Yu et al., "A Wideband Body-Enabled Millimeter Wave Transceiver for Wireless Network-on-Chip," *Proc. IEEE 54th Int'l. Midwest Symp. on Circuits and Systems (MWSCAS)*, Yonsei Univ., Seoul, Korea, Aug. 2011, pp. 1–4.
- [9] G. Singh, "Design Considerations for Rectangular Microstrip Patch Antenna on Electromagnetic Crystal Substrate at Terahertz Frequency," *Elsevier J. Infrared Physics and Technology*, vol. 53, pp. 17–22, 2010.
- [10] D. Daly and A. Chandrakasan, "An Energy-Efficient OOK Transceiver for Wireless Sensor Networks," *IEEE J. Solid-State Circuits*, vol. 42, no. 5, 2007, pp. 1003–11.
- [11] P. Y. Chiang et al., "Short-Range, Wireless Interconnect Within A Computing Chassis: Design Challenges," *IEEE Design and Test of Computers*, vol. 27, no. 4, July 2010, pp. 32–43.
- [12] S. D. Keller, W. D. Palmer, and W. T. Joines, "Digitally Driven Antenna for HF Transmission," *IEEE Trans. Microwave Theory and Techniques*, vol. 58, no. 9, 2010, pp. 2362–67.
- [13] S. Decoutere et al., "Advanced Process Modules and Architectures for Half-Terahertz SiGe:C HBTs," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, Oct. 2009, pp. 9–16.

BIOGRAPHIES

DAVID MATOLAK (matolak@cec.sc.edu) received his B.S. degree from Pennsylvania State University, University Park, his M.S. degree from the University of Massachusetts, Amherst, MA, and the Ph.D. degree from the University of Virginia, Charlottesville, all in electrical engineering. He has worked for over 20 years on communication systems, with the Rural Electrification Administration, Washington, DC, the UMass LAMMDA Laboratory, Amherst, AT&T Bell Laboratories, North Andover, Massachusetts, the University of Virginia's Communication Systems Laboratory, Lockheed Martin Tactical Communication Systems, Salt Lake City, Utah, the MITRE Corporation, McLean, Virginia, and Lockheed Martin Global Telecommunications, Reston, Virginia. From 1999 to August 2012 he was with the School of Electrical Engineering and Computer Science at Ohio University, and since August 2012 he has been with the Department of Electrical Engineering at the University of South Carolina. His research interests are communication over fading channels, radio channel modeling, and ad hoc networking.

AVINASH KARANTH KODI (kodi@ohio.edu) received Ph.D. and M.S. degrees in electrical and computer engineering from the University of Arizona, Tucson in 2006 and 2003, respectively. He is currently an associate professor with the Department of Electrical Engineering and Computer Science at Ohio University, Athens. He is the recipient of the National Science Foundation (NSF) CAREER award in 2011. His research interests include computer architecture, optical interconnects, chip multiprocessors (CMPs), and network-on-chips (NoCs).

SAVAS KAYA (kaya@ohio.edu) obtained his Ph.D. in 1998 from Imperial College of Science, Technology and Medicine, London, United Kingdom, for his work on strained Si quantum wells on vicinal substrates, following his M.Phil. in 1994 from the University of Cambridge. He was a post-doctoral researcher at the University of Glasgow between 1998 and 2001, carrying out research in transport and scaling in Si/SiGe MOSFETs, and fluctuation phenomena in decanano MOSFETs. He is currently with the Russ College of Engineering at Ohio University, Athens. His other interests include transport theory, device modeling and process integration, nanofabrication, nanostructures, and nanosensors.

DOMINIC DiTOMASO (dd292006@ohio.edu) received his B.S. and M.S. degrees in electrical engineering and computer science from Ohio University, Athens in 2010 and 2012. He is currently pursuing his Ph.D. degree in the Department of Electrical Engineering and Computer Science at Ohio University. His research interests include wireless interconnects, network-on-chips (NoCs), and computer architecture.

SOUMYASANTA LAHA (sl922608@ohio.edu) obtained his M.Sc. in embedded digital systems with distinction from the University of Sussex, United Kingdom, in 2007. Since 2008, he is with the Russ College of Engineering, Ohio University pursuing a Ph.D. in electrical engineering in the area of nanoscale energy-efficient RF transceivers. He also has more than three years of industrial work experience in India and the United Kingdom in embedded systems and analog electronics.

WILLIAM RAYESS (wr233608@ohio.edu) received his B.E. in computer and communications engineering from Notre Dame University in Lebanon in 2008, an MCTP from Ohio University in 2009, and is currently pursuing his Ph.D. in electrical engineering at the Russ College of Engineering, Ohio University.

Although there are still significant challenges for WINoC implementation, in terms of channel modeling, transceiver devices, and architecture design, we believe that WINoCs show great promise for enhancing the performance of future multicore ICs.