3-D Topologies for Networks-on-Chip

Vasilis F. Pavlidis, *Student Member, IEEE*, and Eby G. Friedman, *Fellow, IEEE*

Abstract—Several interesting topologies emerge by incorporating the third dimension in networks-on-chip (NoC). The speed and power consumption of 3-D NoC are compared to that of 2-D NoC. Physical constraints, such as the maximum number of planes that can be vertically stacked and the asymmetry between the horizontal and vertical communication channels of the network, are included in speed and power consumption models of these novel 3-D structures. An analytic model for the zero-load latency of each network that considers the effects of the topology on the performance of a 3-D NoC is developed. Tradeoffs between the number of nodes utilized in the third dimension, which reduces the average number of hops traversed by a packet, and the number of physical planes used to integrate the functional blocks of the network, which decreases the length of the communication channel, is evaluated for both the latency and power consumption of a network. A performance improvement of 40% and 36% and a decrease of 62% and 58% in power consumption is demonstrated for 3-D NoC as compared to a traditional 2-D NoC topology for a network size of $N = 128$ and $N = 256$ nodes, respectively.

Index Terms—3-D circuits, 3-D integrated circuits (ICs), 3-D integration, networks-on-chip (NoC), topologies.

I. INTRODUCTION

INTERCONNECT related problems, emerging from technology scaling and the integration limitations of systems-on-chip (SoC), originate from the functional diversity demanded by the electronics market. These issues have triggered a quest for nonconventional IC design paradigms, such as 3-D integration. For example, vertically stacked dies with through-silicon vias [1], together with networks-on-chip (NoC) have been proposed as potent solutions to address these interconnect problems and the design complexity of SoC. Each of these design paradigms offers unique opportunities.

The major advantage of 3-D ICs is the considerable reduction in the length and number of global interconnects, resulting in an increase in the performance and decrease in the power consumption and area of wire limited circuits [2], [3]. Another important advantage of 3-D ICs is that this paradigm enables the integration of CMOS circuits with disparate technologies which can

The authors are with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627 USA (e-mail: pavlidis@ece.rochester.edu; friedman@ece.rochester.edu).

be non-silicon or even electro-mechanical [4]. Despite the significant advantages of three-dimensional integration, important challenges remain such as crosstalk noise analysis and reduction, thermal mitigation, and interconnect modeling.

NoC offers high flexibility and the regularity of a network structure, supporting simpler interconnect models and greater fault tolerance. The canonical interconnect backbone of the network combined with appropriate communication protocols enhance the flexibility of such systems [5]. NoC can include a variety of functional intellectual property (IP) blocks or processing elements (PE) such as processor and DSP cores, memory blocks, FPGA blocks, and dedicated hardware, serving a plethora of applications including image processing, personal devices, mobile handsets, etc [6]–[8]. Note that the terms IP block and PE are interchangeably used in this paper. The intra-PE delay, however, cannot be reduced by the network. Furthermore, the length of the communication channel is primarily determined by the area of the PE, which is typically unaffected by the network structure. By merging these two approaches, many of the individual limitations of 3-D ICs and NoC are circumvented, yielding a robust design paradigm with unprecented capabilities.

Research in 3-D NoC is only now emerging [9]–[11]. Addo-Quaye [10] recently presented an algorithm for the thermal-aware mapping and placement of 3-D NoC including regular mesh topologies. Li *et al.* [11] proposed a similar 3-D NoC topology employing a buss structure for communicating among PEs located on different physical planes. Targeting multi-processor systems, the proposed scheme in [11] considerably reduces cache latencies by utilizing the third dimension. Multidimensional interconnection networks have been studied under various constraints, such as constant bisection-width and pin-out constraints [12]. NoC differ from generic interconnection networks, however, in that NoC are not limited by the channel width or pin-out. Alternatively, physical constraints specific to 3-D NoC, such as the number of nodes that can be implemented in the third dimension and the asymmetry in the length of the channels of the network, have to be considered. In this paper, various possible topologies for 3-D NoC are presented. Additionally, analytic models for the zero-load latency and the power consumption with delay constraints of these networks that capture the effects of the topology on the performance of 3-D NoC are described. Optimum topologies are shown to exist that minimize the zero-load latency and power consumption of a network. These optimum topologies depend upon a number of parameters characterizing both the router and the communication channel, such as the number of ports of the router, the length of the communication channel, and the impedance characteristics of the interconnect. Various
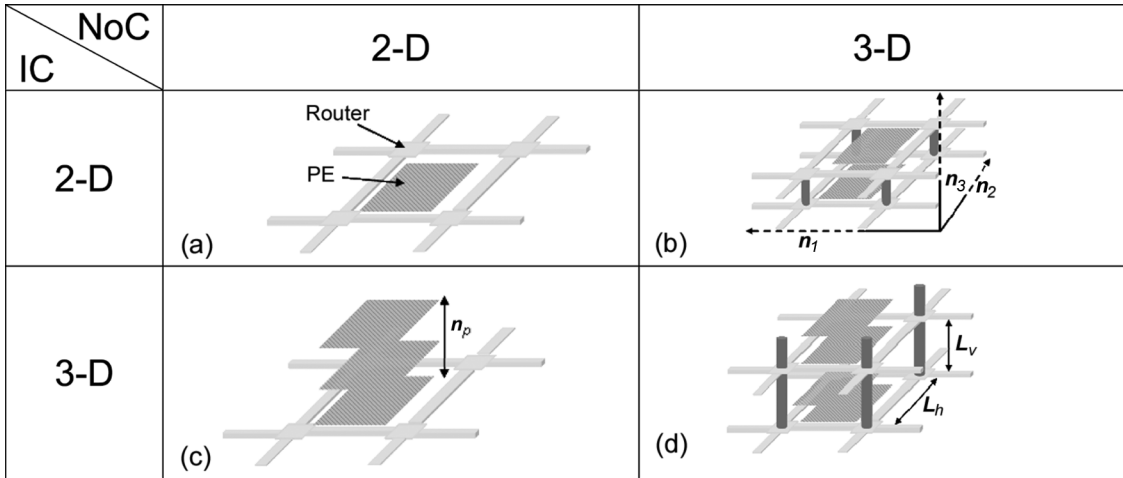
Fig. 1.   Various NoC topologies (not to scale). (a) 2-D IC–2-D NoC. (b) 2-D IC–3-D NoC. (c) 3-D IC–2-D NoC. (d) 3-D IC–3-D NoC.

tradeoffs among these parameters that determine the minimum latency and power consumption topology for a network are investigated for different network sizes.

The paper is organized as follows. In Section II, various topological choices for 3-D NoC are reviewed. In Section III, an analytic model of the zero-load latency of traditional interconnection networks is adapted for each of the proposed 3-D NoC topologies, while the power consumption model of these network topologies is described in Section IV. In Section V, the proposed 3-D NoC topologies are compared in terms of the zero-load network latency and power consumption with delay constraints, and guidelines for the optimum design of performance driven or power driven NoC structures are provided. Finally, some conclusions are offered in Section VI.

II. 3-D NoC TOPOLOGIES

Various topologies for 3-D networks are presented and related terminology is introduced in this section. Mesh structures have been a popular network topology for conventional 2-D NoC [13], [14]. A fundamental element of a mesh network is illustrated in Fig. 1(a), where each processing element (PE) is connected to the network through a router. A PE can be integrated either on a single physical plane (2-D IC) or on several physical planes (3-D IC). Each router in a 2-D NoC is connected to a neighboring router in one of four directions. Consequently, each router has five ports. Alternatively, in a 3-D NoC, the router typically connects to two additional neighboring routers located on the adjacent physical planes. The architecture of the router is considered here to be a canonical router with input and output buffering [15]. The combination of a PE and router is called a network node. For a 2-D mesh network, the total number of nodes N is $N = n_1 \times n_2$, where n_i is the number of nodes included in the ith physical dimension.

Integration in the third dimension introduces a variety of topological choices for NoCs. For a 3-D NoC as shown in Fig. 1(b), the total number of nodes is $N = n_1 \times n_2 \times n_3$, where n_3 is the number of nodes in the third dimension. In this topology, each PE is on a single yet possibly different physical plane (2-D IC–3-D NoC). In other words, a PE can

be implemented on only one of the n_3 physical planes of the system and, therefore, the 3-D system contains $n_1 \times n_2$ PEs on each one of the n_3 physical planes such that the total number of nodes is N. This topology is discussed in [10], [11]. A 3-D NoC topology is proposed in Fig. 1(c), where the interconnect network is contained within one physical plane (i.e., $n_3 = 1$), while each PE is integrated in multiple planes, notated as n_p (3-D IC–2-D NoC). Finally, a hybrid 3-D NoC based on the two previous topologies is proposed in Fig. 1(d). In such an NoC, both the interconnect network and the PEs can span more than one physical plane of the stack (3-D IC–3-D NoC). In the following section, latency expressions for each of the NoC topologies are described, assuming a zero-load model.

III. ZERO-LOAD LATENCY FOR 3-D NoC

In this section, analytic models of the zero-load latency of each of the 3-D NoC topologies are described. The zero-load network latency is widely used as a performance metric in traditional interconnection networks [16]. The zero-load latency of a network is the latency where only one packet traverses the network. Although such a model does not consider contention among packets, the zero-load latency can be used to describe the effect of a topology on the performance of a network. The zero-load latency of an NoC with wormhole switching is [16]

$$T_{\text{network}} = hops \cdot t_r + t_c + \frac{L_p}{b} \tag{1}$$

where the first term represents the routing delay, t_c is the propagation delay along the wires of the communication channel, which is also called a buss here for simplicity, and the third term is the serialization delay of the packet. $hops$ is the average number of routers that a packet traverses to reach the destination node, t_r is the router delay, L_p is the length of the packet in bits, and b is the bandwidth of the communication channel defined as $b = w_c f_c$, where w_c is the width of the channel in bits and f_c is the inverse of the propagation delay of a bit along the longest communication channel.

Since the number of planes that can be stacked in a 3-D NoC is constrained by the target technology, n_3 is also constrained. Furthermore, n_1, n_2, and n_3 are not necessarily equal. The average number of hops in a 3-D NoC is

$$hops = \frac{n_1 n_2 n_3 (n_1 + n_2 + n_3) - n_3(n_1 + n_2) - n_1 n_2}{3(n_1 n_2 n_3 - 1)} \quad (2)$$

assuming dimension-order routing such that the minimum distance paths are used for the routing of packets between any source–destination node pair.

The number of hops in (2) can be divided into two components, the average number of hops within the two dimensions n_1 and n_2, and the average number of hops within the third dimension n_3

$$hops_{2\text{-D}} = \frac{n_3 (n_1 + n_2)(n_1 n_2 - 1)}{3(n_1 n_2 n_3 - 1)} \quad (3)$$

$$hops_{3\text{-D}} = \frac{(n_3^2 - 1) n_1 n_2}{3(n_1 n_2 n_3 - 1)}. \quad (4)$$

The delay of the router t_r is the sum of the delay of the arbitration logic t_a and the delay of the switch t_s, which, in this paper, is considered to be implemented with a classic crossbar switch [16]

$$t_r = t_a + t_s. \quad (5)$$

The delay of the arbiter can be described from [17]

$$t_a = (21(1/4) \log_2 p + 14(1/12) + 9)\tau \quad (6)$$

where p is the number of ports of the router and τ is the delay of a minimum sized inverter for the target technology. Note that (6) exhibits a logarithmic dependence on the number of router ports. The length of the crossbar switch also depends upon the number of router ports and the width of the buss

$$l_s = 2(w_t + s_t)w_c p \quad (7)$$

where w_t and s_t are the width and spacing or, alternatively, the pitch of the interconnect, respectively, and w_c is the width of the communication channel in bits. Consequently, the worst case delay of the crossbar switch is determined by the longest path within the switch, which is equal to (7). The delay of the communication channel t_c is

$$t_c = t_v hops_{3\text{-D}} + t_h hops_{2\text{-D}} \quad (8)$$

where t_v and t_h are the delay of the vertical and horizontal channels, respectively [see Fig. 1(b)]. Note that if $n_3 = 1$, (8) describes the propagation delay of a 2-D NoC. Substituting (5) and (8) into (1), the overall zero-load network latency for a 3-D NoC is

$$T_{\text{network}} = hops(t_a + t_s) + hops_{2\text{-D}} t_h + hops_{3-\text{D}} t_v + \frac{L_p}{w_c} t_h. \quad (9)$$

To characterize t_s, t_h, and t_v, the models described in [18] are adopted, where repeaters implemented as simple inverters are inserted along the interconnect. According to these models,

the propagation delay and rise time of a single interconnect stage for a step input, respectively, are

$$t_{di} = 0.377 \frac{r_i c_i l_i^2}{k_i^2} + 0.693 \left(R_{d0} C_0 + \frac{R_{d0} c_i l_i}{h_i k_i} + \frac{r_i l_i C_{g0} h_i}{k_i} \right) \quad (10)$$

$$t_{ri} = 1.1 \frac{r_i c_i l_i^2}{k_i^2} + 2.75 \left(R_{r0} C_0 + \frac{R_{r0} c_i l_i}{h_i k_i} + \frac{r_i l_i C_{g0} h_i}{k_i} \right). \quad (11)$$

where r_i (c_i) is the per unit length resistance (capacitance) of the interconnect and l_i is the total length of the interconnect. The index i is used to notate the various interconnect delays included in the network (i.e., $i \in \{s, v, h\}$). h_i and k_i denote the number and size of the repeaters, respectively, and C_{g0} and C_0 represent the gate and total input capacitance of a minimum sized device, respectively. C_0 is the summation of the gate and drain capacitance of the device. R_{r0} and R_{d0} describe an equivalent output resistance of a minimum sized device for the propagation delay and transition time of a minimum sized inverter, respectively, where the output resistance is approximated as

$$R_{r(d)0} = K_{r(d)} \frac{V_{dd}}{I_{dn0}}. \quad (12)$$

K denotes a fitting coefficient and I_{dn0} is the drain current of an nMOS device at both V_{ds} and V_{gs} equal to V_{dd}. The value of these device parameters are listed in Table I. A 45-nm technology node is assumed and SPICE simulations of the predictive technology library are used to determine the individual parameters [19], [20].

To include the effect of the input slew rate on the total delay of an interconnect stage, (10) and (11) are further refined by including an additional coefficient γ as in [21]

$$\gamma_r = \frac{1}{2} - \frac{1 - \frac{V_{tn}}{V_{dd}}}{1 + a_n}. \quad (13)$$

By substituting the subscript n with p, the corresponding value for a falling transition is obtained. The average value γ of γ_r and γ_f is used to determine the effect of the transition time on the interconnect delay. The overall interconnect delay can therefore be described as

$$\begin{aligned} t_i &= k_i(t_{di} + \gamma t_{ri}) \\ &= a_1 \frac{r_i c_i l_i^2}{k_i} + a_2 \left(R_0 C_0 k_i + \frac{R_0 c_i l_i}{h_i} + r_i l_i C_{g0} h_i \right) \end{aligned} \quad (14)$$

where R_0, a_1, and a_2 are described in [22] and the index i denotes the various interconnect structures such as the crossbar switch $(i \equiv s)$, the horizontal buss $(i \equiv h)$, and the vertical buss $(i \equiv v)$.

For minimum delay, the size h_i and number k_i of repeaters are determined by setting the partial derivative of t_i with respect to h_i and k_i, respectively, equal to zero and solving for h_i and k_i

$$k_i^* = \sqrt{\frac{a_1 r_i c_i l_i^2}{a_2 R_0 C_0}} \quad (15)$$

$$h_i^* = \sqrt{\frac{R_0 c_i}{r_i C_{g0}}}. \quad (16)$$

TABLE I
INTERCONNECT AND DESIGN PARAMETERS, 45-NM TECHNOLOGY

Parameter	Value	
	NMOS	PMOS
W_{min}	100 nm	250 nm
I_{dsat}/W	1115 µA/µm	349 µA/µm
V_{dsat}	478 mV	-731 mV
V_t	257 mV	-192 mV
a	1.04	1.33
I_{sub0}	48.8 nA	
I_{g0}	0.6 nA	
V_{dd}	1.1 Volts	
Temp.	110 ℃	
K_d	0.98	
K_r	0.63	
C_{g0}	512 aF	
C_{d0}	487 aF	
τ	17 ps	

The expression in (14) only considers RC interconnects. An RC model is sufficiently accurate to characterize the delay of a crossbar switch since the length of the longest wire within the crossbar switch and the signal frequencies are such that inductive behavior is not prominent. For the buss lines, however, inductive behavior can appear. For this case, suitable expressions for the delay and repeater insertion characteristics can be adopted from [23]. For the target operating frequencies (1–2 GHz) and buss length (<2 mm) considered in this paper, an RC interconnect model provides sufficient accuracy [24]. Additionally, for the vertical buss, $k_v = 1$ and $h_v = 1$, meaning that no repeaters are inserted and minimum sized drivers are utilized. Repeaters are not necessary due to the short length of the vertical buss. Note that the proposed latency expression includes the effect of the input slew rate. Additionally, since a repeater insertion methodology for minimum latency is applied, any further reduction in latency is due to the network topology.

The length of the vertical communication channel for the 3-D NoC shown in Fig. 1 is

$$l_v = \begin{cases} L_v, & \text{for 2-D IC–3-D NoC} & \text{(17a)} \\ n_p L_v, & \text{for 3-D IC–3D NoC} & \text{(17b)} \\ 0, & \text{for 2-D IC–2-D NoC and 3-D IC–2–D NoC} & \text{(17c)} \end{cases}$$

where L_v is the length of a silicon-through (interplane) via connecting two routers on adjacent physical planes. n_p is the number of physical planes used to integrate each PE. The length of the horizontal communication channel is assumed to be

$$l_h = \begin{cases} \sqrt{A_{PE}}, & \text{for 2-D IC–2-D NoC and 2-D IC–3-D NoC} & \text{(18a)} \\ 1.12\sqrt{\dfrac{A_{PE}}{n_p}}, & \text{for 3-D IC–2-D NoC} \\ & \text{and 3-D IC–3-D NoC}(n_p > 1) & \text{(18b)} \end{cases}$$

where A_{PE} is the area of the processing element. The area of all of the PEs and, consequently, the length of each horizontal channel are assumed to be equal. For those cases where the PE is implemented in multiple physical planes, a coefficient is included to consider the effect of the interplane vias on the reduction in the ideal wirelength due to utilization of the third dimension. The value of this coefficient (1.12) is based on the layout

of a crossbar switch designed with the FDSOI 3-D technology from MIT Lincoln Laboratory (MITLL) [25]. The same coefficient is also assumed for the design of the PEs on more than one physical plane. In the following section, expressions for the power consumption of a network with delay constraints are presented.

IV. POWER CONSUMPTION IN 3-D NoC

Power dissipation is a critical issue in 3–D circuits. Although the total power consumption of 3-D systems is expected to be lower than that of mainstream 2-D circuits (since the global interconnects are shorter [26]), the increased power density is a challenging issue for this novel design paradigm. Therefore, those 3-D NoC topologies that offer low power characteristics should be of significant interest.

The different power consumption components for interconnects with repeaters are briefly discussed in this section. Due to specified performance characteristics, a low power design methodology with delay constraints for the interconnect in an NoC is adopted from [22]. An expression for the total power consumption per bit of a packet transferred between a source destination node pair is used as the basis for characterizing the power consumption of an NoC for the proposed topologies.

The power consumption components of an interconnect line with repeaters are as follows.

1) *Dynamic power consumption* is the dissipated power due to the charge and discharge of the interconnect and input gate capacitance during a signal transition, and can be described by

$$P_{di} = a_s f(c_i l_i + h_i k_i C_o)V_{dd}^2 \qquad (19)$$

where f is the clock frequency and a_s is the switching factor [27]. A value of 0.15 is assumed here; however, for NoC, the switching factor can vary considerably. This variation, however, does not affect the power comparison for the various topologies as the same switching factor is incorporated in each term for the total power consumed per bit of the network (the absolute value of the power consumption, however, changes).

2) *Short-circuit power* is due to the DC current path that exists in a CMOS circuit during a signal transition when the input signal voltage changes between V_{tn} and $V_{dd} + V_{tp}$. The power consumption due to this current is described as short-circuit power and is modeled in [28] by

$$P_{si} = \frac{4a_s f I_{d0}^2 t_{ri}^2 V_{dd} k_i h_i^2}{V_{dsat} G C_{eff i} + 2H I_{d0} t_{ri} h_i} \qquad (20)$$

where I_{d0} is the average drain current of the nMOS and pMOS devices operating in the saturation region and the value of the coefficients G and H are described in [29]. Due to resistive shielding of the interconnect capacitance, an effective capacitance is used in (20) rather than the total interconnect capacitance. This effective capacitance is determined from the methodology described in [30], [31].

3) *Leakage power* is comprised of two power components, the subthreshold and gate leakage currents. The subthreshold power consumption is due to current flowing in the cut-off region (below threshold), causing I_{sub} current to flow. The

gate leakage component is due to current flowing through the gate oxide, denoted as I_g. The total leakage power can be described as

$$P_{li} = h_i k_i V_{dd}(I_{\text{sub}0} + I_{g0}) \qquad (21)$$

where the average subthreshold I_{sub0} and gate I_{g0} leakage current of the nMOS and pMOS transistors is used in (21).

The total power consumption with delay constraint T_0 for a single line of a crossbar switch P_{stotal}, horizontal buss $P_{h\text{total}}$, and vertical buss $P_{v\text{total}}$ is, respectively,

$$P_{\text{stotal}}(T_0 - t_a) = P_{ds} + P_{ss} + P_{ls} \qquad (22)$$

$$P_{h\text{total}}(T_0) = P_{dh} + P_{sh} + P_{lh} \qquad (23)$$

$$P_{v\text{total}}(T_0) = P_{dv} + P_{sv} + P_{lv}. \qquad (24)$$

The power consumption of the arbitration logic does not appear in (22)–(24), since most of the power is consumed by the crossbar switch and the buss interconnect, as discussed in [32]. Note that for a crossbar switch, the additional delay of the arbitration logic poses a stricter delay constraint on the power consumption of the switch. The minimum power consumption with delay constraints is determined by the methodology described in [22], which is used to determine the optimum size $h_{\text{pow}i}^*$ and number $k_{\text{pow}i}^*$ of repeaters for a single interconnect line. Consequently, the minimum power consumption per bit between a source destination node pair in an NoC with a delay constraint is

$$P_{\text{bit}} = hops P_{\text{stotal}} + hops_{2\text{-D}} P_{h\text{total}} + hops_{3\text{-D}} P_{v\text{total}}. \qquad (25)$$

Note that the proposed power expression includes all of the power consumption components in the network, not only the dynamic power consumption typically considered. The effect of resistive shielding is also considered in determining the effective interconnect capacitance. Furthermore, since the repeater insertion methodology in [22] minimizes the power consumed by the repeater system, any additional decrease in power consumption is only due to the network topology. In the following section, the 3-D NoC topologies that exhibit the maximum performance and minimum power consumption with delay constraints are presented. Tradeoffs in determining these topologies are discussed and the impact of the network parameters on the resulting optimum topologies are evaluated for various network sizes.

V. COMPARISON OF 3-D NoC TOPOLOGIES

Several network parameters characterizing the topology of a network can significantly affect performance and power consumption. The evaluation of these network parameters is discussed in Section V-A. The improvement in network performance achieved by the proposed 3-D NoC topologies is presented in Section V-B. The distribution of nodes that provides the maximum performance is also discussed. The power consumption with delay constraints of a 3-D NoC and the topologies that yield the minimum power consumption of a 3-D NoC are presented in Section V-C.

TABLE II
INTERCONNECT PARAMETERS

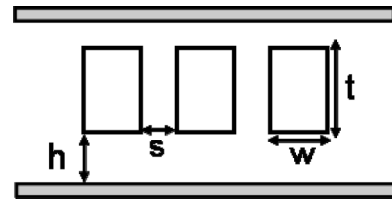| Interconnect Structure	Parameter		
	Electrical		Physical
Crossbar switch	ρ = 3.07 μΩ-cm		w = 200 nm
	k_{ILD} = 2.7		s = 200 nm
	r_s = 614 Ω/mm		t = 250 nm
	c_s = 157.6 fF/mm		h = 500 nm
Horizontal buss	ρ = 2.53 μΩ-cm		w = 500 nm
	k_{ILD} = 2.7		s = 250 (500) nm
	r_h = 46 Ω/mm		t = 1100 nm
	c_h = 332.6 (192.5) fF/mm		h = 800 nm
	$a_{3\text{-D}}$ = 1.02 (1.06)		-
Vertical buss	ρ = 5.65 μΩ-cm		w = 1050 nm
	r_v = 51.2 Ω/mm		L_v = 10 μm
	c_v = 600 fF/mm		-

Fig. 2. Typical interconnect structure.

A. 3-D NoC Parameters

The physical layer of a 3-D NoC consists of different interconnect structures, such as the crossbar switch, the horizontal buss connecting neighboring nodes on the same physical plane, and the vertical buss connecting nodes on different, not necessarily adjacent, physical planes. The device parameters characterizing the receiver, driver, and repeaters are considered to be common to all of these interconnect structures and are listed in Table I. The interconnect parameters reported in Table II, however, are different for each type of interconnect within a network.

A typical interconnect structure is shown in Fig. 2, where three parallel metal lines are sandwiched between two ground planes. Such an interconnect structure is considered for the crossbar switch (at the network nodes) where the intermediate metal layers are assumed here to be utilized. The horizontal buss is implemented on the global metal layers and, therefore, only the lower ground plane is present in this structure for a 2-D NoC. For a 3-D NoC, however, the substrate (back-to-front plane bonding) or a global metal layer of an upper plane (front-to-front plane bonding) behaves as a second ground plane. To incorporate this additional ground plane, the horizontal buss capacitance is multiplied by the appropriate coefficient $a_{3\text{-D}}$. A second ground plane decreases the coupling capacitance to an adjacent line, while the line-to-ground capacitance increases; hence, the total capacitance changes slightly as indicated by $a_{3\text{-D}}$. The vertical buss is different from the other structures in that this buss is implemented by the through silicon vias. These interplane vias can have significantly different impedance characteristics as compared to traditional horizontal interconnect structures, as discussed in [33] and

Parameter	Values
N	16, 32, 64, 128, 256, 512, 1024, 2048
$A_{PE}\,(\mathrm{mm}^2)$	0.5, 0.64, 0.81, 1, 1.5625, 2.25, 4
T_0 (ps)	1000, 500

also verified by extracted impedance parameters. The electrical interconnect parameters are extracted using a commercial impedance extraction tool [34], while the physical parameters are extrapolated from the predictive technology library [19] and the 3-D integration technology developed by MITLL for a 45 nm technology node [20]. The physical and electrical interconnect parameters are listed in Table II. For each of the interconnect structures, a buss width of 64 bits is assumed, while the packet size is assumed to be $L_p = 100w_c$. In addition, n_3 and n_p are constrained by the maximum number of physical planes $n_{\max}$ that can be vertically stacked. A maximum eight planes is assumed here. The constraints that apply for each of the 3-D NoC topologies shown in Fig. 1 are

$$n_3 \leq n_{\max}, \qquad \text{for 2-D IC–3-D NoC} \qquad (26a)$$

$$n_p \leq n_{\max}, \qquad \text{for 3-D IC–2-D NoC} \qquad (26b)$$

$$n_3 n_p \leq n_{\max}, \qquad \text{for 3-D IC–3-D NoC.} \qquad (26c)$$

A small set of parameters is used as variables to explore the performance and power consumption of the proposed 3-D NoC. This set includes the network size or, equivalently, the number of nodes within the network N, the area of each processing element A_{PE}, which is directly related to the buss length as described in (18), and the maximum allowed interconnect delay when evaluating the minimum power consumption with delay constraints. The range of values for these variables is listed in Table III. Depending upon the network size, NoC are roughly divided in this paper as small ($N = 16$ to 64 nodes), medium ($N = 128$ to 256 nodes), and large ($N = 512$ to 2048 nodes) networks. For multi-processor SoC networks, sizes of up to $N = 256$ are expected to be feasible in the near future [11], [35], whereas for NoC with a finer granularity, where the PEs each correspond to hardware blocks of approximately 100 K gates, network sizes over a few thousands nodes are predicted at the 45 nm technology node [36]. Note that this classification of the networks is not strict and is only intended to facilitate the discussion in the following sections.

B. Performance Tradeoffs for 3-D NoC

The performance enhancements that can be achieved in NoC by utilizing the third dimension are investigated in this subsection. Each of the proposed 3-D topologies decreases the zero-latency of the network by reducing different delay components, as described in (9). In addition, the distribution of network nodes in each physical dimension that yields the minimum zero-load latency is shown to significantly change with the network and interconnect parameters.

1) 2-D IC–3-D NoC: Utilizing the third dimension to implement an NoC directly results in a decrease in the average number of hops for packet switching. The average number of hops on the same plane $hops_{2\text{-D}}$ (the intraplane hops) and the average number of hops in the third dimension $hops_{3\text{-D}}$ (the interplane hops) are also reduced. Interestingly, the distribution of nodes n_1, n_2, and n_3 that yields the minimum total number of hops is not always the same as that distribution that minimizes the number of intraplane hops. This situation occurs particularly for small and medium networks, while for large networks, the distribution of n_1, n_2, and n_3 which minimizes the $hops$ also minimizes $hops_{2\text{-D}}$.

In a 3-D NoC, the number of router ports increases from five to seven, increasing, in turn, both the switch and arbiter delays. Furthermore, a short vertical buss generally exhibits a lower delay than that of a relatively long horizontal buss.

The node distribution that produces the lowest latency varies with network size. For example, $n_{3max} = 8$ is not necessarily the optimum for small and medium networks, although by increasing n_3, more hops occur through the short, low latency vertical channel. This result can be explained by considering the reduction in the number of hops that originate from utilizing the third dimension for packet switching. For small and medium networks, the decrease in the number of hops is small and cannot compensate the increase in the routing delay due to the increase in the number of ports of a router in a 3-D NoC. As the horizontal buss length becomes longer, however, (e.g., approaching 2 mm), $n_3 > 1$, and a slight decrease in the number of hops significantly decreases the overall delay, despite the increase in the routing delay for a 3-D NoC. As an example, consider a network with $log_2 N = 4$ and $A_{PE} = 1\ \mathrm{mm}^2$. The minimum latency node distribution is $n_1 = n_2 = 4$ and $n_3 = 1$ (identical to a 2-D IC–2-D NoC as shown in Fig. 3), while for $A_{PE} = 4\ \mathrm{mm}^2$, $n_1 = n_2 = 2$ and $n_3 = 4$.

The optimum node distribution can also be affected by the delay of the vertical channel. The repeater insertion methodology for minimum delay as described in Section III can significantly reduce the delay of the horizontal buss by inserting large sized repeaters (i.e., $h > 300$). In this case, the delay of the vertical buss becomes comparable to that of the horizontal buss with repeaters. Consider a network with $N = 128$ nodes. Two different node distributions yield the minimum average number of hops, specifically, $n_1 = 4$, $n_2 = 4$, and $n_3 = 8$, and $n_1 = 8$, $n_2 = 4$, and $n_3 = 4$. The first of the two distributions also results in the minimum number of intraplane $hops_{2\text{-D}}$, thereby reducing the latency component for the horizontal buss described in (9). Simulation results, however, indicate that this distribution is not the minimum latency node distribution, as the delay due to the vertical channel is non-negligible. For this reason, the latter distribution with $n_3 = 4$ is preferable, since a smaller number of $hops_{3\text{-D}}$ occurs, resulting in the minimum network latency.

2) 3-D IC–2-D NoC: For this type of 3-D network, the PEs are allowed to span multiple physical planes while the network effectively remains 2-D (i.e., $n_3 = 1$). Consequently, the network latency is only reduced by decreasing the length of the horizontal buss, as described in (18). The routing delay component remains constant with such a 3-D topology. Decreasing the horizontal buss length by using multiple physical planes lowers both the communication channel delay and the serialization delay;
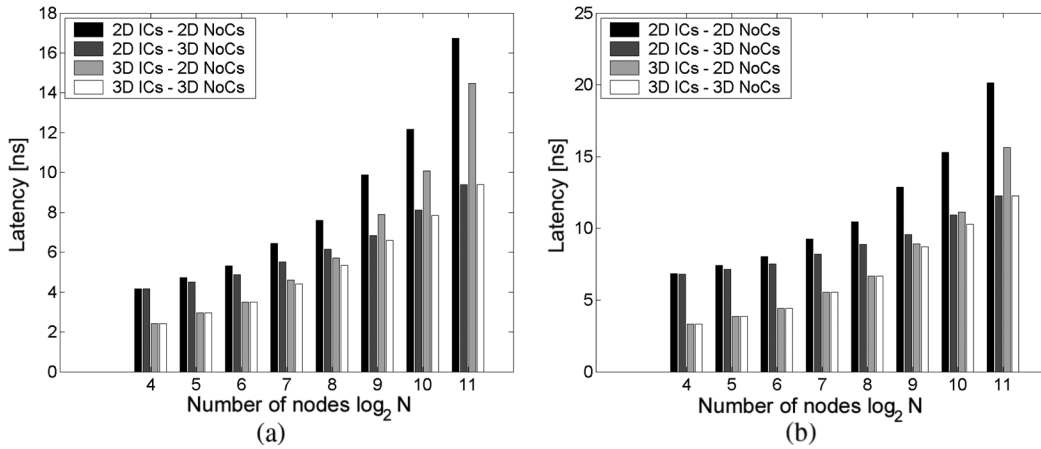
Fig. 3. Zero-load latency for various network sizes. (a) $A_{PE} = 1$ mm^2 and $c_h = 332.6$ fF/mm, (b) $A_{PE} = 4$ mm^2 and $c_h = 332.6$ fF/mm.
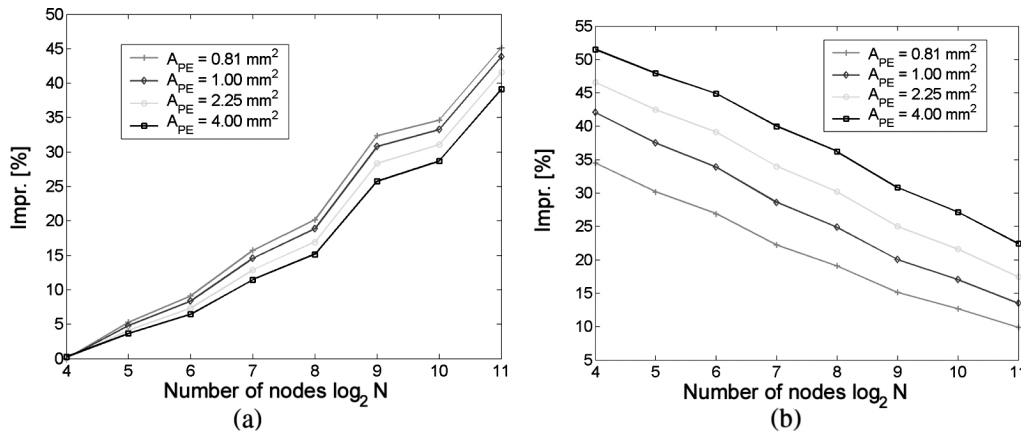

Fig. 4. Improvement in zero-load latency for different network sizes and PE areas (i.e., buss lengths). (a) 2-D IC–3-D NoC and (b) 3-D IC–2-D NoC.

therefore, the optimum value for $n_p = n_{max}$, regardless of the network size and buss length.

In Fig. 4(a) and (b), the improvement in the network latency over a 2-D IC–2-D NoC for various network sizes and for different PE areas (i.e., different horizontal buss length) is illustrated for the 2-D IC–3-D NoC and 3-D IC–2-D NoC topologies, respectively. Note that for the 2-D IC–3-D NoC topology, the improvement in delay is smaller for PEs with larger area or, equivalently, with longer buss lengths independent of the network size. For longer buss lengths, the buss latency comprises a larger portion of the total network latency. Since for a 2-D IC–3-D NoC only the hop count is reduced, the improvement in latency is lower for longer buss lengths. Alternatively, the improvement in latency is greater for PEs with larger areas independent of the network size for 3-D IC–2-D NoC. This situation is due to the significant reduction in PE area (or buss length) that is achieved with this topology. Consequently, there is a tradeoff in the latency of a NoC that depends both on the network size and the area of the PEs. In Fig. 4(a), the improvement is not significant for small networks (all of the curves converge approximately to zero) in 2-D IC–3-D NoC while this situation does not occur for 3-D IC–2-D NoC. This behavior is due to the increase in the delay of the network router as the number of ports increases from five to seven for 2-D IC–3-D NoC, which is a considerable portion of the network latency for small networks.

Note that for 3-D IC-2-D NoC, the network essentially remains two dimensional and therefore the delay of the router for this topology does not increase. To achieve the minimum delay, a 3-D NoC topology that exploits these tradeoffs is proposed in the following subsection.

3) 3-D IC–3-D NoC: This topology offers the greatest decrease in latency over the aforementioned 3-D topologies. The 2-D IC–3-D NoC topology decreases the number of hops while the buss and serialization delays remain constant. With the 3-D IC–2-D NoC, the buss and serialization delay is smaller but the number of hops remains unchanged. With the 3-D IC–3-D NoC, all of the latency components can be decreased by assigning a portion of the available physical planes for implementing the network while the remaining planes of the stack are used for the PE. The resulting decrease in network latency as compared to the basic 2-D IC–2-D NoC and the other two 3-D topologies is shown in Fig. 3. A decrease in latency of 40% and 36% can be observed for $N = 128$ and $N = 256$ nodes, respectively, with $A_{PE} = 4$ mm^2. Note that the 3-D IC–3-D NoC topology achieves the greatest savings in latency by optimally balancing the values of n_3 and n_p.

For certain network sizes, the performance of the 3-D IC–2-D NoC is identical to either the 2-D IC–3-D NoC or 3-D IC–2-D NoC. This behavior occurs because for large network sizes, the delay due to the greater number of hops dominates the total
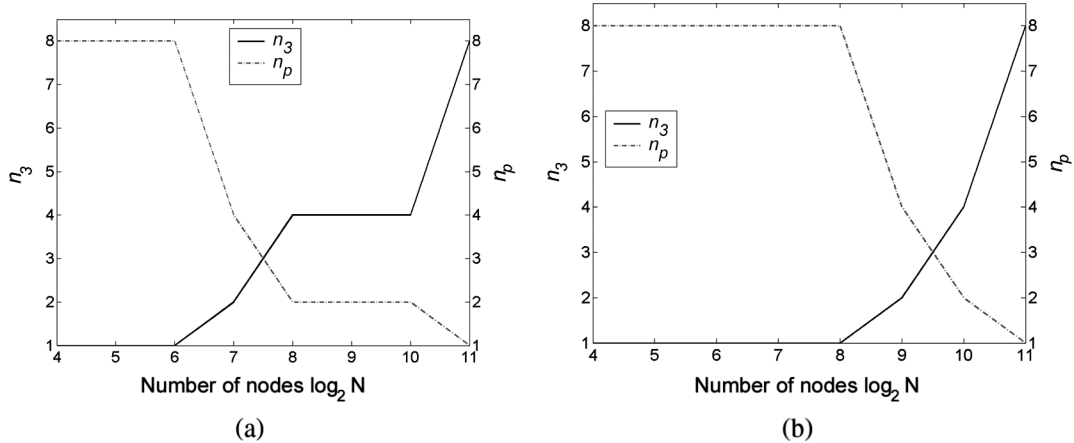
Fig. 5. n_3 and n_p values for minimum zero-load latency for various network sizes. (a) $A_{PE} = 1$ mm^2 and $c_h = 332.6$ fF/mm, (b) $A_{PE} = 4$ mm^2 and $c_h = 332.6$ fF/mm.

delay and, therefore, the latency can be primarily reduced by decreasing the average number of hops. For small networks, the buss delay is large and the latency savings is typically achieved by reducing the buss length ($n_p = n_{max}$). For medium networks, though, the optimum topology is obtained by dividing n_{max} between n_3 and n_p such that (26c) is satisfied. This distribution of n_3 and n_p as a function of the network size and buss length is illustrated in Fig. 5.

Note the shift in the value of n_3 and n_p as the PE area A_{PE} or, equivalently, the buss length increases. For long busses, the delay of the communication channel becomes dominant and therefore the fewer number of hops for medium sized networks cannot significantly decrease the total delay. Alternatively, further decreasing the buss length by implementing the PEs in a greater number of physical planes leads to a larger savings in delay.

The suggested optimum topologies for various network sizes (namely, small, medium, and large networks) also depend upon the interconnect parameters of the network. Consequently, a change regarding the optimum topologies for different network sizes can occur when different interconnect parameters are considered. Despite the sensitivity of the topologies to the interconnect parameters, the tradeoff between the number of hops and the buss length for various 3-D topologies (see Figs. 4 and 5) can be exploited to improve the performance of an NoC. In the following subsection, the topology that yields the minimum power consumption with delay constraints is described. The distribution of nodes for that topology is also discussed.

C. Power Consumption in 3-D NoC

The various power consumption components for the interconnect within an NoC are analyzed in Section IV. The methodology presented in [22] is applied here to minimize the power consumption of these interconnects while satisfying the specified operating frequency of the network. Since a power minimization methodology is applied to the buss lines, the power consumption of the network can only be further reduced by the choice of network topology. Additionally, the power consumption also depends upon the target operating frequency, as discussed later in this section.

As with the zero-load latency, each topology affects the power consumption of the network in a different way. From (25), the power consumption can be reduced by either decreasing the number of hops for the packet or by decreasing the buss length. Note that by reducing the buss length, the interconnect capacitance is not only reduced but also the number and size of the repeaters required to drive the lines are decreased, resulting in a greater power savings. The effect of each of the proposed topologies on the power consumption of an NoC is investigated in this subsection.

1) 2-D IC–3-D NoC: Similar to the network latency, the power consumption is decreased in this topology by reducing the number of hops for packet switching. Again, the increase in the number of ports is significant; however, the impact from this increase is not as important as that on the latency of the network. A 3-D network, therefore, can reduce power even in small networks. The power savings achieved with this topology is greater in larger networks. This situation occurs because the reduction in the average number of hops for a 3-D network increases for larger network sizes.

2) 3-D IC–2-D NoC: With this topology, the number of hops in the network is the same as for a 2-D network. The horizontal buss length, however, is shorter by implementing the PEs in more than one physical plane. The greater the number of physical planes that can be integrated in a 3-D system, the larger the power savings, meaning that the optimum value for n_p with this topology is always n_{max} regardless of the network size and operating frequency. The achieved savings is practically limited by the number of physical planes that can be integrated in a 3-D technology. For this type of NoC, the maximum performance topology is identical to the minimum power consumption topology, as the key element of both objectives originates solely from the shorter buss length.

3) 3-D IC–3-D NoC: Allowing the available physical planes to be utilized either for the third dimension of the network or for the PEs, the 3-D IC–3-D NoC scheme achieves the greatest savings in power in addition to the minimum delay, as discussed in the previous subsection. The distribution of nodes along the physical dimensions, however, that produces either the minimum latency or the minimum power consumption for every network size is not necessarily the same. This non-equivalence
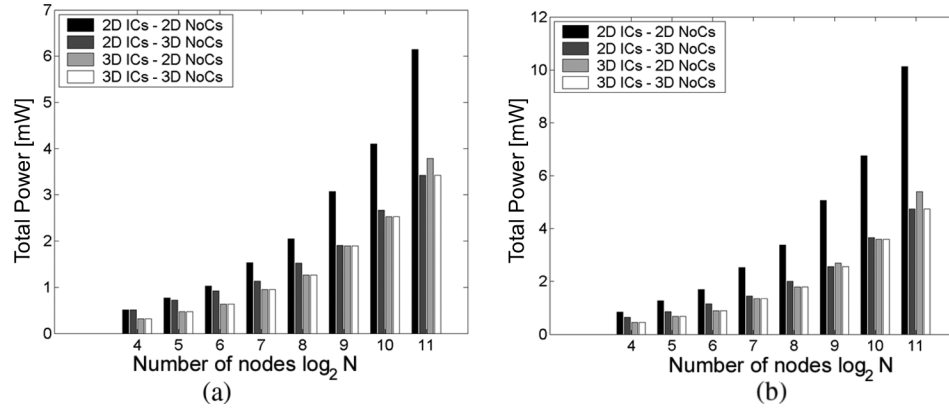
Fig. 6. Power consumption with delay constraints for various network sizes. (a) $A_{\mathrm{PE}} = 1 \text{ mm}^2$, $c_h = 332.6 \text{ fF/mm}$, and $T_0 = 500 \text{ ps}$. (b) $A_{\mathrm{PE}} = 4 \text{ mm}^2$, $c_h = 332.6 \text{ fF/mm}$, and $T_0 = 500 \text{ ps}$.

originates from the different degree of importance of the average number of hops and the buss length in determining the latency and power consumption of a network. In Fig. 6, the power consumption of the 3-D IC–3-D NoC topology is compared to the three-dimensional topologies previously discussed. A power savings of 38.4% is achieved for $N = 128$ with $A_{\mathrm{PE}} = 1 \text{ mm}^2$. For certain network sizes, the power consumption of the 3-D IC–3-D NoC topology is the same as that of the 2-D IC–3-D NoC and 3-D IC–2-D NoC topologies. For the 2-D IC–3-D NoC, the power consumption is primarily decreased by reducing the number of hops for packet switching, while for the 3-D IC–2-D NoC, the NoC power dissipation is decreased by shortening the buss length. The former approach typically benefits small networks, while the latter approach yields lower power consumption for large networks. For medium sized networks and depending upon the network and interconnect parameters, non-extreme values for the n_3 and n_p parameters (e.g., $1 < n_3 < n_{\max}$ and $1 < n_p < n_{\max}$) are required to produce the minimum power consumption topology.

Note that this work emphasizes the latency and power consumption of a network, neglecting the performance requirements of the individual PEs. If the performance of the individual PEs is important, only one 3-D topology may be available; however, even with this constraint, a significant savings in latency and power can be achieved since in almost every case the network latency and power consumption can be decreased as compared to a 2-D IC–2-D NoC topology. Furthermore, as previously mentioned, if the available topology is the 2-D IC–3-D NoC, setting n_3 equal to $n_{\max}$ is not necessarily the optimum choice.

The proposed zero-load network latency and power consumption expressions capture the effect of the topology; yet these models do not incorporate the effect of the routing scheme and traffic load. The effect of the third dimension on the NoC topologies is accurately characterized by the proposed models. Alternatively, these models can be perceived as lower bounds both for the latency and the power consumption of the network. Since minimum distance paths and no contention are implicitly assumed in the proposed expressions, non-minimal path routing schemes and heavy traffic loads will result in increasing both the latency and power consumption of the network. Finally, the

latency and power consumption of the different interconnect structures of the network, which are shown to be significant, are accurately described by the proposed expressions and do not change under any routing conditions and/or traffic load.

VI. CONCLUSION

3-D NoC are a natural evolution of 2-D NoC, exhibiting superior performance. Several novel 3-D NoC topologies are presented. The zero-load latency of the network is modeled for each of these topologies. Expressions for the power consumption per bit with delay constraints are also provided. The minimum latency and power consumption can be achieved by reducing both the number of hops per packet and the length of the communication channels. The 3-D IC–3-D NoC topology provides the optimum choice in terms of minimizing the zero-load network latency and power consumption, as with this topology both the delay and power consumption components can be efficiently reduced. For the case where the impedance characteristics of the buss and crossbar switch within the network are of similar magnitude, the 2-D IC–3-D NoC offers the minimum latency and power consumption, while for large networks, the impedance of the buss determines the delay and power characteristics of the network and, therefore, a 3-D IC–2-D NoC topology yields the best results. For medium sized networks, a 3-D IC–3-D NoC topology is preferable, since in these network sizes both the number of hops and the length of the buss can be decreased to produce the minimum zero-load latency and power consumption.

REFERENCES

[1] R. J. Gutmann *et al.*, "Three-dimensional (3D) ICs: A technology platform for integrated systems and opportunities for new polymeric adhesives," in *Proc. Conf. Polymers Adhesives Microelectron. Photon.*, 2001, pp. 173–180.
[2] J. W. Joyner *et al.*, "Impact of three-dimensional architectures on interconnects in gigascale integration," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 9, no. 6, pp. 922–927, Dec. 2000.

[3] V. F. Pavlidis and E. G. Friedman, "Interconnect delay minimization through interlayer via placement," in *Proc. ACM Great Lakes Symp. VLSI*, 2005, pp. 20–25.

[4] W. R. Davis *et al.*, "Demystifying 3D ICs: The pros and cons of going vertical," *IEEE Design Test Comput.*, vol. 22, no. 6, pp. 498–510, Nov./Dec. 2005.

[5] L. Benini and G. De Micheli, "Networks on chip: A new SoC paradigm," *IEEE Comput.*, vol. 31, no. 1, pp. 70–78, Jan. 2002.

[6] D. Bertozzi *et al.*, "NoC synthesis flow for customized domain specific multiprocessor systems-on-chip," *IEEE Trans. Parallel Distr. Syst.*, vol. 16, no. 2, pp. 113–129, Feb. 2005.

[7] J. C. Koob *et al.*, "Design of a 3-D fully depleted SOI computational RAM," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 13, no. 3, pp. 358–368, Mar. 2005.

[8] S. Kumar *et al.*, "A network on chip architecture and design methodology," in *Proc. Int. IEEE Annu. Symp. VLSI*, 2002, pp. 105–112.

[9] V. F. Pavlidis and E. G. Friedman, "3-D topologies for networks-on-chip," in *Proc. IEEE Int. SOC Conf.*, 2006, pp. 285–288.

[10] C. Addo-Quaye, "Thermal-aware mapping and placement for 3-D NoC designs," in *Proc. IEEE Int. Syst.-on-Chip Conf.*, 2005, pp. 25–28.

[11] F. Li *et al.*, "Design and management of 3D chip multiprocessors using network-in-memory," in *Proc. IEEE Int. Symp. Comput. Arch.*, 2006, pp. 130–142.

[12] W. J. Dally, "Performance analysis of k-ary n-cube interconnection networks," *IEEE Trans. Comput.*, vol. 39, no. 6, pp. 775–785, Jun. 1990.

[13] A. Jantsch and H. Tenhunen, *Networks on Chip*. New York: Kluwer, 2003.

[14] M. Millberg *et al.*, "The nostrum backbone—A communication protocol stack for networks on chip," in *Proc. IEEE Int. Conf. VLSI Des.*, 2004, pp. 693–696.

[15] J. M. Duato, S. Yalamanchili, and L. Ni, *Interconnection Networks: An Engineering Approach*. San Mateo, CA: Morgan Kaufmann, 2003.

[16] W. J. Dally and B. Towles, *Principles and Practices of Interconnection Networks*. San Mateo, CA: Morgan Kaufmann, 2004.

[17] L.-S. Peh and W. J. Dally, "A delay model for router microarchitectures," *IEEE Micro*, vol. 21, no. 1, pp. 26–34, Jan./Feb. 2001.

[18] T. Sakurai, "Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSI's," *IEEE Trans. Electron Devices*, vol. 40, no. 1, pp. 118–124, Jan. 1993.

[19] Predictive Technology Model [Online]. Available: http://www.eas.asu.edu/~ptm

[20] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45 nm design exploration," in *Proc. IEEE Int. Symp. Quality Electron. Des.*, 2006, pp. 585–590.

[21] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, Apr. 1990.

[22] G. Chen and E. G. Friedman, "Low-power repeaters driving RC and RLC interconnects with delay and bandwidth constraints," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 2, pp. 161–172, Feb. 2006.

[23] Y. I. Ismail, E. G. Friedman, and J. L. Neves, "Equivalent Elmore delay for RLC trees," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 19, no. 1, pp. 83–97, Jan. 2000.

[24] ——, "Figures of merit to characterize the importance of on-chip inductance," *IIEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 7, no. 4, pp. 442–449, Dec. 1999.

[25] *"FDSOI Design Guide,"* MIT Lincoln Lab., Cambridge, MA, 2006.

[26] H. Hua *et al.*, "Performance trends in three-dimensional integrated circuits," in *Proc. Int. IEEE Interconnect Technol. Conf.*, 2006, pp. 45–47.

[27] K. Banerjee and A. Mehrotra, "A power-optimal repeater insertion methodology for global interconnects in manometer design," *IEEE Trans. Electron Devices*, vol. 49, no. 11, pp. 2001–2007, Nov. 2002.

[28] H. J. M. Veendrick, "Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 4, pp. 468–473, Aug. 1984.

[29] K. Nose and T. Sakurai, "Analysis and future trend of short-circuit power," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 19, no. 9, pp. 1023–1030, Sep. 2000.

[30] G. Chen and E. G. Friedman, "Effective capacitance of RLC loads for estimating short-circuit power," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2006, pp. 2065–2068.

[31] P. R. O'Brien and T. L. Savarino, "Modeling the driving-point characteristic of resistive interconnect for accurate delay estimation," in *Proc. Int. IEEE/ACM Conf. Comput.-Aided Des.*, 1989, pp. 512–515.

[32] H. Wang, L.-S. Peh, and S. Malik, "Power-driven design of router microarchitectures in on-chip networks," in *Proc. IEEE Int. Symp. Microarch.*, 2003, pp. 105–116.

[33] C. Ryu *et al.*, "High frequency electrical circuit model of chip-to-chip vertical via interconnection for 3-D chip stacking package," in *Proc. IEEE Topical Meeting Elect. Perform. Electron. Packag.*, 2005, pp. 151–154.

[34] *"Metal User's Guide"* OEA Int. Inc., Morgan Hill, CA, 2004. [Online]. Available: www.oea.com

[35] C. Marcon *et al.*, "Exploring NoC mapping strategies: An energy and timing aware technique," in *Proc. ACM/IEEE Des., Automat. Test Eur. Conf. Exhibit.*, 2005, vol. 1, pp. 502–507.

[36] P. P. Pande *et al.*, "Performance evaluation and design trade-offs for network-on-chip interconnect architectures," *IEEE Trans. Comput.*, vol. 54, no. 8, pp. 1025–1039, Aug. 2005.

Vasilis F. Pavlidis (S'04) received the B.S. and M.Eng. degrees in electrical and computer engineering from the Democritus University of Thrace, Xanthi, Greece, in 2000 and 2002, respectively, and the M.S. degree in electrical and computer engineering from the University of Rochester, Rochester, NY, in 2003, where he is currently pursuing the Ph.D. degree.

From 2000 to 2002, he was with INTRACOM S.A., Athens, Greece. His current research interests are in the area of interconnect modeling, 3-D integration, NoC, and related design issues in VLSI.

Eby G. Friedman (S'78–M'79–SM'90–F'00) received the B.S. degree from Lafayette College in 1979, and the M.S. and Ph.D. degrees from the University of California, Irvine, in 1981 and 1989, respectively, all in electrical engineering.

From 1979 to 1991, he was with Hughes Aircraft Company, rising to the position of manager of the Signal Processing Design and Test Department, responsible for the design and test of high performance digital and analog IC's. He has been with the Department of Electrical and Computer Engineering at the University of Rochester since 1991, where he is a Distinguished Professor, the Director of the High Performance VLSI/IC Design and Analysis Laboratory, and the Director of the Center for Electronic Imaging Systems. He is also a Visiting Professor at the Technion-Israel Institute of Technology. His current research and teaching interests are in high performance synchronous digital and mixed-signal microelectronic design and analysis with application to high speed portable processors and low power wireless communications. He is the author of more than 300 papers and book chapters, several patents, and the author or editor of eight books in the fields of high speed and low power CMOS design techniques, high speed interconnect, and the theory and application of synchronous clock and power distribution networks.

Dr. Friedman is the Regional Editor of the *Journal of Circuits, Systems and Computers*, a member of the editorial boards of the *Analog Integrated Circuits and Signal Processing, Microelectronics Journal, Journal of Low Power Electronics*, and *Journal of VLSI Signal Processing*, Chair of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS steering committee, and a member of the technical program committee of a number of conferences. He previously was the Editor-in-Chief of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, a member of the editorial board of the PROCEEDINGS OF THE IEEE and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: ANALOG AND DIGITAL SIGNAL PROCESSING, a member of the IEEE Circuits and Systems (CAS) Society Board of Governors, Program and Technical chair of several IEEE conferences, Guest Editor of several special issues in a variety of journals, and a recipient of the Howard Hughes Masters and Doctoral Fellowships, the University of Rochester Graduate Teaching Award, and a College of Engineering Teaching Excellence Award. He is a Senior Fulbright Fellow.