MP3: Minimizing Performance Penalty for Power-gating of Clos Network-on-Chip

Lizhong Chen[1], Lihang Zhao[2], Ruisheng Wang[1], and Timothy M. Pinkston[1]
[1]*Ming Hsieh Department of Electrical Engineering,* [2]*Information Sciences Institute,*
University of Southern California, Los Angeles, USA
{lizhongc, lihangzh, ruishenw, tpink}@usc.edu

Abstract

Power-gating is a promising technique to mitigate the increasing static power of on-chip routers. Clos networks are potentially good targets for power-gating because of their path diversity and decoupling between processing elements and most of the routers. While power-gated Clos networks can perform better than power-gated direct networks such as meshes, a significant performance penalty exists when conventional power-gating techniques are used. In this paper, we propose an effective power-gating scheme, called MP3 (Minimal Performance Penalty Power-gating)*, which is able to achieve minimal (i.e., near-zero) performance penalty and save more static energy than conventional power-gating applied to Clos networks. MP3 is able to completely remove the wakeup latency from the critical path, reduce long-term and transient contention, and actively steer network traffic to create increased power-gating opportunities. Full system evaluation using PARSEC benchmarks shows that the proposed approach can significantly reduce the performance penalty to less than 1% (as opposed to 38% with conventional power-gating) while saving more than 47% of router static energy, with only 2.5% additional area overhead.*

1. Introduction

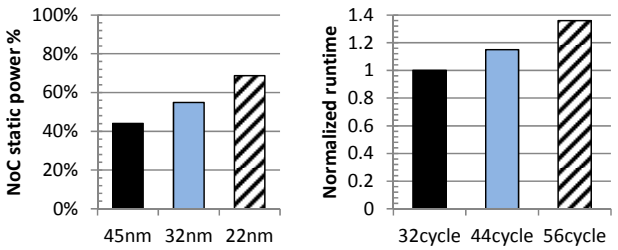
With tightening power constraints and growing demand for high performance, current and future chip multiprocessors (CMPs) need to be designed to optimize both power and performance. As a key component in CMPs for connecting various on-chip resources, the network-on-chip (NoC) can draw a substantial percentage of chip power [1, 9, 10, 26]. In particular, the static power consumption of routers accounts for an increasing percentage of the total NoC power, exceeding 43% for 45nm and beyond. As more cores are integrated on a CMP, the need to reduce on-chip latency will become even more pronounced so as not to degrade system performance. It is thus imperative to devise effective techniques that can dramatically reduce NoC static power without sacrificing performance.

Power-gating is a very useful circuit-level technique to enable trade-offs between static power and performance, especially for circuit blocks that exhibit enough idleness [11]. On-chip routers are potentially good targets for power-gating because of their relatively low average utilization, but recent research shows that it is difficult to power-gate mesh networks effectively [7, 23]. Due to the processor node-router dependence (i.e., sending/receiving packets from/to the local processor node depends on the powered-on status of the connected router), the idle periods of routers in mesh networks are often fragmented and not long enough to compensate for power-gating energy overhead. Moreover, due to limited path diversity (particularly with dimension-order routing), packets are likely to encounter gated-off routers on path(s) to the destination, in which case packets suffer additional latency to wait for routers to wake up, resulting in serious performance degradation. These fundamental but inherent limitations of meshes greatly reduce the usefulness of applying power-gating to network routers.

In this paper, we investigate the largely unexplored power-gating opportunities of Clos networks. The Clos topology has been used in off-chip networks for supercomputers and data-centers [30], and recent studies show promise for adopting Clos as on-chip networks [14, 15, 16, 34]. As Clos belong to the general class of indirect networks, the majority of the routers are not coupled to PEs. Also, Clos have excellent path diversity that can increase the chances of packets avoiding wakeup latency. While power-gating Clos can mitigate the energy and performance overhead compared with meshes, there can still be a significant performance penalty if power-gating is conventionally applied to Clos even with state-of-the-art optimizations (e.g., 38% increase in average packet latency and 15% increase in execution time, as shown in Section 6).

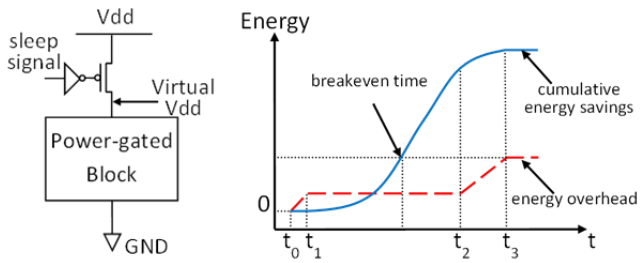
To fully exploit the potential of power-gating when applied to Clos networks, we propose an effective power-gating scheme called *MP3 (Minimal Performance Penalty Power-gating)* for Clos NoCs which can achieve minimal (i.e., near-zero) performance penalty and, at the same time, save more static energy than conventional power-gating. MP3 consists of three techniques which, collectively, are able to completely remove the wakeup latency from the critical path of packet transport and reduce long-term as well as transient contention that may occur during changes in traffic load, thereby addressing all the major sources of performance degradation associated with power-gating. Furthermore, MP3 can steer network traffic based on load conditions and actively create additional power-gating opportunities in a coordinated fashion, thus improving overall energy savings. Full system simulation shows that, compared to an optimized conventional power-gating technique applied to Clos, MP3 achieves a reduction of 36.8% in network performance penalty while saving 9.8% more router static energy. When compared with not using power-gating, MP3 reduces router static energy by 47.7% while incurring only 0.65% increase in execution time.

This research increases understanding of the key factors affecting the effectiveness of power-gating on-chip network routers. The proposed scheme and simulation results provide valuable insights on how to address critical performance and energy issues. While both mesh and Clos networks are evaluated, the main objective is not to establish that one topology is better than the other but, rather, to in-



(a) NoC static power % vs. process technology (b) Normalized runtime vs. average packet latency

Figure 1: Need for reducing NoC static power without increasing packet latency.



(a) Power-gating concept (b) Energy savings and costs

Figure 2: Power-gating technique and breakeven time.

vestigate effective architectural solutions for Clos networks from the perspective of power-gating. To our knowledge, this is the first study to explore power-gating trade-offs for indirect networks such as Clos and demonstrate the viability of realizing minimal performance penalty when applying power-gating to NoCs.

The rest of the paper is organized as follows. Section 2 provides more background on power-gating and identifies fundamental limitations in power-gating mesh networks. Section 3 analyzes the opportunities and challenges of power-gating Clos networks. Section 4 provides details of the proposed MP3 design for optimizing power-gating. Section 5 discusses evaluation methodology, and Section 6 presents simulation results. Finally, related work is summarized in Section 7, and Section 8 concludes the paper.

2. Background and Motivation

2.1 Need for Performance-aware Static Power Reduction

While on-chip networks provide a more scalable interconnection solution for many-core CMPs compared with traditional buses and point-to-point interconnects, the added complexities of buffers, crossbars and control logic in the NoC greatly increase power demand. Industrial and research chips have shown that on-chip networks can draw a substantial percentage of chip power [1, 9, 10, 26]. In particular, a large percentage of the NoC power consumption comes from static power which is trending upward as technology scales. To illustrate, Figure 1(a) plots the percentage of static power of a 64-node NoC at 2GHz for different process generations. Results are obtained from the latest DSENT [32] NoC power simulator fed with statistics from full system simulation (detailed simulation infrastructure is described in Section 5). As shown in the figure, the percentage of static power consumption increases continuously as the transistor feature size shrinks, from 43% at 45nm, to 54% at 32nm, to over 65% at 22nm under representative workloads. This trend only gets worse as technology scales beyond 22nm, indicating a pressing need to reduce NoC static power.

The design of the on-chip network is key to supporting fast communication among various on-chip resources. Care should be taken when trading off NoC performance for power-savings as any non-local data access, coherence messaging and handshaking signaling relies on the on-chip network which is critical to maintaining system performance. Figure 1(b) from our simulations show that, on average, the runtime of PARSEC benchmarks on a 64-node CMP is increased by 15% and 36% when the average on-chip packet latency increases from 32 cycles to 44 cycles and 56 cycles, respectively. With more cores integrated on a chip in the near future, the on-chip network will have an even larger impact on system performance. Given the worsening problem of static power consumption and the growing importance of low packet latency, it is imperative to design effective techniques that can dramatically reduce NoC static power without sacrificing performance.

2.2 Power-gating and Associated Trade-offs

Power-gating is a promising technique for enabling tradeoffs between static energy savings and performance. As depicted in Figure 2(a), it is implemented by inserting appropriately sized header (or footer) transistor(s) – a non-leaky "sleep switch" with high threshold voltage – between Vdd and the block (or the block and GND). By asserting the sleep signal when the power-gated block is idle, the supply voltage to the block can be turned off, thus avoiding static power consumption by removing the leakage currents in both subthreshold conduction and reverse-biased diodes.

The effectiveness of power-gating is determined by two aspects: net energy savings and performance penalty.

Net energy savings: Static energy can be saved during the power-gated period. However, there are energy overheads that come from distributing the sleep signal at the beginning of each power-gating operation (from t_0 to t_1 in Figure 2(b)) and from waking up the gated-off block at the end (from t_2 to t_3 in Figure 2(b)). Consequently, power-gating is useful only when the cumulative static energy savings exceed the energy overhead. This condition of positive net energy savings is reflected in the concept of "breakeven time" (BET) defined to be the minimum number of consecutive cycles that a gated block needs to remain in idle state before being awoken to offset power-gating energy overhead [11, 21, 22]. For on-chip routers, the BET value is around 10 cycles as estimated in prior research using analytical modeling and simulation [5, 11, 23].

Performance penalty: Despite the net energy savings that can come from power-gating, a potential drawback is the detrimental impact on system performance it may have. Whenever a gated-off block needs to be used again, it first has to be awoken by restoring virtual Vdd. Under typical technology parameters, the wakeup latency for on-chip routers is usually a few nanoseconds (or around 5-15 cycles depending on the frequency) according to previous studies

[7, 23, 25]. Since a powered-off block cannot perform the assumed operations until it becomes fully functional, stalling in the system may occur that can result in serious performance penalty if power-gating is performed frequently and the gated-off periods are short.

Therefore, to utilize power-gating effectively, we need to maximize net energy savings by increasing the idleness of unneeded functional blocks and, at the same time, minimize performance penalty by partially or even completely reducing/hiding the wakeup latency.

2.3 Limitations in Power-gating Mesh Networks

Power-gating has been applied successfully in cores and execution units [11, 21, 22] for some time and shown to enable viable trade-offs between performance and energy. Only recently has research efforts started to consider the application of power-gating in on-chip network routers [5, 7, 23, 24, 25, 29], all of which assume mesh-based topologies. Owing to its planar topology, the mesh is a popular network used in chip multiprocessors. However, there are several fundamental limitations in applying power-gating usefully to meshes and other direct networks. As shown in Figure 3(a), in direct networks such as the mesh, every router (denoted by the labeled square) is connected to a processing element (PE, denoted by the circle); whereas in indirect networks such as the Clos of Figure 3(b), only the input and output routers at the edge of the network are associated with PEs, so that packets sent from PEs are forwarded indirectly through the middle-stage routers. Compared with Clos, there are two distinctive properties of mesh networks that greatly limit the effectiveness of applying power-gating: 1) dependence between each PE-router pair and 2) less path diversity.

From the energy perspective, due to the PE-router dependence, a mesh router must be awoken whenever the connected PE needs to send a packet to the network or receive a packet from the network, thus breaking the potentially long idle period of the router into fragmented intervals that may fall below the required BET. Moreover, the BET limitation is further intensified in meshes due to the fewer alternative paths as more non-local packets have to be forwarded through the local router, making the idle intervals even shorter. For example, any packet sent from router 0-5 in Figure 3(a) needs to be forwarded through router 6 to get to router 7 assuming a minimal routing algorithm. Our full system evaluation on PARSEC benchmarks shows that, for an 8x8 mesh, the number of idle periods having a length less than the BET constitutes more than 67.2% of the total number of idle periods, which severely limits the potential to achieve large net energy savings.

In addition, from the performance perspective, power-gating of mesh routers can have a considerable negative impact on NoC performance. When a PE needs to send/receive a packet, due to the PE-router dependence, a wakeup is inevitable if the associated router is in the powered-off state, and the wakeup latency is exposed directly to the critical path of the packet's transport to the next hop. Furthermore, a packet routed over multiple hops can experience wakeup latency multiple times as routers at many hops along the path could be gated-off. This *cumulative wakeup latency* problem is severe in meshes as there are few alternative node-disjoint paths from which to choose at any particular hop.

To improve the effectiveness of power-gating mesh networks, several optimization techniques can be used. However, they all have limited capability in mitigating the above energy and performance issues. For example, *early-wakeup* signal generation [23] can only hide up to 3 cycles of the entire wakeup latency, assuming a canonical 3-stage router with look-ahead routing. The *Idle-detect* [11] technique can usually only filter out idle intervals that are shorter than around 4 cycles [7] without substantially losing static power saving opportunities. It is also possible to implement power-gating for smaller circuit blocks within each router, such as per input port or per virtual channel [24, 25]. However, individual components have only slightly longer idle periods, and this method requires prohibitive implementation overhead (e.g., 35 power domains are needed in a single router [25] to implement this method in addition to the complex coordination among different components). These techniques have only limited effectiveness as they can neither remove the inherent dependence between the PE and router in a mesh nor increase path diversity. We address these issues by exploring power-gating on another class of topology that expands the possibility for power-performance tradeoffs.

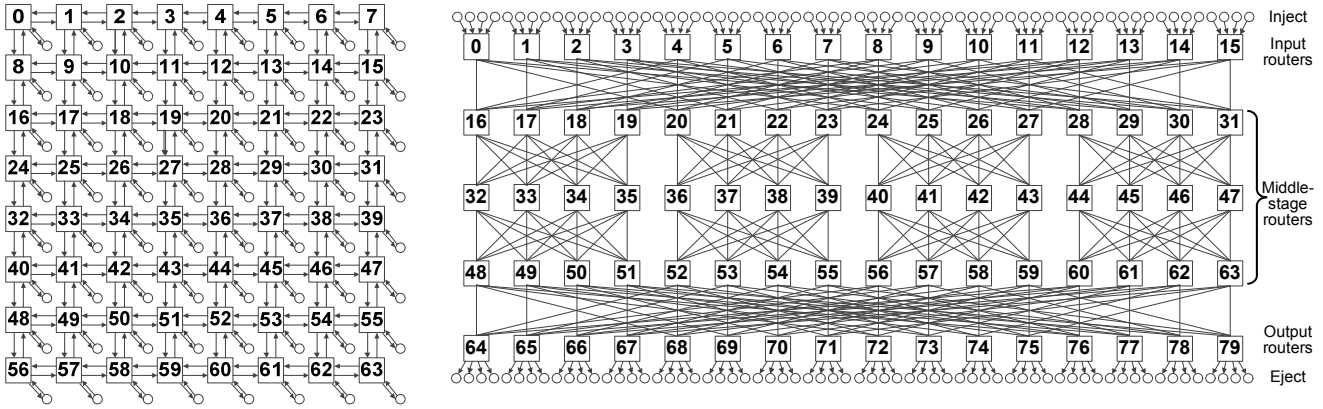
3. Analysis of Power-gating Clos NoC

3.1 Clos Networks

Whereas most of the NoC power-gating work to-date tries to combat critical problems in applying power-gating to mesh networks, very little research has explored the opportunities of power-gating Clos NoCs belonging to the large class of indirect networks. The Clos topology has long been studied since first being proposed in 1953 [6]. Early applications of Clos were for circuit switching in telephone exchange systems due to the topology's superior capability for establishing many concurrent connections. More recently, packet-switched Clos and its variants have been proposed for off-chip networks in supercomputers as well as on-chip networks for chip multiprocessors [16, 30, 34].

A packet-switched Clos network consists of three types of routers: input routers (IRs) that receive input packets from PEs through the injection channels, output routers (ORs) that output packets to PEs through ejection channels, and middle-stage routers (MRs) that do not connect to any PE and only perform forwarding functions. In general, a Clos network can be made of any odd number of stages. Figure 3(b) shows an example of a 5-stage Clos composed of 4x4 routers to connect 64 PEs using unidirectional links. The respective top and bottom PEs are the same repeated for simplicity of representation, per usual convention.

In the past, the main concern for adopting Clos NoCs was long wires. With specialized floor-planning optimizations to reduce total wire length of the Clos [16, 33] and with routing-over-logic techniques to largely remove the area overhead of long wires [28, 34], the hardware complexity of Clos NoCs can be greatly mitigated. Moreover, Clos also has the flexibility to be implemented with lower radix routers (e.g., 2x2 router) to increase clock frequency or with higher radix routers (e.g., 8x8 routers) to reduce hop count, making Clos very competitive to other traditional and ad



(a) Mesh (direct network, 64 5x5 routers) (b) 5-stage Clos (indirect network, 80 4x4 routers)

Figure 3: Direct network (mesh) vs. indirect network (Clos) for connecting 64 PEs; all links are unidirectional. PE-router dependence only at input and output routers in indirect networks.

hoc topologies [16] (Section 6.8 provides more discussion). These recent optimizations and flexibility on implementation make it interesting to explore Clos NoCs and their power-savings capabilities.

3.2 Opportunities

As an indirect network, Clos has at least three major advantages for applying power-gating. First, except for the input and output routers, all the middle-stage routers are not coupled to PEs. Therefore, sending and receiving packets in PEs do not necessarily trigger the wakeup of most routers. This not only reduces the number of router wakeups but also mitigates the energy overhead and delay associated with the wakeup. It also increases the chances of routers being idle longer than the required BET.

Second, for a given network size, the number of stages in a reasonably designed Clos is usually smaller than the average hop count in a mesh. As routers at each hop could be gated-off, the Clos topology can essentially alleviate the aforementioned cumulative wakeup latency problem by reducing the total number of encountered routers that are in gated-off state. The result is accelerated packet forwarding.

Third, the Clos provides path diversity, so that in-transit packets have multiple routing options and can avoid waiting for router wakeup as long as one of the downstream routers allowed by the routing algorithm is not in the gated-off state. It is possible, in theory, for packets to avoid all wakeups along a packet's entire path from source to destination, thereby eliminating the wakeup delay and minimizing the overall performance penalty of applying power-gating.

3.3 Challenges

Although the above opportunities suggest that indirect Clos networks are promising candidates for power-gating, applying the circuit-level power-gating technique conventionally (or *conventional power-gating* for short) to Clos can have limited effectiveness, especially in terms of reducing performance penalty. Our simulations show that conventional power-gating of the Clos, even with early-wakeup and idle-detect optimizations mentioned in Section 2.3, can still incur 38% increase in average packet latency and 15% increase in execution time. This significant performance penalty is caused by a number of reasons, as explained below.

First, as can be observed, besides the middle-stage routers, there are still a sizable number of input or output routers (e.g., 32 out of 80 routers in the 5-stage Clos). Similar to the mesh, these routers connect to PEs directly, thus suffering from the same problem: either the router idleness is upper-bounded by the local PE's traffic, or packets from/to PEs have to experience the wakeup latency of the directly associated router. Therefore, a way to allow packets to be forwarded through the input and output routers with low overhead is needed while allowing part or all of the static energy of these routers to be saved.

Second, even though the Clos has better path diversity and smaller average hop count, the wakeup latency is still on the critical path of packet transport. Moreover, the cumulative wakeup latency remains as packets at some particular routers may be left with one unique path to the destination. For instance, there are 16 different paths from *R1* to *R64* overall (we use *Ri* to denote the labeled router in Figure 3). However, if a packet is currently in *R32* and destined to PEs connected to *R64*, then the only reachable path is *R32 => R48 => R64*. If both *R48* and *R64* are powered-off, the packet will experience wakeup latency twice with no alternative paths. To make things worse, power-gating saves more static energy when network load is low, in which case routers are also likely to be powered-off, making packets more likely to encounter multiple wakeups. One effective approach to solve this problem is to completely remove the wakeup latency from the critical path by always providing a minimal set of carefully selected powered-on paths between any PE pair, as proposed in the next section.

Third and most importantly, conventional power-gating of the Clos is uncoordinated in the sense that every router makes routing decisions unaware of the global network status, thus switching between powered-on and off states independently based only on local traffic information. This wastes energy-saving opportunities and incurs unnecessary performance penalties in various ways. For example, even when the overall network load is low, packets in the upstream router can still be routed to multiple downstream routers, requiring more powered-on routers that could otherwise be gated-off. Also, due to the unhidden portion of the wakeup latency, if a gated-off router starts to wake up only

after it receives a wakeup signal, the router will not be ready by the time the packets actually arrive, unless some hints about the traffic between the up and downstream routers can be exchanged in advance. In addition, as some routers in the network may be in sleep state, a sudden increase in the amount of injecting packets are forwarded temporarily only through the remaining powered-on routers, which may cause transient congestion and performance degradation until more routers are gradually awoken. To address these issues, we need a more coordinated way to efficiently power-gate all the routers in the network.

In summary, while Clos networks have great potential to reap energy benefits without incurring excessive performance overhead, this is hard to achieve through conventional power-gating approaches but, instead, requires considerable support at the architecture level as proposed in this work.

4. Minimal Performance Penalty Power-gating

In this section, we propose an effective power-gating scheme called *MP3 (Minimal Performance Penalty Power-gating)* for Clos NoCs which is able to achieve minimal (i.e., near-zero) performance penalty and, at the same time, save more static energy than conventional power-gating. The basic idea is to first guarantee network connectivity by constructing a minimum resource set that is always powered-on so that regardless of the on/off status of other resources, packets always have the last resort of using this resource set for transporting packets without suffering any wakeup latency. Then, *dynamic traffic diversion* actively steers traffic between the minimum and maximum available resources of the network in a coordinated fashion based on load conditions. In this way, contention at any particular load level is kept low while more resources can be powered-off through increased power-gating opportunities. Finally, *rapid wakeup* further reduces any transient contention that may occur during sudden load increases by powering on a selective and necessary set of downstream routers in advance. This enables those routers to be ready when packets arrive. The following subsections describe these techniques in detail.

4.1 Guaranteed Connectivity

To minimize the performance penalty of power-gating, the foremost task is to remove wakeup latency from the critical path of packet transport. We achieve this by providing guaranteed connectivity in the Clos. The basic idea is to turn ON a minimal set of resources, *S*, to ensure that at least one powered-on path always exists between any source and destination PE pair. The set *S* can be composed of routers or components within routers. As this set of resources is always ON regardless of the network load, the key is to minimize *S* to maximize energy savings, with low implementation overhead. We use the example in Figure 4 to explain our method of constructing *S*. The procedure is generally applicable to other Clos instances.

There are two main steps. The first step is to reduce the number of powered-on routers in the NoC to a minimum, and the second step is to reduce the amount of ON components within that minimum set of routers. Specially, as can be seen immediately from the figure, no PE is disconnected even if all the black routers are gated-off. Hence, the resources associated with the 39 black routers are not needed in *S*. However, this is not sufficient as every input or output router is still needed. To reduce *S* further, notice that when all the black routers are turned off, each input router only needs to forward packets from four input ports to one output port (e.g., *R0* only forwards packets to *R16*). Based on this observation, we split the resources of input routers into two power domains.

As depicted in Figure 5, the striped components are in one power domain and are needed in *S*, whereas the rest of the components form the other power domain. Essentially, to maintain the connectivity from four input ports to one output port, only one of the four 4-to-1 multiplexers in the crossbar is needed in *S*. Also, only one virtual channel (VC) for each message class is needed in an input port to correctly buffer packets without message-dependent deadlock. In general, assuming the original router has m dependent message classes, p input ports, and v VCs per class per port, the minimal number of VCs needed in *S* is $m \times p$ – one VC for each message class per port. Hence, the amount of VC resources in *S* is $1/v$ of the total VC resources. The higher the value of v, the more static energy that can be saved. In most wormhole routers, the value of v is typically two or more in order to mitigate head-of-line blocking effectively (e.g., Intel's 48-core SCC chip has 8 VCs for two message classes [10]). In addition to the four input ports and one output port, we also conservatively put all the router arbitrator components into *S* given that arbitrators usually consume a very small portion (less than 5%) of the total router energy. A two-domain separation for router arbitrators can also be used if some customized router designs employ very large arbitrators. The two-domain split approach incurs much lower hardware overhead than implementing power-gating at per port or per VC level, and allows the majority of router components to be powered off without losing the required forwarding functionality.

Likewise, *R16-R19* perform the same 4-to-1 minimal forwarding and can follow the same two-domain design. Similarly, all the output routers and *R48-R51* only need to forward packets from one input port to four output ports, so the minimal resources in *S* for these routers include one input port with m (out of $m \times v$) VCs, one-fourth of the crossbar, four output ports with m (out of $m \times v$) latches per port, and control logic. All the remaining resources are put in the other power-domain.

Overall, the above approach based on identifying a minimal resource set enables a wide range of power-gating configurations. At one end of the spectrum, all 80 routers can be turned on to support high network load during data-intensive phases of an application's execution. At the other end, when the load intensity allows it, only 1 router (white) needs to be fully powered-on while 40 routers (gray) can be partially powered-off and 39 routers (black) can be fully powered-off, allowing maximum static energy savings. More importantly, network connectivity is guaranteed at all times, so that any packet can always use the resource set *S* as the last resort for transporting packets regardless of the on/off status of other resources, thus eliminating wakeup latency from the critical path of packet forwarding.

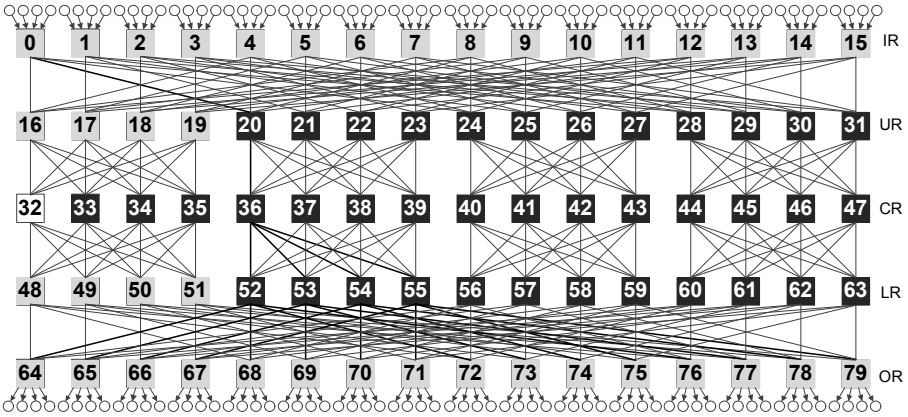


Figure 4: Illustration of MP3 for guaranteeing connectivity, dynamically steering traffic, and reducing transient contention by rapid wakeup.

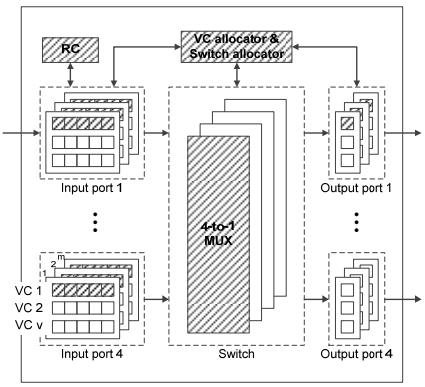


Figure 5: Partially-on design for saving energy while maintaining minimal forwarding functions.

4.2 Dynamic Traffic Diversion

While the guaranteed connectivity approach lays the foundation for effective power-gating of Clos NoCs, it accomplishes only part of our objective as packets are not automatically concentrated to only those resources needed for a specific load. In order to perform power-gating in a more coordinated fashion, we propose *dynamic traffic diversion* which systematically steers traffic to certain resources based on prevailing load conditions to 1) allow non-essential resources to be powered off via concentration and 2) gradually power on more resources as load increases to reduce contention and balance performance via distribution. To achieve these objectives, an appropriate metric is first selected for monitoring traffic intensity and then, based on the load status, the routing algorithm is augmented to enforce a network-wide consistent order of concentrating and distributing traffic to resources. Finally, a handshaking mechanism is carefully designed to power on/off resources correctly. The details are explained below.

Traffic Intensity Metric: An appropriate metric is needed as an indicator of traffic intensity. R. Das, *et al.* found that several intuitive metrics are actually ineffective in assessing load status [7]. For example, the metric of average buffer occupancy per router does not perform well as some input buffers along the congested paths may be heavily occupied while the average occupancy is still low. Injection rate also is not satisfactory as there is no universal threshold that works well for all traffic patterns (e.g., uniform random and transpose saturate at different injection rates, making it difficult to choose a predetermined threshold). In addition, the average blocking delay per flit is theoretically accurate but prohibitively expensive to implement in practice. Therefore, the use of the maximum buffer occupancy as an appropriate metric is suggested [7], where the occupancy of each input port is counted, and then the maximum value among all the input ports is computed and compared with predetermined thresholds. We use a similar metric with a slight difference in that the thresholds are adjusted based on the number of powered-on VCs to make the metric suitable for both partially-on and fully-on routers. This metric allows the threshold to be determined empirically and performs well for different traffic patterns and benchmarks.

Routing: After an appropriate metric is selected, the next augmentation is to allow the routing algorithm to become aware of the load status reflected in the metric and to steer traffic accordingly. For Clos networks, packets in earlier stages have more routing freedom than packets in later stages. For example, in Figure 4, packets in input routers (IRs) or upper routers (URs) have up to four output port choices, but packets in center routers (CRs) or lower routers (LRs) can choose only one output port to reach the destination. Therefore, steering traffic is achieved during earlier stages of packet forwarding.

In the case of the example depicted in Figure 4, we assume the metric threshold is divided into four ranges to create a 4-level configuration that corresponds to increasing load conditions. The threshold increases one level when the load condition makes the packet latency exceed 15% of the zero-load latency under the router on/off configurations for the previous level. For each router belonging to IR or UR, its four downstream routers are numbered 1 to 4 from left to right. When the load condition of a router reaches level k (k = 1, 2, 3, 4), the router is allowed to forward packets to its downstream routers numbered from 1 to k, but not above k (i.e., the leftmost k downstream routers). Adaptive routing among the k options is used based on the number of available credits (or any other commonly used criteria). In this way, 1) no downstream routers are used beyond the minimally needed k routers corresponding to current load conditions and 2) among the downstream routers, utilization is maximized through load-balancing adaptive routing. At the highest load, all four downstream routers can be used in this method, which is the same as the no-power-gating case with no sacrifice in throughput. It is worth noting that, by enforcing the left-to-right order at every router, the entire network agrees on a consistent order of which resource set to concentrate or expand (e.g., when load is on level-1, *R0*, *R4*, *R8* and *R12* will consistently all forward packets to *R16*), thus avoiding the inefficiencies of uncoordinated power-gating. Also, if some routers at a particular stage, e.g., CRs, are accidently turned off, the upstream stage routers, URs, will experience higher maximum buffer occupancy and consequently wake up more downstream routers, which are the exact same stage CRs. This will restore the balance between load intensity and powered-on routers.

Handshaking: Finally, we discuss the details of the required conditions and handshaking mechanism for routers to correctly transition between power states. No extra signal is needed between up and downstream routers besides what is already provided in conventional power-gating. Based on the types of routers, there are four cases.

Case 1 – white routers: Since white routers are always powered-on, no transition is need.

Case 2 – black routers: Black routers transition from on to off if 1) the datapath of the router is empty and 2) all of the wakeup signals from its upstream routers are de-asserted. The router transitions from off to on if any of its upstream routers asserts the wakeup signal. Here, an upstream router asserts the wakeup signal to a downstream router if a packet needs to be forwarded to that router. An optimization for wakeup signal generation will be presented in the next subsection, but the conditions for state transitions are the same.

Case 3 – gray routers in IRs and URs: A gray router in this category transitions from fully-on to partially-on if 1) the metric indicates load is on level-1 and 2) the datapath to the three rightmost downstream routers are empty (any new incoming packets will be forwarded only to the leftmost downstream router after detecting the low load). The router transitions from partially-on to fully-on if the load is on level-2 or above. Note that a fully-on router does not necessarily need to forward packets to all its downstream routers.

Case 4 – gray routers in LRs and ORs: A gray router in this category transitions from fully-on to partially-on if 1) the datapath of the resources that are not in *S* is empty and 2) all of the wakeup signals from its three rightmost upstream routers are de-asserted. The router transitions from partially-on to fully-on if any of its three rightmost upstream routers asserts the wakeup signal.

4.3 Rapid Wakeup

One effect of dynamic traffic diversion is to reduce performance degradation caused by power-gating, as more resources will eventually be turned on to accommodate the increase of traffic in the long run. However, transient contention may still be possible during the time that the new resource is being awoken. For example, suppose the network load suddenly jumps from level-1 to level-2 at *R0*, so that *R0* tries to wake up *R20* to distribute the traffic. Yet, *R20* will not be ready until it is fully awake after the unhidden portion of wakeup latency, during which the packets still have to be forwarded to *R16*. Then, after *R20* is powered on, packets routed through *R20* will find that *R36* is asleep. So again, packets need to wait for *R36* to wake up, and so on. In such cases, while *R20*, *R36*, *R52* are sequentially waking up, incoming packets are queued in the input buffers along this path. When the backpressure propagates back to *R0*, most of the new packets of level-2 load are still forwarded through *R16*. This leads to transient contention since the resources from *R16* and beyond are supposed to handle only level-1 load without contention.

To avoid this type of pathological performance degradation, we propose *rapid wakeup*, which relays the wakeup signal from upstream routers to downstream routers in a chained fashion to wake up needed downstream routers in advance so that those routers will be ready when packets arrive. In order to realize rapid wakeup effectively, the key is to minimize the needed router set, which is achieved by limiting the breadth and depth of the signal relay tree from the upstream router. First, to limit the breadth of the relay tree, the wakeup signal is relayed to only one downstream router if multiple options are available. For instance, *R20* relays the wakeup signal only to *R36* as *R37*-*R39* are not additionally needed for packets to reach any destination provided that *R36* is powered on. In contrast, *R36* relays the wakeup signal to *R52*-*R55* as they are indispensable for packets to reach any destination. This is because the destination of a particular packet is unknown beforehand and, more importantly, most of the destinations will in fact be visited since a batch of packets likely will arrive due to the load increase.

Second, notice that packets themselves take a few cycles to traverse each router, so downstream routers that are several hops away do not need to be awoken too early. In general, an N_{hop}-away downstream router can wake up in time if

$$N_{hop} \times T_{link} + T_{unhidden\_wakeup} \leq N_{hop} \times (T_{router} + T_{link})$$

This reduces to

$$N_{hop\_min} = \lceil T_{unhidden\_wakeup} / T_{router} \rceil$$

which is about 2-3 hops depending on actual parameter values. This means that the relay depth only needs to be 2-3 hops. After limiting the breadth and depth, the remaining relay tree from a particular router is minimal in the sense that all the remaining relays are necessary and any delay in waking up these downstream routers will cause some performance penalty. Note that, although the gated-off time of these routers may be slightly reduced, the reduced amount is only a few cycles upper-bounded by $T_{unhidden\_wakeup}$ assuming the above formula to limit the depth while still being able to wake up in time. The majority of routers are not affected. Hence, rapid wakeup can largely remove the transient contention penalty while retaining most of the power-gating opportunities. Moreover, since the wakeup signal is required for power-gating anyway, no additional signaling network is needed.

4.4 Impact of MP3 on Performance and Energy

Putting the three techniques together, the proposed MP3 scheme fully exploits the power-gating potential offered by indirect Clos networks while effectively addressing its performance and energy challenges.

From the performance perspective, the guaranteed connectivity technique is first used to remove the wakeup latency from the critical path of packet forwarding. Then, dynamic traffic diversion is used to guard against contention in the long-run and rapid wakeup is used to reduce transient contention. Therefore, MP3 removes all the major sources of possible performance degradation, thereby minimizing performance penalty of power-gating the Clos.

From the energy perspective, guaranteed connectivity enables a wide spectrum of energy-performance tuning opportunities by constructing a minimally needed resource set. Dynamic traffic diversion then utilizes these opportunities to coordinate router power-gating by steering traffic and turning on/off resources dynamically. This not only extends the idle periods of the majority of routers, but also reduces the number of wakeups and the associated energy overhead that causes BET limitation in the first place. As a result, MP3 is able to save more static energy with less energy overhead, thus effectively increasing the net energy savings.

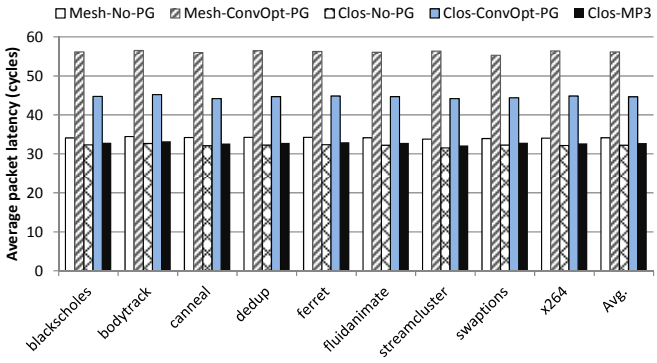


Figure 6: Average packet latency.

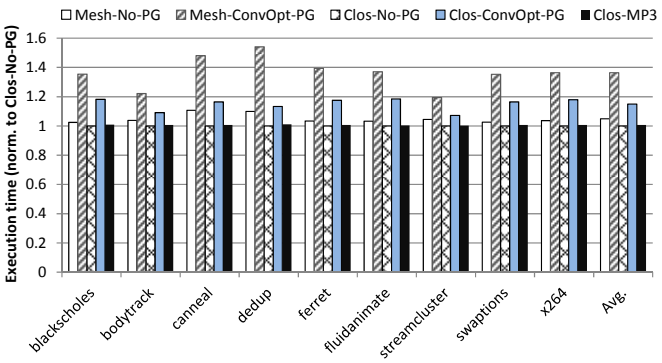


Figure 7: Execution time.

5. Evaluation Methodology

The proposed MP3 scheme is evaluated quantitatively under full system simulation with the combined use of multiple architecture-level and circuit-level simulators. Cycle-accurate SIMICS and GEMS are used for processor functional and memory timing simulation. GARNET [2] is used for detailed NoC performance evaluation, from which the network activity statistics are collected and fed into DSENT [32] for network power estimation. We modify the simulators to model all the key additional hardware in MP3, such as handshaking logic, buffer occupancy comparators, wakeup signal relay, and so on. Each PE in the network contains an UltraSPARC III+ core running at 2GHz, a 32KB I/D private L1 cache and a 256KB shared L2 cache slice. Coherence is managed by the MOESI protocol. Four memory controllers are provided. The cache and memory controller on a PE share the injection channel in the network interface. All topologies under comparison have the same bisection bandwidth of 1TB/s. To accurately reflect the link delays of the Clos, we follow the floorplan optimization in [16] to estimate every link length. GARNET is configured to have the delay of each link to be proportional to its length. We model a canonical 3-stage router with look-ahead routing [18]. Two virtual channels per message class are provided, though MP3 can achieve more energy savings with more VCs, as mentioned in Section 4.1. Also, as the Clos has more bisection links than the mesh, for comparison purposes, both the Clos and mesh networks are configured with the same total bisection bandwidth and the same total buffer sizes.

Given that the metric of maximum buffer occupancy is insensitive to traffic patterns (one of the main benefits), thresholds for congestion levels are determined empirically. However, router wakeup latency has a large impact on system performance. To estimate wakeup latency accurately, we generate the physical layout of a router at 45nm technology with 1.0V voltage using a standard VLSI design flow. Synopsys Design Compiler is used for logic synthesis, and Cadence Encounter is used to process the gate-level netlist to generate the power grid, floorplan, clock trees and routes. Parasitic extraction is performed on a 451um-by-451um layout to obtain the parasitic resistance and capacitance as well as the cell load on the Vdd wiring. Finally, the extracted data is fed into a SPICE RC model, providing a wakeup latency of 8 cycles. Because of the criticality of wakeup latency, additional sensitivity studies are also conducted to shed more light on the applicability of different schemes.

The following schemes are compared on a 64-core system: (1) Mesh-No-PG: mesh network with no power-gating; (2) Mesh-ConvOpt-PG: conventional power-gating of mesh optimized with early-wakeup and idle-detect – these optimizations not only improve performance by hiding 3 cycles of wakeup latency, but also reduce energy overhead by avoiding powering-off all idle periods that are shorter than 4 cycles; (3) Clos-No-PG: Clos network with no power-gating; (4) Clos-ConvOpt-PG: conventional power-gating of Clos with early-wakeup and idle-detect optimizations; (5) Clos-MP3: Clos with the proposed power-gating scheme. All the five schemes allow adaptive routing for fair comparison. While the mesh is included in the evaluation as a point of reference, the main objective is to evaluate the power-gating opportunities of Clos and how its power-gating potential can be exploited by our proposed scheme.

6. Results and Analysis

6.1 Impact on Performance

As one of the primary targets, we first examine the performance impact of different schemes by running multi-threaded PARSEC benchmarks [4]. Figure 6 compares the average packet latency, and Figure 7 shows the execution time of the five schemes normalized to Clos-No-PG. Results are consistent across the range of benchmarks. As Mesh-No-PG and Clos-No-PG do not use power-gating, they provide a lower bound of performance for the mesh and Clos, respectively. As can be seen from Figure 6, even with early-wakeup and idle-detect optimizations, the conventional power-gating scheme for the mesh, Mesh-ConvOpt-PG, still significantly increases the average packet latency by 64.5% on average compared with Mesh-No-PG; whereas Clos-ConvOpt-PG causes 38.6% increase in the average packet latency compared with Clos-No-PG. This indicates that the indirect network nature of Clos indeed helps to reduce performance degradation as compared to the mesh, but it still cannot entirely mitigate the negative effects of wakeup latency. In contrast, Clos-MP3 completely removes the wakeup latency from the critical path and reduces both long-term and transient contention. Consequently, Clos-MP3 achieves a remarkable reduction of average packet latency, having only 1.8% increase on average. This is equivalent to

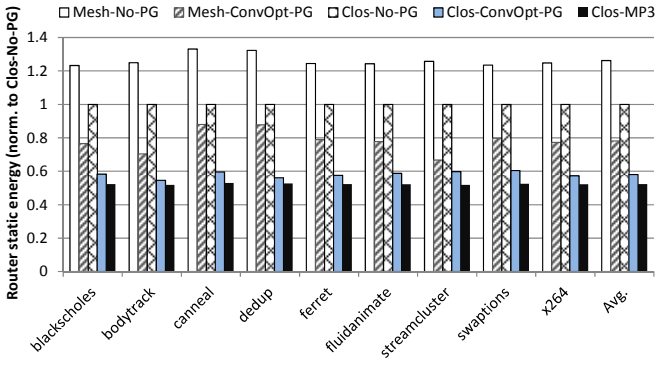


Figure 8: Router static energy.

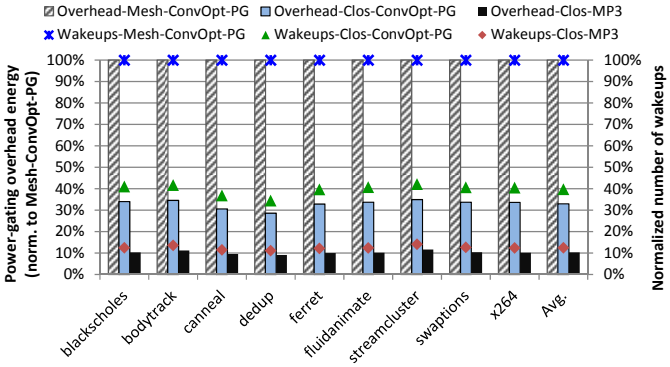


Figure 9: Comparison of power-gating overhead.

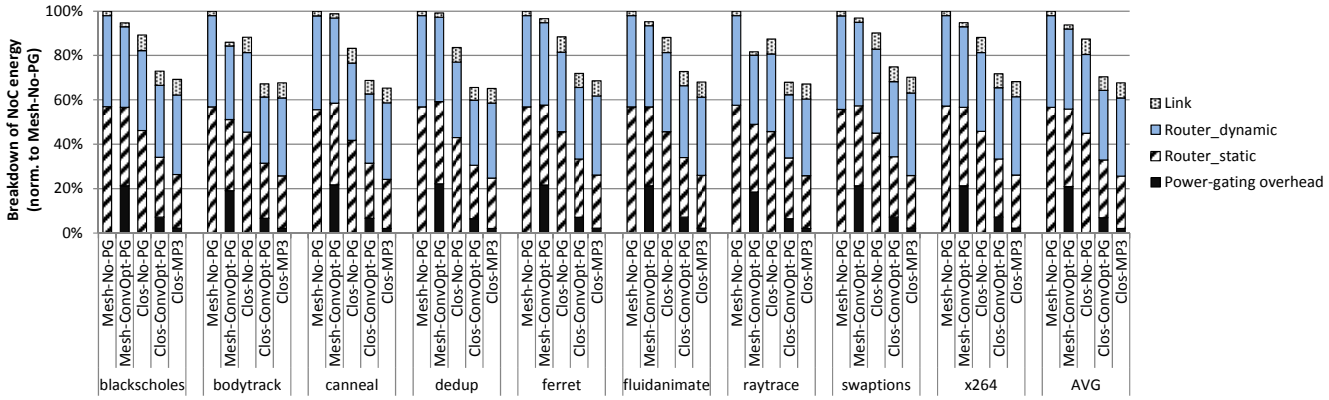


Figure 10: Breakdown of NoC energy.

a 36.8% improvement compared with Clos-ConvOpt-PG. Similar trends are also reflected in execution time. As shown in Figure 7, Mesh-ConvOpt-PG and Clos-ConvOpt-PG increase execution time by 36.3% and 14.9% on average, respectively. In comparison, the proposed Clos-MP3 incurs a minimal increase of only 0.65% in execution time, effectively realizing near-zero performance penalty.

6.2 Impact on Router Static Energy

The performance advantage of Clos-MP3 does not sacrifice its energy savings at all. Figure 8 presents the results of router static energy of different designs normalized to Clos-No-PG. As can be seen, Mesh-ConvOpt-PG reduces the router static energy by 38.2% relative to Mesh-No-PG. In comparison, Clos-ConvOpt-PG is able to reduce the static energy by 41.1% relative to Clos-No-PG, which is slightly better than Mesh-ConvOpt-PG due to the inherent suitability of Clos for power-gating. The lowest router static energy is achieved in the proposed Clos-MP3, with an average reduction of 47.7%. This improvement mainly is attributed to the ability of Clos-MP3 to dynamically concentrate traffic and actively create power-gating opportunities. When compared relatively, the proposed Clos-MP3 saves 9.8% more router static energy than Clos-ConvOpt-PG. This highlights the effectiveness of Clos-MP3 in both providing higher performance and lower energy simultaneously.

6.3 Comparison of Power-gating Overheads

To gain more insight on the feature of Clos-MP3 to reduce unnecessary wakeups by steering traffic, Figure 9 shows the energy overhead (left vertical axis) caused by router wakeup for the conventional power-gating schemes and the Clos-MP3 scheme, normalized to Mesh-ConvOpt-PG. As can be observed, the power-gating overhead in Clos-MP3 is substantially lower than the other two schemes. Figure 9 further compares the reduction in the total number of wakeups in the different schemes (right vertical axis). Whereas Clos-ConvOpt-PG decreases the number of wakeups by 60.3% compared to Mesh-ConvOpt-PG, Clos-MP3 is able to reduce wakeups by 87.6%, on average, owing to its coordinated power-gating among all routers in the network. This explains the large reduction in energy overhead and demonstrates the usefulness of Clos-MP3.

6.4 Impact on NoC Energy

Figure 10 plots the breakdown of NoC energy across the benchmarks normalized to Mesh-No-PG, showing the relative impact of each energy component. Several observations can be drawn from the figure. First, although Clos may consume more link energy than mesh for the no power-gating cases, the total NoC energy of the Clos is still lower than that of the mesh, indicating that Clos is a competitive NoC topology. Second, the large power-gating overhead in Mesh-ConvOpt-PG makes it very ineffective, leading to a less than 6.3% reduction in overall NoC energy; whereas Clos with conventional power-gating saves 19.4% of overall NoC energy. Third, the proposed Clos-MP3, while significantly reducing the performance penalty, can also save 22.5% of NoC energy. This means that, compared with the state-of-the-art power-gating scheme for Clos (i.e., Clos-ConvOpt-

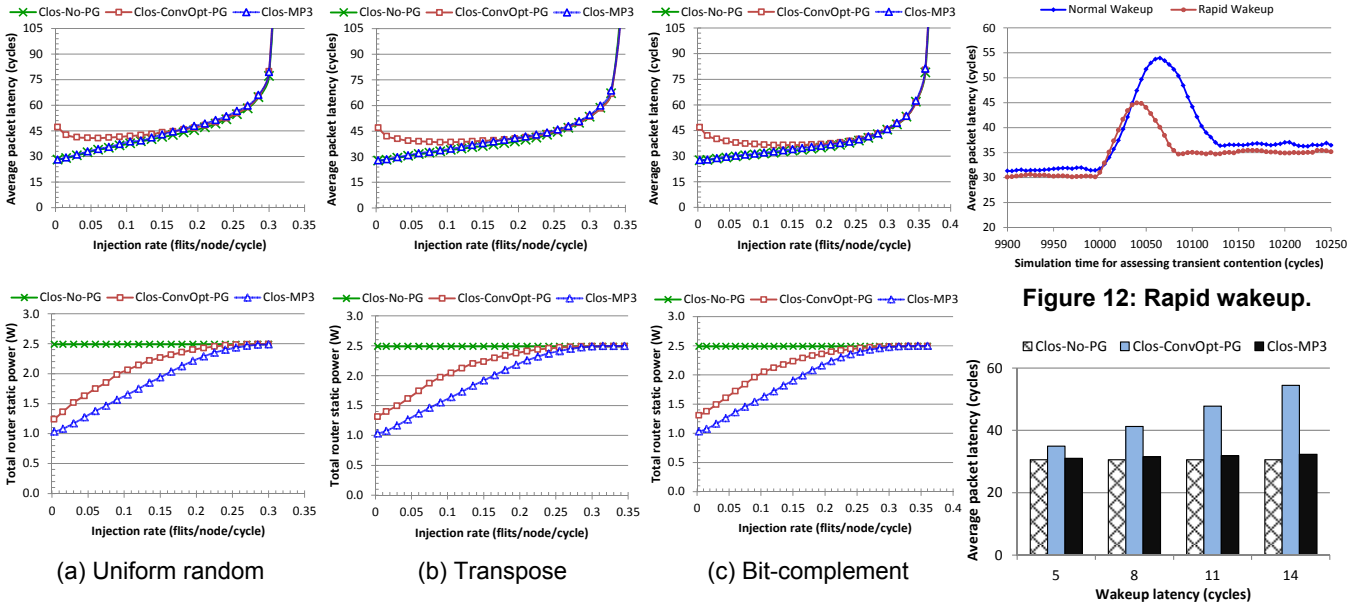


Figure 11: Behavior across full range of network loads.

Figure 12: Rapid wakeup.

Figure 13: Wakeup latency tolerance.

PG), MP3 is better in terms of both performance and energy.

6.5 Comparison across Full Network Load Range

In order to understand the behavior of different schemes more fully, we leverage synthetic traffic and vary the network load across the full range: from zero load to saturation load. Figure 11 presents the performance and power results for uniform random, transpose and bit-complement traffic patterns. On the performance side, while typical behavior is observed for Clos-No-PG, interesting results are found for Clos-ConvOpt-PG. It can be seen that, at low load, the average packet latency of Clos-ConvOpt-PG is actually very high. This is because many routers are gated-off at this load, so packets are likely to experience wakeup latency multiple times (i.e., cumulative wakeup latency), which increases the packet latency considerably. When load increases, the average packet latency first decreases as more routers are awoken, and then starts to rise again as load approaches saturation. In contrast, the average packet latency of Clos-MP3 follows Clos-No-PG closely across the entire load range, showing that it only incurs minimal performance penalty. It is important to note that Clos-MP3 can indeed reach the maximum throughput of the no-power-gating case. This means that all routers can be correctly woken up in Clos-MP3 if needed, which is important and necessary for supporting high network load phases of application execution.

When comparing static power, the proposed Clos-MP3 clearly has a significant advantage for various traffic patterns. As shown in the figure, the static power savings of Clos-ConvOpt-PG are less than 10% when the load rate only reaches 50% of saturation. In comparison, Clos-MP3 saves more than 10% of the static power even when load passes 75% of saturation. These results suggest that Clos-MP3 is much more energy-proportional than conventional power-gating.

6.6 Effect of Rapid Wakeup

We also perform simulations to demonstrate the ability of rapid wakeup to reduce transient contention. To assess this effect quantitatively, the injection rate is quickly increased from 5% to 25% when the simulation time passes the 10k-cycle mark (sufficient for reaching steady state in synthetic uniform random traffic). Figure 12 plots the changes of average packet latency as more resources are waking up to accommodate the new load. Compared with the normal wakeup, rapid wakeup mitigates transient contention in two ways: 1) rapid wakeup stabilizes the packet latency within 75 cycles, which is 42% shorter than the normal wakeup and 2) rapid wakeup also reduces the peak increase of average packet latency during the transition by 34%. Due to these features, rapid wakeup is very helpful in minimizing performance penalty of Clos-MP3.

6.7 Wakeup Latency Tolerance

As mentioned previously, cumulative wakeup latency is a big obstacle for reducing the performance overhead of conventional power-gating, particularly in multi-hop networks. The evaluation so far has shown that Clos-MP3 incurs minimal performance penalty when the wakeup latency is 8 cycles (which is obtained from our detailed circuit-level simulation). To illustrate that Clos-MP3 can effectively address the challenge of wakeup latency, Figure 13 compares the average packet latency of Clos-No-PG, Clos-ConvOpt-PG and Clos-MP3 with varying values of wakeup latency. The load rate is set to the average load rate of PARSEC benchmarks. As can be seen, the average packet latency of Clos-ConvOpt-PG increases by 56% when the wakeup latency increases from 5 to 14 cycles; whereas the latency of Clos-MP3 remains very similar (less than 3.5% increase) for different wakeup latencies. This demonstrates the ability of Clos-MP3 to hide wakeup latency and its wide applicability to various designs (e.g., under different frequencies).

6.8 Discussion

Hardware overhead: When configured with the same total buffer size and total bisection bandwidth, the hardware cost of the 5-stage Clos (80 4x4 routers plus links) is 17% lower than that of mesh (64 5x5 routers plus links), so the

Clos is a viable option for on-chip networks in terms of implementation cost. Second, for any power-gating technique, there is hardware overhead for sleep switch and state retention, which is typically within 4-10% depending on circuit-level optimizations [11, 13]. More of a concern is the additional hardware induced by Clos-MP3. Our simulation results show that the added components in Clos-MP3, including the modified routing logic, handshaking control, wakeup signal relay and so on, have a hardware overhead of less than 2.5% compared with conventional power-gating.

Scalability: The proposed Clos-MP3 does not have any particular element that limits its scalability (e.g., no central controller, no global signaling, etc.) and can be used for any size of Clos NoCs. Thus, the scalability of Clos-MP3 is only bounded by the Clos topology itself which has been shown to have similar scalability as mesh NoCs [16].

Other topologies: Thus far, a 5-stage Clos is used as a case-in-point to illustrate the proposed MP3 scheme. This Clos example is compared to a traditional mesh network that is also composed of low radix routers. When higher radix routers are allowed under design constraints (e.g., to meet certain clock frequency criteria), several other topologies are available to increase network performance. For example, with an 8x8 router radix, mesh can use a concentration degree of 4 to reduce the network diameter to 6 for a 64-node system [3]. Flattened butterfly [19] can further reduce this diameter to 2 by directly connecting the nodes in a dimension, with a router radix of 10x10. In addition, folded Clos (fat-tree) also has a network diameter of 2 with 8x8 router radix. While these topologies are able to reduce packet latency considerably, Clos remains competitive given that a 3-stage Clos can also achieve a network diameter of 2 with 8x8 router radix, resulting in a similar reduction of packet latency.

Prior research has shown that high-radix Clos has comparable hardware complexity but higher power efficiency (assuming no power-gating) than several other mainstream topologies [16]. However, in terms of static power savings potential, the aforementioned topologies (i.e., concentrated mesh, flattened butterfly and fat-tree) are much more limited by the router-PE coupling than Clos. This is because, with high-radix routers, a packet to any of the many input ports needs to wake up the router, which reduces the router idleness and causes wakeup delay. In contrast, this effect is greatly mitigated in Clos with our proposed MP3 technique. For example, all the 8 input routers and 8 output routers in a 3-stage Clos can benefit from the two-domain partial power-gating (which is even better than the 5-stage Clos as now roughly only 1/8th of the router needs to be turned on minimally). Dynamic traffic diversion also works better due to increased adaptivity (8 outputs to choose from at each router). Rapid wakeup may have reduced benefits but can still hide the majority of wakeup latency. Hence, high-radix Clos can have similar packet latency advantages as other high-radix topologies while being a better target for power gating. These findings together with the 5-stage Clos example presented in previous sections lead to the conclusion that Clos is a competitive topology for both low-radix and high-radix networks.

Applicability: The proposed MP3 scheme is an important extension to enhance the power-gating capabilities of a variety of interconnection networks. First, MP3 is applicable to both on-chip and off-chip Clos networks with different radices and network sizes. Second, MP3 can also be applied to other topologies that have multiple node-disjoint paths (excluding edge routers), which are often provided in indirect networks such as Benes, Omega, and non-flattened Butterfly with extra stages. A similar methodology of dividing the edge routers into two power domains and dynamically turning on and off edge/middle routers with wakeup signal relay can be applied. The proposed MP3 scheme, however, has limited applicability to direct networks (even with multiple paths) due to direct coupling between router and PE that may trigger the power-state to transition very frequently and because the two power domains may not be sufficient to maintain network connectivity (e.g., concentrated mesh requires all inputs and outputs to be on).

7. Related Work

A couple of related works have already been mentioned in previous sections. In addition, a multiple network-on-chip power-gating design is proposed in [7] and a power-gating bypass design is proposed in [5]. Both designs mainly target power-gating mesh networks, and the increase in average packet latency is considerable. A router parking scheme is introduced in [29] to power-gate routers in meshes when the core is idle, but it needs to flush private caches before turning off routers, which may cause serious performance issue. Some research is also conducted to power-gate individual components within a router [24, 25], but this approach is very costly (16% hardware overhead) with still limited energy-savings and non-negligible performance degradation even if the wakeup latency is only 4 cycles. Our work differs from these works in that we explore power-gating opportunities for Clos networks, and the proposed scheme is able to entirely remove wakeup latency from the critical path, thus achieving near-zero performance penalty.

Some work has gone into improving Clos for off-chip interconnects [30], and recent research has shown that it is also very promising to adopt Clos for on-chip networks [14, 15, 16, 34] as new floorplan and layout techniques emerge. However, none of the off-chip or on-chip works looked into the power-gating of Clos. Our work provides insight on the power-gating characteristics of Clos networks and helps to facilitate more efficient use of Clos NoC.

Much research has been conducted to reduce buffer requirements and improve buffer utilization, which directly or indirectly saves buffer static power [12, 27]. Aggressive bufferless routers can even eliminate buffers and their associated power consumption at the complexity of potential livelock, misrouting and packet reassembly [8]. However, besides buffers, there are other components in a router that also consume a substantial percentage of the total static power (42% as observed in our simulation), which are not addressed by the bufferless approach but can be avoided using our approach.

Prior research has also proposed various techniques to save dynamic and static power of links [17, 31]. DVFS [20] is also extensively studied to reduce power consumption. These works and other dynamic power-saving techniques (such as clock-gating) are largely orthogonal and comple-

mentary to this work, and can be used together with MP3 to provide more efficient Clos on-chip networks.

8. Conclusion

Current and future many-core systems require on-chip networks to be designed with both power and performance awareness. While mesh networks present several fundamental limitations for effective power-gating, this paper investigates the power-gating opportunities and challenges of Clos networks. To combat the various limitations and inefficiencies in conventional power-gating of Clos, a minimal performance penalty power-gating scheme (MP3) is proposed in this work. MP3 not only removes the wakeup latency from the critical path and reduces long-term and transient contention, but also actively steers network traffic to create increased power-gating opportunities coordinated globally across the network. Simulation results show significant reduction in the performance penalty while saving more router static energy than conventional power-gating. These results demonstrate the viability of using power-gating in NoCs with only minimal performance overhead.

References

[1] Adapteva. http://www.adapteva.com/epiphanyiv/, 2013.

[2] N. Agarwal, T. Krishna, L.-S. Peh, and N. K. Jha, "GARNET: A detailed on-chip network model inside a full-system simulator," in *International Symposium on Performance Analysis of Systems and Software (ISPASS)*, Boston, USA, pp. 33-42, 2009.

[3] J. Balfour and W. J. Dally, "Design tradeoffs for tiled CMP on-chip networks," in *20th Annual International Conference on Supercomputing (ICS)*, Cairns, Australia, pp. 187-198, 2006.

[4] C. Bienia and K. Li, "Parsec 2.0: A new benchmark suite for chip-multiprocessors," in *Proceedings of the 5th Annual Workshop on Modeling, Benchmarking and Simulation*, 2009.

[5] L. Chen and T. M. Pinkston, "NoRD: Node-Router Decoupling for Effective Power-gating of On-Chip Routers," in *45th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2012.

[6] C. Clos, "A study of non-blocking switching networks," *Bell System Technical Journal,* vol. 32, pp. 406-424, 1953.

[7] R. Das, S. Narayanasamy, S. K. Satpathy, and R. G. Dreslinski, "Catnap: Energy Proportional Multiple Network-on-Chip," in *40th International Symposium on Computer Architecture (ISCA)*, 2013.

[8] M. Hayenga, N. E. Jerger, and M. Lipasti, "SCARAB: A single cycle adaptive routing and bufferless network," in *42nd IEEE/ACM International Symposium on Microarchitecture (MICRO)*, New York, NY, USA, pp. 244-254, 2009.

[9] Y. Hoskote, S. Vangal, A. Singh, N. Borkar, and S. Borkar, "A 5-GHz mesh interconnect for a Teraflops processor," *IEEE Micro,* vol. 27, pp. 51-61, 2007.

[10] J. Howard, S. Dighe, S. R. Vangal, G. Ruhl, N. Borkar, S. Jain*, et al.*, "A 48-Core IA-32 Processor in 45 nm CMOS Using On-Die Message-Passing and DVFS for Performance and Power Scaling," *IEEE Journal of Solid-State Circuits,* vol. 46, pp. 173-83, 2011.

[11] Z. Hu, A. Buyuktosunoglu, V. Srinivasan, V. Zyuban, H. Jacobson, and P. Bose, "Microarchitectural techniques for power gating of execution units," in *International Symposium on Lower Power Electronics and Design (ISLPED)*, pp. 32-37, 2004.

[12] S. A. R. Jafri, Y.-J. Hong, M. Thottethodi, and T. N. Vijaykumar, "Adaptive flow control for robust performance and energy," in *43rd IEEE/ACM International Symposium on Microarchitecture (MICRO)*, Atlanta, GA, USA, pp. 433-444, 2010.

[13] H. Jiang, M. Marek-Sadowska, and S. R. Nassif, "Benefits and costs of power-gating technique," in *International Conference on Computer Design (ICCD)*, San Jose, CA, USA, pp. 559-566, 2005.

[14] A. Joshi, C. Batten, Y.-J. Kwon, S. Beamer, I. Shamim, K. Asanovic*, et al.*, "Silicon-photonic clos networks for global on-chip communication," in *3rd International Symposium on Networks-on-Chip (NOCS)*, San Diego, USA, pp. 124-133, 2009.

[15] A. Joshi, B. Kim, and V. Stojanovic, "Designing energy-efficient low-diameter on-chip networks with equalized interconnects," in *17th IEEE Symposium on High Performance Interconnects (HOTI)*, New York, NY, USA, pp. 3-7, 2009.

[16] Y.-H. Kao, N. Alfaraj, M. Yang, and H. J. Chao, "Design of high-radix Clos Network-on-Chip," in *4th International Symposium on Networks on Chip (NOCS)*, Grenoble, France, pp. 181-188, 2010.

[17] E. J. Kim, K. H. Yum, G. M. Link, N. Vijaykrishnan, M. Kandemir, M. J. Irwin*, et al.*, "Energy Optimization Techniques in Cluster Interconnects," in *Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED)*, Seoul, Korea, Republic of, pp. 459-464, 2003.

[18] J. Kim, D. Park, T. Theocharides, N. Vijaykrishnan, and C. R. Das, "A low latency router supporting adaptivity for on-chip interconnects," in *42nd Design Automation Conference (DAC)*, Anaheim, CA, USA, pp. 559-564, 2005.

[19] J. Kim, J. Balfour, and W. J. Dally, "Flattened butterfly topology for on-chip networks," in *40th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, Chicago, USA, pp. 172-182, 2007.

[20] S. Li, P. Li-Shiuan, and N. K. Jha, "Dynamic voltage scaling with links for power optimization of interconnection networks," in *9th International Symposium on High-Performance Computer-Architecture (HPCA)*, Los Alamitos, USA, pp. 91-102, 2003.

[21] A. Lungu, P. Bose, A. Buyuktosunoglu, and D. J. Sorin, "Dynamic power gating with quality guarantees," in *International Symposium on Low Power Electronics and Design (ISLPED)*, San Fancisco, USA, pp. 377-382, 2009.

[22] N. Madan, A. Buyuktosunoglu, P. Bose, and M. Annavaram, "A case for guarded power gating for multi-core processors," in *17th International Symposium on High-Performance Computer Architecture (HPCA)*, San Antonio, TX, USA, pp. 291-300, 2011.

[23] H. Matsutani, M. Koibuchi, W. Daihan, and H. Amano, "Run-time power gating of on-chip routers using look-ahead routing," in *13th Asia and South Pacific Design Automation Conference (ASP-DAC)*, Piscataway, NJ, USA, pp. 55-60, 2008.

[24] H. Matsutani, M. Koibuchi, D. Wang, and H. Amano, "Adding slow-silent virtual channels for low-power on-chip networks," in *2nd ACM/IEEE International Symposium on Networks-on-Chip (NOCS)*, Newcastle upon Tyne, UK, pp. 23-32, 2008.

[25] H. Matsutani, M. Koibuchi, D. Ikebuchi, K. Usami, H. Nakamura, and H. Amano, "Ultra fine-grained run-time power gating of on-chip routers for CMPs," in *4th International Symposium on Networks on Chip (NOCS)*, Grenoble, France, pp. 61-68, 2010.

[26] H. McGhan, "Niagara 2 opens the floodgates," *Microprocessor Report,* vol. 20, pp. 1-12, 2006.

[27] C. A. Nicopoulos, D. Park, J. Kim, N. Vijaykrishnan, M. S. Yousif, and C. R. Das, "ViChaR: a dynamic virtual channel regulator for network-on-chip routers," in *39th International Symposium on Microarchitecture (MICRO)*, Los Alamitos, USA, p. 12 pp., 2006.

[28] D. Pamunuwa, J. Öberg, L.-R. Zheng, M. Millberg, A. Jantsch, and H. Tenhunen, "Layout, Performance and Power Trade-Offs in Mesh-Based Network-on-Chip Architectures," in *VLSI-SOC*, p. 362, 2003.

[29] A. Samih, W. Ren, A. Krishna, C. Maciocco, C. Tai, and Y. Solihin, "Energy-efficient interconnect via router parking," in *19th International Symposium on High Performance Computer Architecture (HPCA), 23-27 Feb. 2013*, pp. 508-19, 2013.

[30] S. Scott, D. Abts, J. Kim, and W. J. Dally, "The BlackWidow high-radix clos network," in *33rd International Symposium on Computer Architecture (ISCA)*, Boston, MA, USA, pp. 16-27, 2006.

[31] V. Soteriou and P. Li-Shiuan, "Design-space exploration of power-aware on/off interconnection networks," in *2nd International Conference on Computer Design (ICCD)*, pp. 510-17, 2004.

[32] C. Sun, C.-H. O. Chen, G. Kurian, L. Wei, J. Miller, A. Agarwal*, et al.*, "DSENT - A tool connecting emerging photonics with electronics for opto-electronic networks-on-chip modeling," in *6th International Symposium on Networks-on-Chip (NOCS)*, Copenhagen, Denmark, pp. 201-210, 2012.

[33] Z. Wang, J. Xu, X. Wu, Y. Ye, W. Zhang, M. Nikdast*, et al.*, "Floorplan Optimization of Fat-Tree Based Networks-on-Chip for Chip Multiprocessors," *IEEE Transactions on Computers,* vol. PP, pp. 1-1, 2012.

[34] A. Zia, S. Kannan, G. Rose, and H. J. Chao, "Highly-scalable 3D CLOS NOC for many-core CMPs," in *8th IEEE International NEWCAS Conference*, Montreal, QC, Canada, pp. 229-232, 2010.