

Small-World Power-Law Interconnects for Nanoscale Computing Architectures

Christof Teuscher

Los Alamos National Laboratory, CCS-1, MS-B287, Los Alamos, NM 87545, USA

E-mail: christof@teuscher.ch

Abstract—Interconnects on today’s chips have become more important than transistors and this trend is predicted to continue with ongoing miniaturization and forthcoming beyond-silicon technologies. Current ad hoc interconnect technology is not suitable for multi-billion-transistor chip integration and we argue that radically new approaches potentially offer better performance for a lower price. The goal of this paper is to explore certain design and performance tradeoffs of both regular and irregular interconnect fabrics for self-assembled nanoscale electronics in a realistic framework. We show that fabrics with small-world-like properties have major advantages in terms of performance and robustness over purely regular fabrics. Our findings support the unconventional approach of self-assembling nanoscale electronics in a mostly random manner, which is believed to be more fabrication friendly and thus cheaper to realize.

I. INTRODUCTION AND MOTIVATION

The physical realization of computations from an abstract computing machine is a challenging task, which is usually guided by a number of major tradeoffs in the design space, such as the number and characteristics of the resources available, the required performance, the energy consumption, and the reliability. The lack of systematic understanding of these issues and of clear design methodologies makes the process still more of an art than of a scientific endeavor. The appearance of novel and non-standard physical computing devices for nanoscale and molecular electronics (such as for example array-based [4] architectures or random assemblies of molecular gates [17]) only aggravates these difficulties.

In recent years, the importance of interconnects on chips has outrun the importance of transistors as a dominant factor of chip performance [3], [5], [10]. The ITRS roadmap [1] lists a number of critical challenges for interconnects and states that “[i]t is now widely conceded that technology alone cannot solve the on-chip global interconnect problem with current design methodologies.” The major problems are related to delays of non-scalable global interconnects and reliability in general, which leads to the observation that simple scaling will no longer satisfy performance requirements as feature sizes continue to shrink [5].

The main challenge of interconnect fabrics—seen from a bird’s eye view—consists in transferring data between two points of the chip with a minimal latency, minimal energy consumption, and maximal reliability. This job can obviously be done in a wide variety of ways. As opposed to the traditional ad hoc and monolithic interconnect networks used

in most of today’s chip designs, people in recent years started to look at novel *Network-on-Chip* (NoC) [2], [12] paradigms and regular interconnects as commonly used in FPGAs and in the parallel-computing field.

The goal of this paper is to experimentally explore certain design and performance tradeoffs of unconventional and irregular, but physically plausible 3D interconnect fabrics, that are targeted for self-assembled nanoscale electronics, and to compare them to regular and nearest-neighbor connected 2D and 3D fabrics. The motivations for looking into radically different approaches are based on the following observations and facts: (1) long-range and global connections are costly and limit system performance [5]; (2) it is unclear whether a precisely regular and homogeneous arrangement of components is needed and possible on a multi-billion-component nanoscale assembly [17]; (3) “[s]elf-assembly makes it relatively easy to form a random array of wires with randomly attached switches” [20]; and (4) building a perfect system is very hard and expensive. We will only focus on interconnects in this paper and neither on the processing elements nor on how to map and solve problems on such a self-assembled architecture.

By using an abstract, speculative, yet realistic, and fabrication-friendly nanoscale computing framework, we will show that interconnect fabrics with small-world-like [18] properties have major advantages in terms of performance and robustness over purely regular and nearest-neighbor connected fabrics. Our findings support the unconventional approach of self-assembling nanoscale electronics in a mostly random manner, which is likely to be cheaper and easier to realize.

II. COMPLEX NETWORKS

Most real networks, such as brain networks [15], electronic circuits [6], the Internet, and social networks, share the so-called *small-world* (SW) property [18]. Compared to fully regular networks (such as the cellular automata interconnect), small-world networks have a very short average distance between any pair of nodes, which makes them particularly interesting for efficient communication.

The classical Watts-Strogatz small-world network [18] is built from a regular lattice with only nearest neighbor connections. Every link is then rewired with a *rewiring probability* p to a randomly chosen node. Thus, by varying p , one can obtain a fully regular ($p = 0$) and a fully random ($p = 1$) network topology. The rewiring procedure establishes “shortcuts” in the network, which significantly lower the average distance (i.e.,

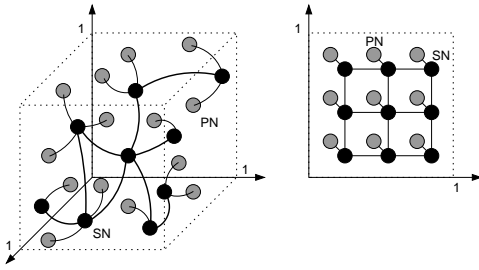


Fig. 1. Left: a random multitude (RM) example composed of processing nodes (PNs), switch nodes (SNs), and interconnections. Right: a 2D CA-like architecture.

the number of edges to traverse) between any pair of nodes. In the original model, the length distribution of the shortcuts is uniform since a node is chosen randomly. If the rewiring of the connections is done proportional to a power law, $l^{-\alpha}$, where l is the wire length, then we obtain a *small-world power-law network*. The exponent α affects the network’s transport characteristics [9] and navigability [8], which is better than in the uniformly generated SW network.

In a real network, it is fair to assume that local connections have a lower cost (in terms of resources required and wire delay) than long-distance connections. Physically realizing small-world networks with uniformly distributed long-distance connections is thus not realistic and distance, i.e., the wiring cost, needs to be taken into account [7], [13], [14]. The same holds for random Erdős-Rényi networks.

On the other hand, a network’s topology also directly affects how efficiently problems can be solved. For example, it has been shown that both SW topologies [16] as well as random Erdős-Rényi topologies [11] have better performance than regular lattices and are easier to evolve to solve the global synchronization and density classification task. Thus, although generally easy to physically realize, local-neighborhood networks are mostly not suitable for solving problems efficiently because of their poor global transport characteristics.

In this paper we are interested in small-world power-law networks because they present a realistic model of a fabrication-friendly, self-assembled nano-scale interconnect fabric with great transport characteristics.

III. DESCRIPTION OF THE FRAMEWORK

In order to compare representative regular nearest-neighbor and irregular small-world interconnect fabrics, we use an abstract, yet physically realistic system-on-chip framework and an evaluation methodology inspired by Pande et al. [12]. We will compare selected measures that are relevant for real systems on three interconnect architectures, which shall be briefly described in the following sections.

A. Regular 2D and 3D CA-like Architectures

Both 2D and 3D *cellular automata* (CA) like architectures are used as representatives of regular nearest-neighbor interconnect fabrics. The basic system-on-chip-like architecture is composed of programmable computing elements, called

processing nodes (PNs), and of an associated switch-based interconnect fabric, which is itself composed of *switch nodes* (SNs) and bi-directional point-to-point interconnects. We use an unfolded version (see Figure 1, right), called *CLICHÉ* in [12], since folding requires long-distance connections. The PNs are regularly arranged in 2D or 3D Euclidean space inside a unitary square, respectively cube. The number of PNs is equal to the number of SNs, and each PN is connected to its associated SN by a single connection of 0.01 unit length. For our purposes, the PNs are able to send and receive messages, whereas the SNs perform basic routing only.

B. The Irregular Random Multitude (RM)

In a *random multitude* (RM), both PNs and SNs are randomly arranged in 3D space, as illustrated in Figure 1 (left). To make comparisons with the CA-like architectures easier, we assume that each PN is connected to the nearest SN in space by a single connection. The SNs are connected among themselves by a small-world power-law network [7], [13], [14] with average connectivity $\langle k \rangle$, i.e., each node establishes connections with its neighbors proportional to $l^{-\alpha}$, where l is the Euclidean distance between the two SNs in question. Thus, the bigger α , the more local the connections. If $\alpha = 0$, we obtain the original Watts-Strogatz SW topology.

C. Physical Realization of a Random Multitude

There exists an abundance of abstract computing models which are either hard or impossible (e.g., when infinite resources or time is involved) to physically realize. Because of fewer constraints, we argue that computing architectures that are “assembled” in a largely random manner are easier and cheaper to build than highly regular architectures, such as crossbars or CA-like assemblies. Self-assembly, for example, is particularly well suited for building random structures [20]. Power-law connection-length distributions have been observed in many systems created through self-organization, such as the human cortex or the Internet. Also, Tour et al. [17], for example, explored the possibility of computing with randomly assembled, easily realizable molecular switches.

We imagine that a random multitude would be best realized in a hybrid way first, where the PNs and SNs—which might be considered as very simple IP blocks—are for example made of current nanoscale silicon. The interconnect fabric would then be gradually self-assembled using techniques such as directed assembly [19] by means of electrodeposition or vapor deposition, or any other suitable technique. To obtain a power-law distribution of connection lengths, one might imagine fabricating a large amount of wires first, whose lengths follow a power law distribution. In a second step, they would be immersed in a solution together with the nodes, randomly aligned (e.g., by means of electric fields), and soldered as described in [19].

IV. EXPERIMENTS

We performed a number of experiments to compare and contrast the three interconnect architectures as described in Section III.

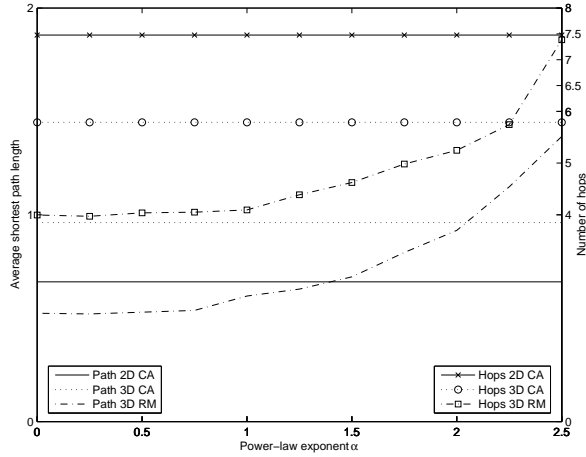


Fig. 2. Average length of shortest path and number of hops as a function of α . Shortest path routing (SPR), average values over 2 runs.

A. Methods

For our purposes, we assume that each SN can only perform either random (RR) or shortest path (SPR) routing. Many other and more efficient routing techniques exist, but we consider these two as representatives of the least and the most effective methods. Each node has a variable number of virtual channels, which allows to process messages in parallel. As a simplification, all buffer sizes are considered unlimited. All nodes are updated asynchronously. Our simulations use a simplistic random traffic model, which generates a message with a random destination with a certain probability $trLoad$ in each PN. If $trLoad = 1$, one message will be generated in each node at each update. To be able to compare the results with the CA-like topologies, we keep the number of SNs and PNs equal in the random multitude architecture. We measured the following performance metrics: (1) average message latency (in clock cycles); (2) the average shortest path length (in distance units); (3) the average number of hops; (4) and the throughput (in messages/number of updates/switch node). For all experiments, we used $|SN| = |PN| = 64$, 6 virtual channels per node (i.e., a 3D CA-node could in principle send a message into all directions simultaneously), an average SN connectivity of $\langle SN_k \rangle = 6$, an exact PN connectivity of 1, and a traffic load of $trLoad = 0.1$. The average connectivity was chosen to have roughly the same connectivity as a 3D local-neighborhood interconnect. For our purposes, we kept all these parameters constant as a detailed analysis would have been beyond the scope of this paper.

B. Results

Figure 2 shows the average length of the shortest paths (left y-axis) between each pair of PNs and the number of hops (right y-axis) as a function of the power-law exponent α . As one can see, the average shortest path gets shorter the smaller α , i.e., the more long-distance connections exist. The effect on the number of hops is identical.

According to [14], a network is in a small-world regime if

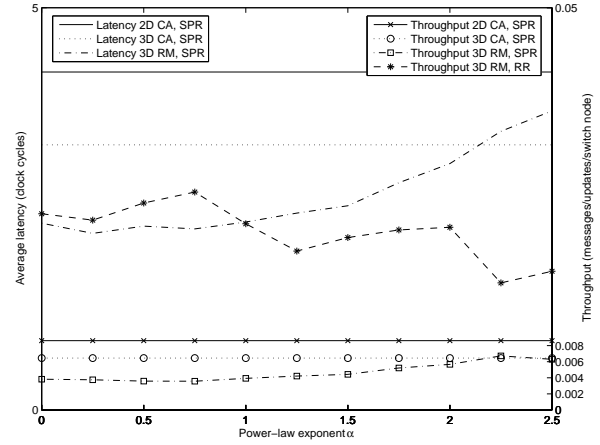


Fig. 3. Latency and throughput as a function of α . Shortest path routing (SPR), random routing (RR), average values over 2 runs.

$\alpha < 2D$, where D is the dimension of the original lattice. However, compared to [14], we allow multiple connections and the nodes are randomly arranged, but the above equation (without doing the math) should still approximately held for $D = 3$.

Figure 3 shows both latency and throughput as a function of the power-law exponent α . Due to the long-distance connections, the random multitude has a lower latency than both 3D and 2D local-neighborhood interconnects. Rather surprisingly, the throughput of the random multitude is the worst if one uses shortest path routing (SPR). This is because SPR uses the same nodes for numerous paths and thus creates more congestion because of the limited number of channels per node. If one uses random routing (RR), as also shown in Figure 3, the random multitude performs best because the SNs are used more evenly and there is thus less congestion. In reality, a routing algorithm which also considers traffic and queue-lengths should be used.

Finally, Figure 4 illustrates what happens when a certain number of links is removed randomly. We have chosen $\alpha = 1.25$ for this experiment because such a random multitude just has a shorter average path length than both a 2D and 3D CA, as Figure 2 shows. As one can see, the latency of the random multitude is the lowest and is basically unaffected by the random removal of a limited number of links. We used random routing (RR) to illustrate an extreme case. The 2D CA is not shown because it performs much worse than both the 3D CA and the RM. As one can see, the number of hops is affected by the link removal in a similar way than the latency.

C. Discussion

We have seen that small-world power-law networks perform better and are more robust than 2D and 3D local-neighborhood interconnects in our simple network-on-chip framework. The main reason for this are the few long(er)-distance connections, which provide short-cuts in the network. Compared to a random network (which also has the small-world property), the small-world power-law networks we use are more fabrication

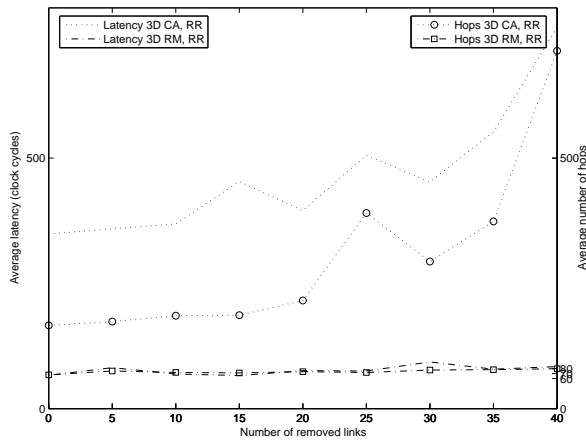


Fig. 4. Latency and hops as a function of the number of removed links between SNs. Random routing (RR), $\alpha = 1.25$, average values over 4 runs.

friendly and very resource economical because they only use a very limited number of longer-distance connections. As Jespersen and Blumen [7] state, networks with $\alpha < 2$ differ significantly from those of regular lattices, which our experiments confirm. We have further seen that small-world power-law networks are more robust with respect to link deletions. Petermann and De Los Rios [13] have further shown that the mean distance increases by removing links and that the system becomes more fragile as α increases. Finally, as our experiments show, 3D local-neighborhood interconnects require a lower number of hops and have a lower average latency than their 2D counterpart. 3D fabrics thus presents an obvious solution to interconnect problems. It has also been shown elsewhere [3] that the average total wire lengths are shorter and that fewer and shorter semi-global and global wires are required in 3D interconnects.

V. CONCLUSION

We have investigated several relevant metrics of three types of speculative, yet physically plausible system-on-chip computing architectures, namely 2D and 3D local-neighborhood as well as random small-world power-law interconnects. The small-world power-law architecture is physically realizable, could be built very economically by self-assembling mechanisms, has great transport characteristics, and is robust against link failures. While regular and local-neighborhood interconnects are easier and more economical to build than interconnects with lots of global or semi-global long-distance connections, they have been proven not to be as efficient for global communication, which is very important and also directly affects how efficient problems can be solved in general. Small-world networks with a uniform distribution of long-distance connections or pure random networks, on the other hand, are not physically plausible, although they have very interesting properties. As our results have shown by means of our simplistic, yet realistic framework, small-world power-law interconnects offer a unique balance between performance, robustness, and physical plausibility.

We believe that computation in random self-assemblies of components (e.g., [17]) is a highly appealing paradigm, both from the perspective of fabrication as well as performance and robustness. This is obviously a radically new technological and conceptual approach with many open questions. For example, there are basically no methodologies and tools that would allow (1) to map an arbitrary architecture on a randomly assembled physical substrate, (2) to do arbitrary computations with such an assembly, and (3) to systematically analyze performance and robustness within a rigorous mathematical framework. There are also open questions regarding fabrication techniques.

Future work will concentrate on the computational aspects of such assemblies and not solely on the interconnects, as in the present work. We also plan to evaluate further measures, such as energy consumption and area used, and to develop special localized and adaptive routing strategies.

REFERENCES

- [1] International technology roadmap for semiconductors (ITRS). Semiconductor Industry Association, <http://public.itrs.net>, 2003.
- [2] L. Benini and G. de Micheli. Networks on chips: A new SoC paradigm. *IEEE Computer*, 35(1):70–78, 2002.
- [3] J. A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif, and J. D. Meindl. Interconnect limits on gigascale integration (GSI) in the 21st century. *Proceedings of the IEEE*, 89(3):305–324, 2001.
- [4] A. DeHon. Array-based architecture for FET-based nanoscale electronics. *IEEE Transactions on Nanotechnology*, 2(1):23–32, 2003.
- [5] R. Ho, K. W. Mai, and M. A. Horowitz. The future of wires. *Proceedings of the IEEE*, 89(4):490–504, 2001.
- [6] R. Ferrer i Cancho, C. Janssen, and R. V. Sole. Topology of technology graphs: Small world patterns in electronic circuits. *Physical Review E*, 64:046119, 2001.
- [7] S. Jespersen and A. Blumen. Small-world networks: Links with long-tailed distributions. *Physical Review E*, 62(5):6270–6274, 2000.
- [8] J. K. Kleinberg. Navigation in a small world. *Nature*, 406:845, 2000.
- [9] B. Kozma, M. B. Hastings, and G. Korniss. Diffusion processes on power-law small-world networks. *Physical Review Letters*, 95:018701, 2005.
- [10] J. D. Meindl. Interconnect opportunities for gigascale integration. *IEEE Micro*, 23(3):28–35, 2003.
- [11] B. Mesot and C. Teuscher. Deducing local rules for solving global tasks with random Boolean networks. *Physica D*, 211(1–2):88–106, 2005.
- [12] P. P. Pande, C. Grecu, M. Jones, A. Ivanov, and R. Saleh. Performance evaluation and design trade-offs for network-on-chip interconnect architectures. *IEEE Transactions on Computers*, 54(8):1025–1040, 2005.
- [13] T. Petermann and P. De Los Rios. Spatial small-world networks: A wiring-cost perspective. [arXiv:cond-mat/0501420](https://arxiv.org/abs/cond-mat/0501420), 2005.
- [14] T. Petermann and P. De Los Rios. Physical realizability of small-world networks. *Physical Review E*, 73:026114, 2006.
- [15] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgtag. Organization, development, and function of complex brain networks. *Trends in Cognitive Sciences*, 8(9):418–425, 2004.
- [16] M. Tomassini, M. Giacobini, and C. Darabos. Evolution and dynamics of small-world cellular automata. *Complex Systems*, 15(4):261–284, 2005.
- [17] J. Tour, W. L. Van Zandt, C. P. Husband, S. M. Husband, L. S. Wilson, P. D. Franzon, and D. P. Nackashi. Nanocell logic gates for molecular computing. *IEEE Transactions on Nanotechnology*, 1(2):100–109, 2002.
- [18] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442, 1998.
- [19] H. Ye, Z. Gu, T. Yu, and D. H. Gracias. Integrating nanowires with substrates using directed assembly and nanoscale soldering. *IEEE Transactions on Nanotechnology*, 5(1):62–66, 2006.
- [20] V. V. Zhirmov and D. J. C. Herr. New frontiers: Self-assembly in nanoelectronics. *IEEE Computer*, pages 34–43, January 2001.