Open access • Journal Article • DOI:10.1109/TC.2012.272

Modeling and Tools for Power Supply Variations Analysis in Networks-on-Chip

— Source link

Nizar Dahir, Terrence Mak, Fei Xia, Alexandre Yakovlev

Institutions: Newcastle University, The Chinese University of Hong Kong

Published on: 01 Mar 2014 - IEEE Transactions on Computers (IEEE)

Topics: Router and System on a chip

Modelling and Tools for Power Supply Variations Analysis in Networks-on-Chip

Nizar Dahir, *Student Member, IEEE,* Terrence Mak, *Member, IEEE,* Fei Xia, Alex Yakovlev, *Senior Member, IEEE,*

Abstract—Power supply integrity has become a critical concern with the rapid shrinking feature size and the ever increasing power consumption in nanometre scale integration. In particular, on-chip communication, in platforms such as networks-on-chip (NoC), dictates the power dissipation and overall system performance in multi-core systems and embedded computing architectures. These architectures require a dedicated tool for analyzing the power supply noise which must embed distinctive communication characteristics and spatial parameters. In this paper, we present a tool dedicated for determining the on-chip V_{DD} drops due to communication workload in NoCs. This tool integrates a fast power grid model, a NoC simulator, an on-chip link model and a microarchitectural power model for router. The model has been rigorously verified using SPICE simulations. The proposed model and tools are further exemplified through analyzing the impact of power supply noise for NoC links. Statistical timing analysis of NoC links in the presence of power supply noise was performed to evaluate the bit error rates. This work would enable better understanding of the tradeoffs existing in the design of NoCs, and the induced power supply noise due to on-chip communication. This understanding is crucial for the analysis of the quality of service (QoS) of communication fabrics in NoCs at the early design stages.

Index Terms—Networks-on-chip, power supply noise, power grid simulation, on-chip routing, timing analysis, power grid granularity, probability of error, bit error rate.

1 INTRODUCTION

POWER supply noise has adverse effects on digital circuit performance and reliability. It could cause signal deterioration and create soft errors. Recently, it has been reported that variation in power supply would have significant impacts on operational frequency and system power dissipation [1], [2]. Both resistive (*IR*) and inductive (ΔI) voltage drops are sources of power supply noise. The resistive voltage drop occurs mainly due to the resistance of power delivery wires in the power grid network and increases with the amount of current delivered through these wires. On the other hand, the inductive drop is mainly due to wire inductance in the package as well as in the grid wires and is proportional to the rate of change of current.

Technology scaling exacerbates the problem of power supply noise for many reasons. Firstly, wire thickness in the power network is rapidly shrinking. This substantially increases the resistance of the power delivery wires. Also, the demand for power delivery is rapidly increasing. These facts promotes higher *IR* drop. Secondly, higher switching frequency increases ΔI drop. Thirdly, lower operating voltage decreases the noise margin. Consequently, voltage drop as a percentage of supply voltage is rapidly increasing. For example, the voltage drop can be up to 30% of nominal supply voltage in 65 nm technology if the necessary precautions are not taken [3]. Mitigating power supply noise becomes a grand challenge for the sustainability of future large-scale integration development.

Manuscript received xx, 201x; revised xx, 201x. Nizar Dahir, Fei Xia, Alex Yakovlev are with the School of Electrical, Electronic & Computer Engineering, Newcastle University, UK. E-mail: {nizar.dahir, terrence.mak, fei.xia, alex.yakovlev}@newcastle.ac.uk. Terrence Mak is with the Department of Computer Science and Engineering, Chinese University of Hong Kong email: stmak@cse.cuhk.edu.hk.

Some research efforts have focused on studying the impact of power supply noise on the performance of VLSI circuits, while others are aimed to mitigate this noise. The traditional technique for power supply noise mitigation is the use of on-chip decoupling capacitors. In [4], [5] techniques are proposed to determine the optimal values and positions of decoupling capacitors for minimizing the power supply noise during floor-planning. Optimal power-gating scheduling is also proposed to minimize the voltage drop caused by the switching activities in the gated blocks [6].

Power delivery network design optimized for Dynamic Voltage Scaling systems (DVS) is also considered in [7], where a model that uses the Markov decision process (MDP) to minimize the total energy demand when the system works in this power management mode.

The above techniques are examples of power supply noise mitigation at the circuit level. At higher levels and for multi-core systems, workload assignment can have significant impact on the induced power supply noise in the system. In [8], a simulated-annealing approach is employed to optimize assignment of workloads to the cores, such that the resulting power supply noise can be minimized. However, they considered independent tasks running on these cores and, thus, ignored intra-chip communication.

The emerging multi-core systems require dedicated and high-performance on-chip communication systems. Network-on-chip (NoC) had been proposed as a promising infrastructure to deliver scalable and high-performance on-chip communication [9].

The power budget of the NoC takes up significant overall portion in NoC-based systems. For instance, the routers in MIT-RAW CMP network consumes about 40% of the tile power, and the communication network takes up to 35% of the overall system power [10]. A measurement for the Intel's 80-tile TeraFLOPS CMP reported that the communication power budget is about 28% [11]. This imply that on-chip communication workload is responsible for a considerable portion of the overall power supply noise. In contrast to conventional models for logic or microprocessors, this portion of power supply noise would have an interesting correlation to the temporal and spatial distributions of the traffic load. This load can be determined in the early design stages, once the application characteristics are known.

Due to aggressive technology scaling, multi-core systems and, particularly, on-chip interconnection networks are more and more prone to various sources of noise. Apart from power supply noise, which is a major source of errors, process variation, crosstalk, thermal and leakage are other examples of error sources. All of these noises can cause errors and contribute to degradation of performance. Thus, error control techniques are needed for fault tolerance and to provide the quality of service (QoS) required by the target application [12], [13]. More importantly, it requires accurate estimations of error and fault rates from various sources at early design stage [14]. For independent noise sources, fault rates can be modelled separately and their effects can be added to account for a general estimation of fault tolerance metrics.

In this work, a tool for analyzing power supply noise dedicated for networks-on-chip is presented. It captures the supply voltage variations caused by communication loads across the chip. This model will allow us to better understand the tradeoffs existing in the design of communication links and, in particular, evaluate the relationship between the voltage or frequency and fault rate or bit error rate (BER). This relationship is crucial for early stage analysis of the quality of service (QoS) of communication fabrics in NoCs. The major contributions of this paper are summarized as follows:

1) Develop a tool which employs an integrated model of power supply noise in NoCs. Detailed circuit level design parameters and application-specific on-chip communication dynamics, including traffic pattern and link bandwidth, are considered. This tool provides a compact integration of NoC power and area model, an NoC simulator, on-chip link model and a power grid model.
2) Rigorous evaluation of the model accuracy and the impact of power grid granularity on this accuracy has been carried out using SPICE verifications. This, also, gives an insight on the scalability and the trade-offs between simulation time and accuracy of the power grid model.
3) The model has been employed to analyze the power supply noise in networks-on-chip. Novel observations about power supply noise distribution and variation due to different routing algorithms and traffic patterns are found.
4) The impact of the resulting power supply noise on the performance has been studied. Statistical timing analysis of the link delay caused by the power supply noise is performed. Moreover, high level fault metrics such as the probability of timing errors and bit error rates are, also, evaluated based on real world and synthetic communication scenarios.

2 Background and Related Work

NoCs are used to connect components on the same chip. The transfer of data is achieved in a way similar to conventional computer networks where packet switching is used and packets are routed from the source to the destination. A packet is split into smaller data units called *flits*. The interconnected components can be general purpose microprocessors, memory blocks or control circuitry. Each component (IP) is attached to a router which is used as a gateway to connect the IP to other IPs and to route information for the overall system. The term *tile* is often used to stand for the IP core and the corresponding router.

Many tools have been developed to model NoC power and area for early-stage design space exploration [15], [16], [17]. These tools aim to help designers evaluate their design in the early stages and explore different design strategies and techniques which will result in an initial estimation of the significance of a specific design technique. Other researchers focused on optimizing floor planning and topology [16], [18], [19], and application mapping [20], [21]. The majority of these efforts aim to minimize area and power and do not consider the ever increasing problem of power supply noise which is directly affected by the output of these design strategies. Optimizing NoC design for power supply noise requires a tool for modelling this noise to guide and evaluate the optimization process. This paper's aim is to provide such a tool.

Power noise modelling requires models for both workload and power delivery grid. The workload models determine the values and locations of the power consuming modules in the chip, while the power grid model determines the supply voltage profile across the power delivery grid in the presence of these workloads. This section surveys the state-of-the-art in these two areas.

2.1 Power Grid Model

To analyze the power delivery grid in VLSI circuits, the grid is modelled as an RLC network while loads are often modelled as independent current sources [22] or equivalent passive elements [23]. Determining the node voltages for an m node power grid model requires solving the following system of partial differential equations (PDEs).

$$Gv(t) + Cv'(t) = i(t) \tag{1}$$

where $G, C \in \mathbb{R}^{m\times m}$ are matrices representing memoryless elements (resistors) and memory elements (inductors and capacitors) respectively, while, $v(t), i(t) \in \mathbb{R}^m$ are vectors of voltages and imposed independent current sources at the grid nodes respectively. In this model, the independent current sources are used to represent the circuit activity across the chip.

Due to the enormous number of elements and nodes in the grid, solving for the node voltages $v(t)$ using traditional circuit simulators, such as SPICE, is impractical in terms of both memory and simulation time. This problem has been considered by many researchers. Several solutions have been proposed to solve the power grid size problem during both simulation and modelling.

For modelling, model order reduction (MOR) approaches were used to reduce model order before simulation. Multigrid-like [24], [25], hierarchical [26], [27], partition-based [28] and Krylov subspace-based methods [29] are examples of MOR. Other works focused on reducing simulation time, for instance random walk-based simulation [30]. Most of these methods are based on iterative computation, and they are difficult to use for analysis that involves solving the model several times due to their high computational demand. Alternative direct methods are preferable in this case, particularly when the interest is in the peak noise rather than the time profile of the noise.

In [31] a fast and direct model to determine the peak power supply noise is proposed. The power grid is modelled as a distributed RLC network excited by constant voltage sources and switching capacitors (C^{Load}) are used to model the on-chip circuit activity. The amount of these capacitors, for a particular circuit, is determined by the amount of charge (or energy) delivered to this circuit during the switching time period t_s. Approximating the noise impulse with a linear ramp which reaches its maximum at $t = t_s$, the minimum voltage at node j in the grid, V_j^{min}, is given by

$$V_j^{min} = \frac{1}{\lambda_j}\left(\sum_{i=1,i\neq j}^{k} x_{i,j}V_i^{min} + \frac{1}{2}\sum_{i=1,i\neq j}^{k} C_{i,j}V_{DD}\right) \tag{2}$$

where $\lambda_j = \sum_{i=1,i\neq j}^{k} x_{i,j} + \frac{1}{2}\sum_{i=1,i\neq j}^{k} C_{i,j} + C_j^{Load}$ and $x_{i,j} = t_s^2/(6L_{i,j} + 3R_{i,j}t_s)$. $R_{i,j}$, $L_{i,j}$ and $C_{i,j}$ are the resistance, inductance and capacitance between nodes i and j in the power grid, respectively, t_s is the switching time, and C_j^{Load} is the equivalent capacitance of the load at node j.

This model provides an accurate estimation of peak power supply noise and a maximum error of 5% was reported [31]. This power grid model can significantly reduce the simulation time. It would be an ideal candidate for developing a tool for real-time supply variations evaluation. However, this model assumes known load equivalent capacitances. In this work, these capacitances dynamically change in real time, thus, this power grid model can be adopted after introducing a technique to determine these switching load capacitances for the communication fabric of NoC. This load includes both links and router switching as detailed in Section 3.2.

2.2 Workload Model

Considerable amount of literature have been published on workload modelling for power grid analysis. Techniques that represent the workload with equivalent passive elements were reported. For example, in [23], a macro model based on the effective impedance of the current consumer is proposed. Other techniques that are based on independent current source models are also used. For example macro-models can be used to determine the waveforms of these current sources. In [22], a frequency domain current macro-model is proposed where the input vector pairs of the circuits are partitioned according to the hamming distance and a current macro-model is built for each distance using regression. However, these workload models assume computation workloads are independent current sources or passive elements. Task dependencies and correlation between the computational cores are ignored and hence workload due to communication cannot be captured in these models.

Independent current sources or passive elements can be a reasonable representation of workloads for simple logic or, to an extent, for microprocessors. For complex on-chip communication systems, such as networks-on-chip, dynamic power consumption through the on-chip communication infrastructure imposes a significant variation in power and load. Therefore, an effective model that captures the communication dynamics is required. This model must integrate both the router circuit and links workloads.

For the router circuit workload, a router microarchitectural power model [17] can be integrated with a circuit activity simulator to characterize this workload.

A model for on-chip link is essential for determining on-chip communication loads. In literature, a number of current [32], energy [33], and power [34], [35] based models for on-chip interconnects have been proposed. On-chip interconnects are modelled as capacitively and inductively coupled distributed RLC lines. In [32], an analytical model for on-chip link current based on decoupling techniques is presented. On-chip link wires are driven by exponential voltage source (V_S) and loaded by capacitor (C_L). A two-port network with source and load impedances is employed to derive a closed form for wire current. The link current can be obtained using decoupling transformation [34].

3 METHODOLOGY

On-chip communication traffic produces a considerable portion of the overall power supply noise in NoCs. In contrast to conventional models for logic or microprocessors, this portion of engendered noise or V_{DD} variation would have an interesting correlation to the spatial distribution of the traffic load which is a direct result of the network-level design outcomes, such as application mapping and routing path allocation (or routing algorithm).

Power supply noise model requires a detailed consideration of both workload and power grid models. Fig. 1 illustrates the inputs, outputs and the models used for power supply noise computation in NoCs in this work. The technology and architectural parameter files, in addition to the floorplan information are taken as input to the model. These files are given to the NoC power model to compute the power traces of NoC components. For router, we used the well known NoC router power and area model Orion [17]. This model is fast, accurate and easy to integrate with other models. More importantly, this is an architecture-level model which enables the macro-modelling of power grid workload for NoCs. Application characteristics files are fed to an NoC simulator to generate traffic information. The open-source SystemC-based NoC simulator, Noxim [36], is modified and employed here because it supports a wide range of traffic distributions and routing algorithms and due to its efficiency and ease of configuration. Also, for power grid model solution, the fast peak noise power grid model proposed in [31] is adopted.

Integrating all these models in an automated flow enables the computation of dynamic voltage variations in NoCs. The tools and the benchmarks used in this work are made available online [37]. The methods and tools presented in this paper are applicable to a wide range of topologies, including tree, mesh, torus etc. However, for convenience we will describe it in the context of a regular mesh topology.

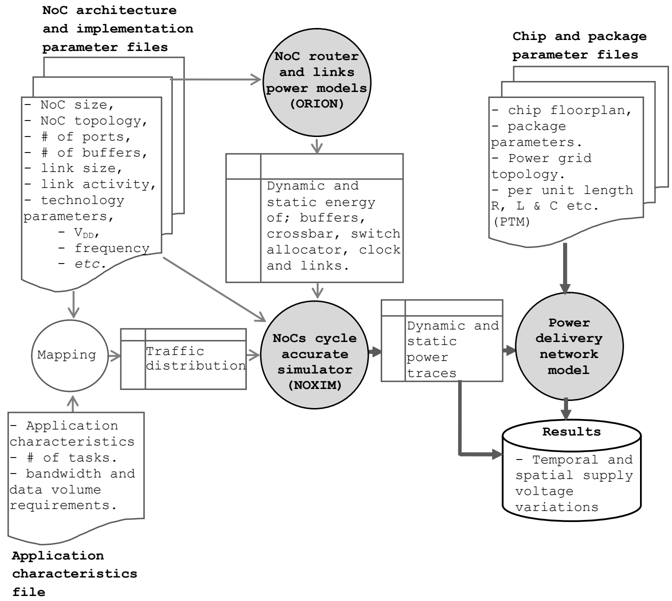


Fig. 1. Computational flow for NoC power supply noise modelling.

implies that segment length and width decreases (grid granularity increases) as we go from more global to more local power grid nodes. Some nodes in the upper layers are connected to package V_{DD} (and ground) pads. These pads are modelled as RL segments here.

As mentioned earlier, power grid analysis for the whole grid is not practical due to the huge size of the resulting model. Thus, exploiting the hierarchical nature of power grids, power grid analysis can be performed using a macro modelling approach. A lumped model is used to characterize individual blocks in the chip and power grid of appropriate granularity is considered [26]. In this work a macro model is considered for routers. The

3.1 Power Noise in Networks-on-Chip

Many design parameters can affect the spatial and temporal distribution of communication loads in NoCs. Routing algorithms, traffic patterns, and packet injection rates are examples of these parameters. Consequently, NoC communication workloads have spatial and temporal distributions both are determined by the design entities of the system [38], [39], [40]. This distribution in time and space of communication loads is reflected as spatial and temporal power supply noise distribution in the power delivery grid.

Fig. 2 shows a general overview of a network-on-chip and its power grid. The power grid is a grid of metal wires, which can be modelled as an RLC network. The power grid can have different topologies, e.g. mesh, tree, and irregular. These grids usually span several metal layers and they are hierarchical in nature. This

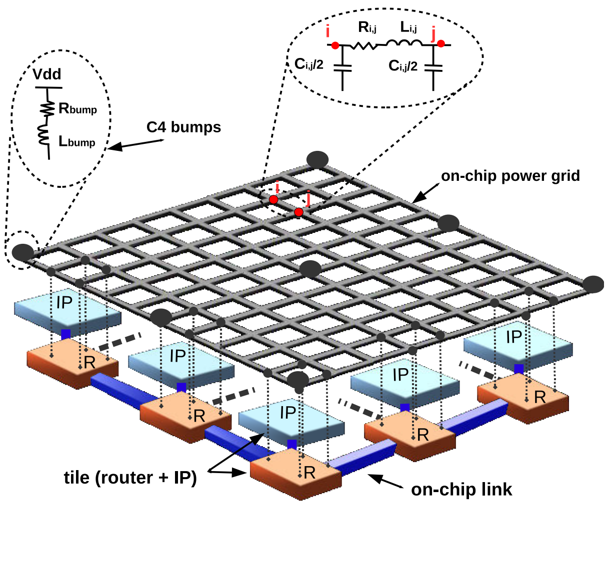


Fig. 2. NoC power delivery network.

$C_r^{total}(k)$	Total load equivalent capacitance for router r.
$C_r^{links}(k)$	Load equivalent capacitance for router r at cycle k due to link traversals for all the links in the router.
$C_{ch}(k)$	Load equivalent capacitance at cycle k due to link traversal of router channel ch.
$C_r^{circuit}(k)$	Load equivalent capacitance of router r at cycle k due to circuit activity.
n	Number of wires in the NoC data link.
R	The set of all routers in the NoC.
CH_r	The set of all channels in router r.
G_r	The set of power grid nodes responsible for delivering power to router r.
SW	Vector of size n with elements representing the wire switching direction, 0 for quiet, 1 for switching up and -1 for switching down.
Ψ	The set of microarchitectural-level processes executed by NoC router.
$E_r(k)$	Total energy delivered to router r at cycle .
E_ψ	Energy consumed when executing micro process ψ.
$\alpha_\psi(r,k)$	The number of occurrences of process ψ in router r at cycle k.
I_ch	The link current profile for channel ch.
g	Power grid granularity multiple used to map fine grid to a coarse grid.
l_f	The fine power grid segment length.
l_X	The coarse power grid segment length.
$Error_g$	Power supply noise error due to grid granularity reduction.
V_X	Voltage for the coarse grained grid model.
V_f	Voltage for the fine grained grid model.
G_f	Fine grained (original) power grid.
G_X	Coarse grained (reduced) power grid.
$t^i_{clk\_Q}$	Clock-to-Q delay of latch i.
t^i_{setup}	Critical setup time of latch i.
$t^{i,j}_{wire}$	The delay of wire i,j.
$Pr(Err_i)$	Probability of error for link i.
γ_i	Utilization of link i

TABLE 1
Definitions of symbols.

router is matched with a region of the power grid which is determined by the floorplan information.

Fig. 3 illustrates links and routers of network-on-chip. Router functional units and link's equivalent circuit including the drivers and loads are also shown. In this work, the link wires are modelled as RLC interconnects driven by exponential voltage sources and loaded by capacitances (C_L). NoC workload consists of the switching workloads of both router circuits and links. Router workload is due to internal router processes. These processes are: receiving a flit, route computation, switch allocation and switch traversal. Link workload is due to the switching of its drivers and repeaters.

3.2 Compartmental Modelling for Communication Fabrics

The data in NoCs are routed using routers. Routers include a cross-bar switch (see Fig. 3) which, for mesh topology, comprises four input/output channels for global communication (north, east, south and west) and one input/output channel for local communication. The link is driven by drivers, which are part of the sending router circuitry, this implies that link traversal power is supplied by the flit forwarding router and the repeaters along the link path (if any). In our model, the workload of both routers and links are characterized by capacitance. This capacitance is determined by the charge delivered to the circuit during the switching time period.

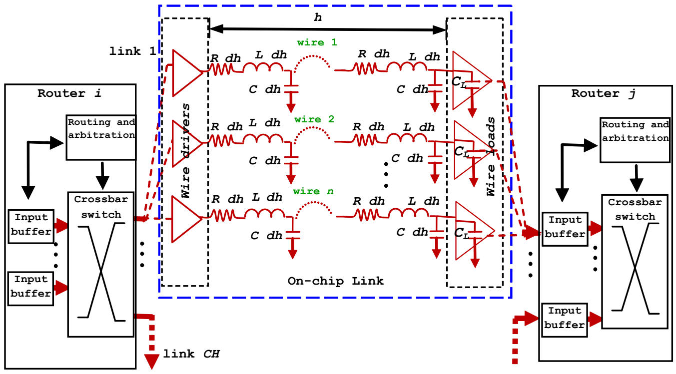


Fig. 3. Illustration of the NoC routers connecting tiles i and j, and the equivalent circuit of on-chip links from router i (sender) to j (receiving). n: link size, h: link length, CH: number of router channel.

Based on the above, we compute the router capacitive load for router $r \in R$ (see Table 1) in the power grid at the k^{th} switching cycle ($C_r^{Load}(k)$) as:

$$C_r^{load}(k) = C_r^{links}(k) + C_r^{circuit}(k). \quad (3)$$

where $C_r^{links}(k)$ is load equivalent capacitance due to link traversal of the flits at cycle k, and $C_r^{circuit}(k)$ is the load equivalent capacitance due to router circuit activity at cycle k. The capacitances here vary over time to reflect dynamic communication load changes in the network.

3.2.1 Link Workload

Links traversal load is the summation of loads of all cahnnel links of the router i.e.

$$C_r^{links}(k) = \sum_{\forall ch \in CH_r} C_{ch}(k) \quad (4)$$

where, for mesh topology, the set of router channels are $CH_r = \{North, East, South, West, Local\}$.

Channel's link load capacitance at cycle k, $C_{ch}(k)$, can be computed from link's current profile at that cycle as follows:

$$C_{ch}(k) = \frac{Q_{ch}(k)}{V_{DD}} = \frac{\int_0^{T_{clk}} I_{ch}(k,t)dt}{V_{DD}} \quad (5)$$

where $Q_{ch}(k)$ is the total charge delivered to the channel link at cycle k, $I_{ch}(k,t)$ is the current profile of the channel link at cycle k, and T_{clk} is the clock frequency period. In this work the link model proposed in [32] is employed to compute this current profile after modifying the formula for total channel link current I_{ch} to include wire switching.

Let SW be a vector of size n (the number of link wires) with elements sw_i that takes values of 0, 1 or -1, when wire i is quiet, switching up $(0 \rightarrow 1)$, or switching down $(1 \rightarrow 0)$, respectively. The time profile of the current draw of channel link at cycle k, $I_{ch}(k,t)$, can be expressed as:

$$I_{ch}(k,t) = \sum_{h=1}^{n}\sum_{i=1}^{n} M_{h,i}^{T} sw_{i,ch,k} \sum_{j=1}^{n} M_{i,j}^{T} I_{i,ch,k}(t) \qquad (6)$$

where $I_{i,ch,k}$ is the current of wire $i \in \{1...n\}$, for the channel link at cycle k, M is the decoupling transformation matrix [34] and $sw_{i,ch,k}$ is the switching direction of wire i at cycle k [32].

3.2.2 Router Workload

Router's circuit capacitive load for a switching time period can be computed from the energy consumed by router's circuit in this time period. This energy is determined by integrating an NoC simulator, to determine the processes are taking place in the router at each cycle, and a router power model, which determines the energy consumed by each of these processes.

Let Ψ={RECEIVE, ROUTE, FORWARD, STANDBY} be the set of microarchitectural-level processes that can be executed by the router. These processes are receiving a flit, route computation, forwarding a flit, and no activity (static energy), respectively. The energy of the RECEIVE process is the energy required for writing to the input buffer. ROUTE process energy is required for route computation, which is only performed for header flits (for wormhole routing). FORWARD process energy is the summation of energies required for reading from input buffer , switch allocation, switch traversal and link traversal. Also, let $\alpha(r,k)$ be the number of occurrences of process $\psi \in \Psi$ in router r at cycle k. Now, the total energy delivered to router r at cycle k, $E_r(k)$, can be expressed as:

$$E_r(k) = \sum_{\forall \psi \in \Psi} E_{\psi}.\alpha_{\psi}(r,k). \qquad (7)$$

where E_{ψ} is the energy required by the router circuit to execute process ψ. This energy, for a particular router design, can be computed using a router microarchitectural power model, while, α_{ψ} can be determined using a cycle accurate NoC simulator. The router load equivalent capacitance at cycle k ($C_r^{circuit}(k)$) can now be computed as follows:

$$C_r^{circuit}(k) = \frac{E_r(k)}{V_{DD}^2} \qquad (8)$$

The load capacitance which results from Eq. 3 is used to characterize the load in Eq. 2.

The following sections are based on the assumption that router r is supplied with power through a set of nodes (G_r) in the power grid, and in line with [41], [42], the resulting router load is divided equally over the set G_r. The set G_r is determined by the floorplan information, i.e. based on the geometrical structure of the grid and areas and positions of the routers.

3.3 Power Grid Granularity

In power grid simulation there are two techniques for model solution, iterative and direct [27]. Due to the very large grid size, the iterative technique is more suitable for analyses which involve obtaining single system solution. For instance, the DC analysis of power grids. On the other hand, direct technique is more convenient when multiple model solutions are necessary. This is the case for this work, thus, we used a coarse-grained lumped model for the power grid in order to achieve significant results in a practical simulation time. Power grids designed for real VLSI circuits may contain tens of thousands or even millions of nodes [43] which results in impractical simulation time and memory requirements. Thus, a coarse grid approach was used in many previous works [24], [25], [44]. In these works a multi-grid based model order reduction is used and the number of nodes in the power grid is reduced by node elimination. The analysis is performed on the reduced coarse grid. Then, the solution is mapped back to the original (fine-grained) grid using linear interpolation, taking into account the values of conductances between the nodes.

During the mapping from the fine grid (G_f) to the coarse grid (G_X) the geometrical coordinates of the nodes in G_X must be equal to their counterparts in G_f to preserve the structure of the grid. For regular mesh topology, this mapping takes the ratio ($g = \frac{l_X}{l_f}$) of the segment length of the fine grid, l_f, and the segment length of the coarse grid, l_X, as input (see Fig. 4). To keep the total resistance equal, a segment's width must be increased in proportion to increasing its length.

The relative error of the voltage drop for node $j' \in G_X$ due to grid granularity reduction can be expressed as

$$Error_g(j') = \left| \frac{\Delta V_f(j') - \Delta V_X(j)}{\Delta V_f(j)} \right| \times 100\%, \qquad (9)$$

where ΔV_f and ΔV_X are the voltage drops in the fine and coarse grids, respectively, j and j' are the nodes in the fine grid and its counterpart in the coarse grid, respectively.

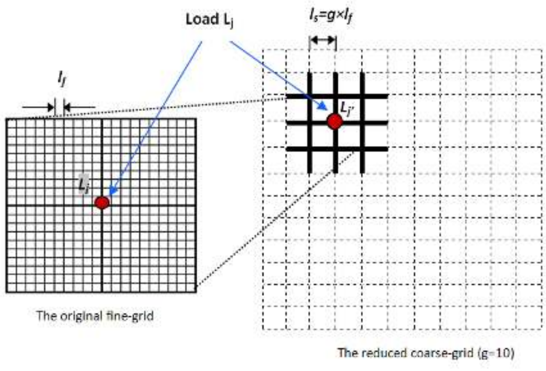


Fig. 4. Illustration of the mapping from fine-grained to a coarse-grained models of the power delivery grid.

4 Experimental Results and Discussion

4.1 Experimental Setup

To evaluate our model, we adopt the floorplan and architecture of Intel's TeraFlop tile [11]. A 38-bit communication links are assumed with 3 GHz frequency. The technology used is 65 nm with nominal V_{DD}=1V. Tile dimensions are 2mm height and 1.5 mm width. The computational units' power traces are estimated using results presented in [11].

For the power delivery network (PDN), we used a lumped model which includes both on-chip and off-chip power delivery network models. The on-chip PDN consists of a global level mesh structure routed in the top metal layers. Unless otherwise mentioned, the on-chip power network is modelled as RLC mesh with a grid segment length such that we have 5×5 granularity per NoC tile. Based on our analysis (see Section 4.3) and as suggested by [41], [45], this is enough for capturing the power supply voltage variations across the chip with reasonable accuracy and simulation time. The RLC values of the grid segments and link wires were determined using PTM [46].

Orion [17] is used for router power computation. Communication traffic simulation is done using Noxim [36], a SystemC-based NoC simulator. A Gaussian random distribution is assumed for the link switching activities and a Poisson distribution for the packet injection. A uniform buffer size of 16 flits and a packet length of 3 flits are assumed. These values are in line with Intel's TeraFlop NoC configuration [11]. These models are integrated in an automated flow to compute the power supply voltage variations as a function of activity for on-chip networks (Fig. 1).

4.2 Model Verification

Firstly, we performed an experiment to evaluate the accuracy of our model. The power trace of a 3×3 NoC is computed under the $Transpose$ traffic (in which $tile(i,j)$ sends packets to $tile(j,i)$) that results from the NoC simulator for a packet injection rate of $0.015\ packets/cycle/node$. Then, a SPICE netlist of the circuit is generated. In this netlist the workload is modelled as triangular current sources. The peaks of these current sources are computed based on power traces taken from the router and link power models and the activity of the NoC. This will enable the evaluation of the voltage variations resulting from integrating all the components of the model together (activity, power, and grid models).

This circuit netlist is simulated in SPICE to obtain the resulting grid node voltages. The same scenario is also simulated using our model following the methodology described in Section 3. Power grid node voltages obtained from both the model and SPICE are presented in Fig. 5, which shows good matching with a mean relative error of only 4.7%.

4.3 Granularity Analysis

The impact of power grid granularity on both model accuracy and simulation time is also evaluated. A power grid for an area of $1mm^2$ is simulated assuming a load of 38-wire link computed using Eq. 6. V_{DD} drop across the grid is determined with different power grid granularities. Coarser grids are generated by doubling the grid segment lengths and widths to preserve the resultant resistance.

Table 2 shows the results of this analysis for a grid granularity starting from 40×40 (6400 nodes) down to 5×5 (25 nodes). Taking SPICE simulation of the 40×40 grid as a baseline, both relative error (Eq. 9) and simulation time are shown.

Results show that model accuracy decreases when the granularity of the model is quartered. On the other hand, simulation speed rapidly increases with granularity due to higher model order reduction. Considering fixed grid granularity per tile, the simulation time increases linearly with the NoC size (number of tiles) due to the fact that the number of power grid nodes will increase linearly with the number of tiles.

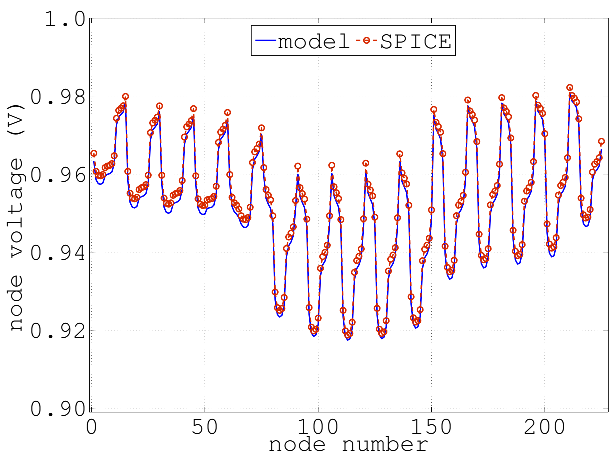


Fig. 5. Comparison of node voltages between the proposed computational model with SPICE simulation for a 3×3 NoC configuration.

l_X (μm)	l_f (μm)	#of nodes	$g = l_X/l_f$	$Error_g$ (%)	time (s)	using
25	25	40 × 40 (1600)	1	-	8.9	SPICE
25	25	40 × 40 (1600)	1	1.98	1.152	model
50	25	20 × 20 (400)	2	6.07	0.155	model
100	25	10 × 10 (100)	4	8.6	0.022	model
200	25	5 × 5 (25)	8	11.85	0.006	model

TABLE 2
Comparison between the proposed model with different power grid granularities with SPICE simulation.

4.4 Synthetic Traffics and Routing Algorithms

In this section we present and discuss the results of power supply noise caused by different synthetic traffic patterns and routing algorithms. A 6×6 NoC is considered here. Traffics used are *Random*, *Transpose*, and *Hotspot*. For the Random traffic each tile sends data to all other tiles with equal probability. For the Transpose case $tile(i, j)$ sends packets to $tile(j, i)$. For the Hotspot traffic pattern the four central tiles receive an extra 5% in addition to the uniform (Random) traffic. We also considered four routing algorithms; *XY*, *Odd-Even* (OE), *Fully-Adaptive* and *Negative-First* (NF).

The number of clock cycles necessary for capturing the characteristics of the workload differs from one application to another. For synthetic traffic it is noticed that the power traces have been constantly repeated after 10,000 clock cycles. However, the model is run for 100,000 clock cycles to guarantee the coverage of all workload characteristics. The resulting V_{DD} drop is plotted for different routing algorithms (Fig. 6) and traffic distributions (Fig. 7) for a range of packet injection rates. The peak and mean drops are shown for both figures. The achieved throughputs for these routing algorithms and traffics are also shown in Fig. 8. Spatial V_{DD} drops under these traffics and routing algorithms are given in Figures 9 and 10 respectively for a PIR of 0.015 $packets/cycle/node$.

Note that in general there is a considerable increase in V_{DD} drop with PIR. This is expected since a higher packet injection rate leads to a higher throughput which increases the switching activity of the routers and data links and, in turn, raises the current draw causing a higher V_{DD} drop. However, different routing algorithms and traffics behave differently in terms of V_{DD} drop with the increase of PIR.

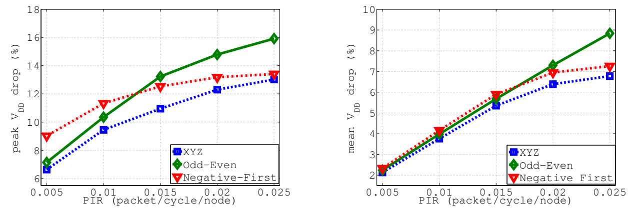
For instance, consider the NF routing algorithm. For low PIR (< 0.015) it can be noticed that it causes higher peak (Fig. 6(a)) and mean (Fig. 6(b)) drops than the XY and OE, although it achieves the same throughput (and thus consumes the same power) within this PIR range (see Fig.8(a)). This is due to the fact that NF algorithm tends to migrate the traffic to the negative quarter of the NoC mesh as can be seen in Fig. 9(d). This can create hotspots that would suffer higher supply drop due to unbalanced power density. On the other hand, both XY and OE have more harmonic spatial workloads compared to NF which leads to lower supply drop.

At high PIR (> 0.015) the V_{DD} drop for both NF and XY is less than that of OE due to the fact that OE achieves higher throughput compared to XY and NF at this PIR range. This is becuase the NoC starts to saturate and throughput decreases for the latter two which is not the case for OE (see Fig. 8(a)).

Looking at the traffic patterns, we can see that the Hotspot traffic causes higher V_{DD} peak drop due to the centric nature of this traffic distribution and for the same reasons discussed above. This drop reduces at higher PIR (> 0.015) due to the reduced throughput caused by saturation. Random traffic results in higher throughput at this range of PIR (Fig. 8(b)) which causes higher peak and mean drops.

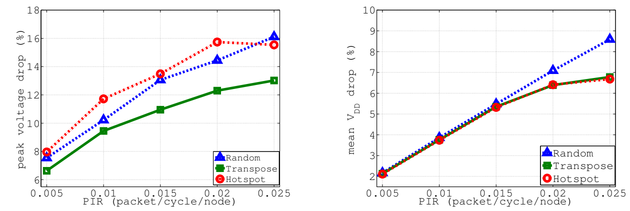
Figures 9 and 10 show the spatial distributions of selected routing algorithms and traffics, respectively. In general, it can be observed that the power supply drop which results from a traffic/routing is determined by the amount of the workload of this traffic/routing and increases with this workload. Also, the spatial distribution of a traffic/routing workload plays an important role here. Highly unbalanced traffic/routing can lead to significantly higher power supply noise compared to the balanced traffic/routing, even for the same amount of workload.

Table 3 summarizes the results of peak and mean V_{DD} drops for the considered set of traffics and routing algorithms.



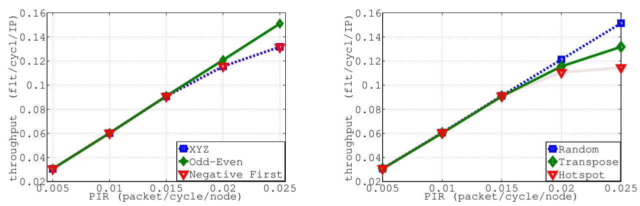
(a) Peak V_{DD} drop (b) Mean V_{DD} drop

Fig. 6. V_{DD} drop for different routing algorithms versus PIR.



(a) Peak V_{DD} drop (b) Mean V_{DD} drop

Fig. 7. V_{DD} drop for traffic patterns versus PIR.

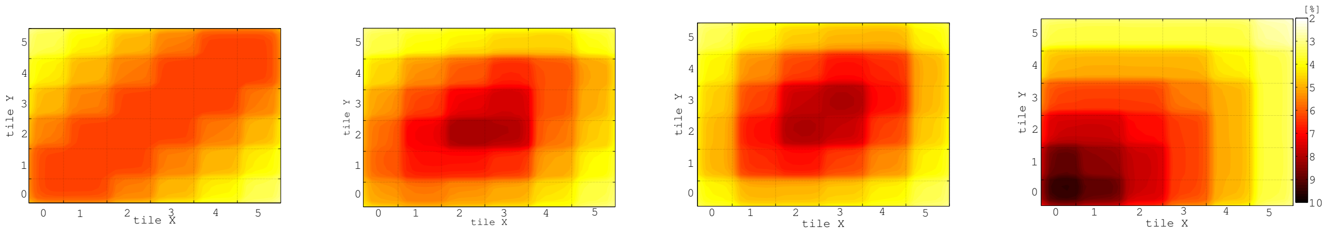


(a) (b)

Fig. 8. Throughput for differnet (a) routing algorithms, and (b) traffic patterns.

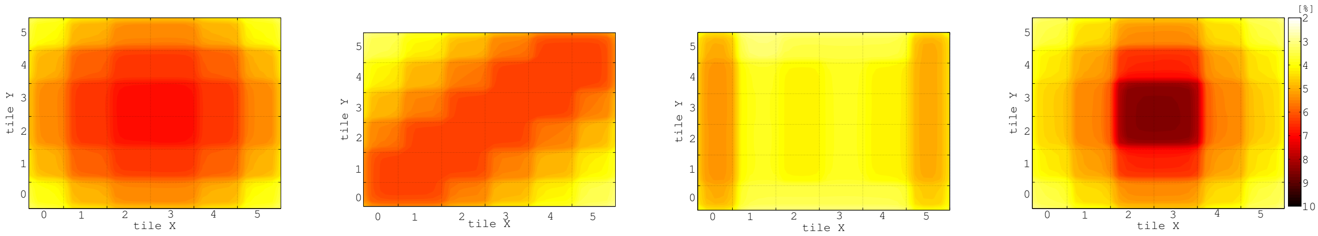
4.5 Real Traffic

To generate a realistic communication scenario, a generic complex MultiMedia System (MMS) which comprises



(a) XY (b) Odd-Even (c) Fully-Adaptive (d) Negative-First

Fig. 9. Spatial distribution of mean V_{DD} drop (%) for different routing algorithms and Transpose traffic.



(a) Random (b) Transpose (c) Butterfly (d) Hotspot (5%)

Fig. 10. Spatial distribution of mean V_{DD} drop (%) for different synthetic traffics with XY routing.

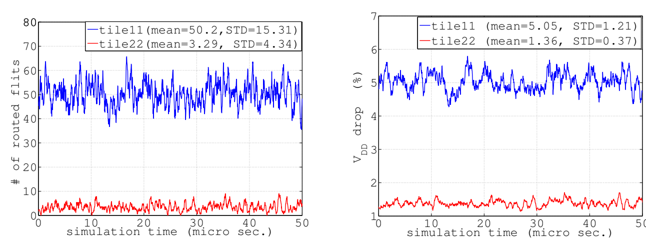
	TRAFFIC					
	Random		Tranpose		Hotspot	
ROUTING	peak	mean	peak	mean	peak	mean
XY	13.63	5.6	12.95	5.6	13.18	5.6
Odd-Even	11.51	5.48	13.82	5.5	12.53	5.45
Negative-First	14.15	5.31	13.81	5.31	14.21	5.3
Fully-Adaptive	12.96	2.92	13.56	5.48	12.79	5.32

TABLE 3
Summary of V_{DD} drop (%). Results of four different routing algorithms and three traffic patterns.

H263 video encoder, an H263 video decoder, an MP3 audio encoder, and an MP3 audio decoder is used [20]. We considered three mapping strategies to map this benchmark to a 5×5 NoC; maximizing performance (minimizing packet latency) [21], minimizing energy [47] and a random mapping. The resulting V_{DD} drops in the presence of the resulting three traffic patterns are computed using our tool. The power trace for this benchmark is found to be periodic with a period of nearly 70,000 clock cycles. The simulations are run for 100,000 cycles to guarantee the coverage of workload characteristics.

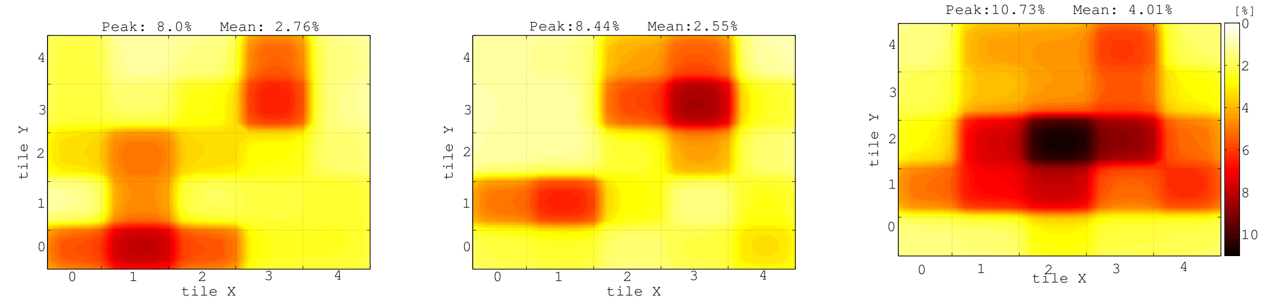
Fig. 11 shows the spatial distribution of power supply drop. It can be seen that performance and energy mappings are relatively close to each other in terms of V_{DD} drop. However, the energy-aware mapping has slightly higher peak drop compared to performance-aware mapping. It can also be seen that in this instance of random mapping there is considerably higher drop compared to performance and energy mappings. This is caused by not only higher power for this mapping ($14W$ compared to $10.4W$ for the performance mapping and $9.8W$ for the energy mapping), but also due to the spatial distribution of this power profile which results in higher power (and thus current) density in the central tiles and leads to higher voltage drop.

Fig. 12 plots the traffic (in terms of routed flits) and the corresponding V_{DD} drop with time for two tiles from the MMS benchmark with performance mapping. This figure illustrates very high correlation between traffic and supply drop. This can be explained by the fact that power dissipation in NoCs is strongly correlated with the network traffic load. Thus higher traffic is reflected as higher current draw increasing supply drop on the power grid. This also implies that higher/lower temporal variation of traffic (Fig. 12(a)) results in higher/lower variation of V_{DD} increasing power supply noise (Fig. 12(b)). It can be noticed that the DC component of V_{DD} drop is highly correlated to the average of the traffic while the AC component is correlated to the temporal variation of traffic.



(a) Traffic (b) V_{DD} drop

Fig. 12. Temporal variations of routed traffic and the corresponding V_{DD} drop for buzy (tile 11) and quite (tile 22) tiles, under the MMS traffic.



(a) Maximum performance mapping (b) Minimum energy mapping (c) Random mapping

Fig. 11. Spatial distribution of mean V_{DD} drop (%) for the MMS application traffic with three mapping strategies.

5 Case Study: Timing and Error Analysis of Links in the Presence of Power Supply Variations

Power supply noise modelling for NoCs can find many applications in the evaluation and design space exploration at various levels. Power grid integrity analysis, power supply noise-aware application mapping, and floor planning, are examples of these applications. However, our tool can also be used to analyze the impact of the resulting V_{DD} variations on timing accuracy for the circuit dominated paths [1], or link dominated paths, such as communication links and clock distribution networks [48], [49]. Here we perform a power supply variations-aware statistical timing analysis of NoC links which enables the computation of the probability of switching errors for data links comprising the NoC communication fabric.

5.1 Power Supply Variations Impact on Link Delay

A major impact of power supply noise on performance can be seen by its impact on delay. Power supply voltage drops can cause significant increases in delay for circuit dominated as well as interconnect dominated paths [48], [50]. This delay could lead to violations of timing constrains in these paths and thus generate soft errors. It has been reported that for these timing constraints to be met for a 20% of supply variation, a 42% decrease in frequency is required for 65nm technology [48]. To compute the probability of switching error due to timing delays under power supply variations, a full knowledge of the V_{DD} variation distribution is needed. Also, a delay versus V_{DD} relationship for various components of the link is necessary. The delay components of on-chip global interconnects are illustrated in Fig. 13. Considering a synchronous data path between two tiles, i and j, of a NoC data link and assuming a zero clock skew between these two, the sum of the clock-to-Q delay of the sending Flip-Flop ($t^i_{clk\_Q}$), the wire delay between the two FFs ($t^{i,j}_{wire}$) in addition to the setup time of the receiving FF (t^j_{setup}) must not exceed the link clock period T_{clk} [51] i.e.

$$t^i_{clk\_Q} + t^{i,j}_{wire} + t^j_{setup} < T_{clk}. \quad (10)$$

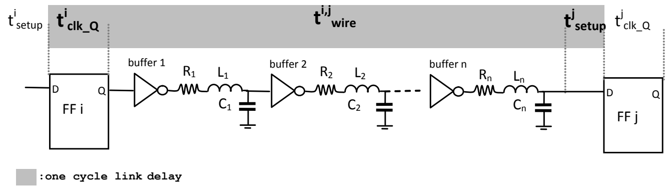


Fig. 13. A model of on-chip link illustrating the delay components and timing constrains [51].

Violation to this timing constraint would lead to switching errors in the link.

To perform timing analysis, and in line with [1], [50], we adopt a quadratic approximation to determine the impact of V_{DD} drop on these delay components, i.e.

$$t_d(\Delta V_{DD}) = k_1 + k_2(\Delta V_{DD}) + k_3(\Delta V_{DD})^2 \quad (11)$$

where, $t_d(\Delta V_{DD})$ is any of the link timing components on the left hand side of Eq. 10 and k_i (i=1,2,3) are technology dependant constants. Assuming 65 nm technology, we simulated an edge triggered D FF, which comprises two master-slave D latches, in SPICE and obtained the clock-to-Q and setup times of the FF under V_{DD} variation. We also obtained the wire delay for a $2mm$ length buffered wire. Fig. 14 plots the delay obtained for t_{clk_Q}, t_{wire} and t_{setup} when V_{DD} drop is varied from 0% ($V_{DD} = 1.0V$) to 25% ($V_{DD} = 0.75V$) of nominal supply voltage with a step of 5% ($50mV$). Using these results, analytical formulas relating the clock-to-Q delay, setup time and wire delays to V_{DD} drop are obtained using regression.

5.2 Probability of Timing Violation Errors and Bit Error Rates

Using our model, we obtained the distribution of V_{DD} for all chip components. In general the obtained V_{DD} distribution consists of a DC component (or IR) drop, which results in shifting the mean of the the V_{DD} distribution and an AC component(or ΔI droop), which results in variation of V_{DD}. Using the formulas for link delay (Eq. 11), a statistical timing analysis can be performed to obtain the distribution of the resulting link

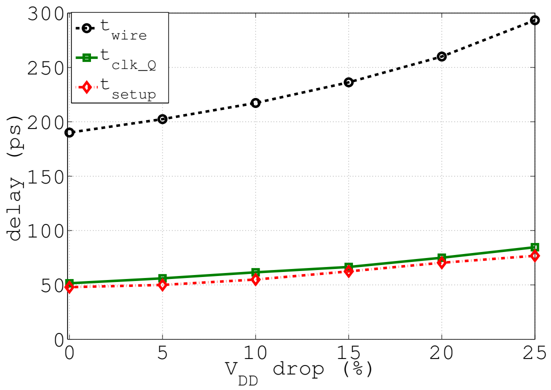


Fig. 14. The t_{wire}, t_{clk_Q} and t_{setup} link delays versus V_{DD} drop.

delay variations due to power supply variations for all NoC links. Thus, IR drop translates into a delay skew and ΔI droop translates into delay jitter for these links.

For FFs, (t_{clk_Q}) and t_{setup} are computed from the V_{DD} of the sending and receiving tiles respectively. The wire delay (t_{wire}) is computed using the V_{DD} between the sending and receiving tiles.

The resulting delay distribution of a link can be used to estimate the probability of timing error due to power supply variations for that link. We estimate this probability as the portion of the delay distribution that does not satisfy the constraint in Eq. 10. In other words the probability of timing error on link l, $Pr(Err_l)$, can be expressed as:

$$Pr(Err_l) = Pr(t_l > T_{clk}) \tag{12}$$

where t_l is the total link delay which is computed in the presence of V_{DD} variations for that link using Eq. 11. Fig. 15 shows the results of this analysis for the MMS benchmark with maximum performance mapping. Fig. 15(a) shows the distribution of delay means (skews) for all links and Fig. 15(b) shows the distribution of delay STDs (jitters) for these links. It is found that links with the highest error probability belong to the tile, which suffers the highest V_{DD} drop, tile 2, as can be seen in Fig. 11(a).

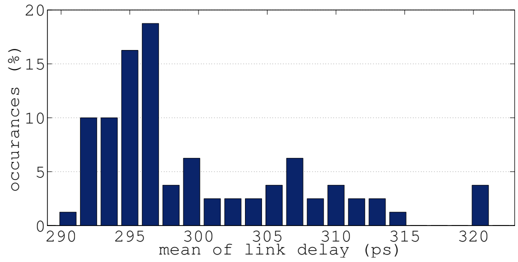
Using the probability of error for each link the average bit error rate (BER) for the NoC can be computed. Given the application and mapping characteristics, let the relative utilization of a NoC link l be γ_l which is the ratio of the data volume communicated through this link (ν_l) to the total data volume communicated in the NoC. Thus, the relative utilization of the link can be characterized as:

$$\gamma_l = \frac{\nu_l}{\sum_{\forall i,j \in \{N\}} \nu_{i,j}} \tag{13}$$

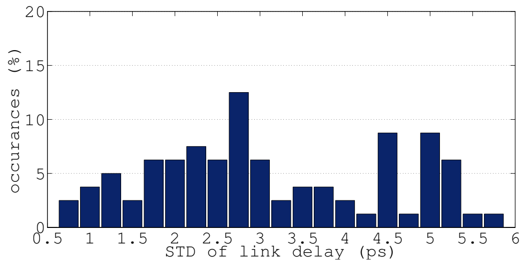
where $\{N\}$ is the set of all nodes (tiles) in the NoC and $\nu_{i,j}$ is the data volume that needs to be communicated between tile i (as a source) and tile j (as a consumer). Now, the the bit error rate (BER) for all the NoC links (assuming links to be independent) can be computed as follows:

$$BER = \sum_{\forall l \in \{L\}} \gamma_l.\alpha_l.Pr(Err_l) \tag{14}$$

where $\{L\}$ is the set of all links in the NoC and α_l is the average switching activity of link l. The switching activity α_l is the average spatial activity of the link which is determined by the average hamming distance between consecutive flits, while γ_l can be seen as the average switching activity in time. Both are in the range of 0-1.



(a) Distribution of links delay (mean)



(b) Distribution of links delay (std)

Fig. 15. Links delay statistics for the MMS benchmark using maximum performance mapping.

For the MMS benchmark with performance mapping and assuming α_l is 25% the average BER is found to be 5.9476×10^{-6}. On the other hand, when energy mapping is considered the NoC expereinces slightly higher drop (see Fig. 11) and BER increases to 2.9×10^{-5}.

To illustrate the impact of increased communication workloads in terms of network throughput on BER for various traffic scenarios, Fig. 16 plots BER for different synthetic traffics with throughput. It can be noticed that BER increases exponentially with throughput for all traffics. However, due to different hotspot and traffic distributions, different traffics experience different BERs. Higher and more concentrated traffics lead to higher BER as can be seen for the Hotspot traffic in Fig. 16.

6 Conclusion

In this work, an integrated tool to capture the impact of on-chip communication workloads on power delivery

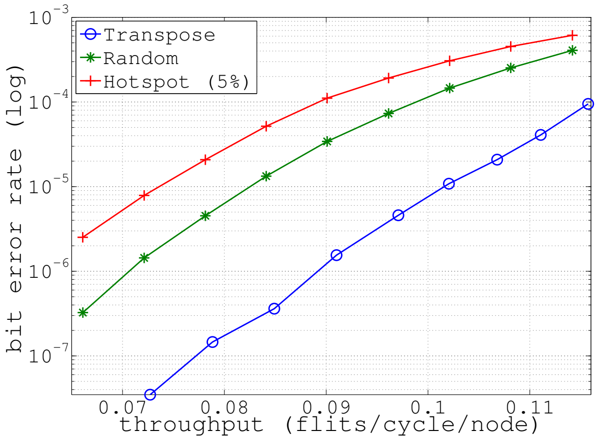


Fig. 16. Bit error rate versus throughput for various synthetic traffics.

grid is presented. This tool is dedicated for NoCs. It integrates a NoC simulator, on-chip link model, NoC power and area models, and a fast power grid model to provide a comprehensive simulation and system analysis. The granularity of the power grid model would contribute to the degree of accuracy of the analysis. Compared to the SPICE simulation, error of the power grid model is less than 2% and increases linearly with the granularity of the grid. The developed tool also provides detailed analysis for power supply variation based on the traffic distributions and routing algorithms. The practicality of the proposed model is further exemplified through a case study. Detailed statistical timing analysis for communication links delay is presented. This enables the study of impact of power supply noise based on communication delay for different communication workloads and traffic patterns. The results from such analyses can be used to determine a high-level performance metric, such as probability of error and bit error rates. Comprehensive analyses of power grid and accurate communication link models are crucial to power supply integrity evaluation. This proposed method would lead to a robust evaluation of NoC-based multi-core systems in early design stages.

References

[1] M. Saint-Laurent and M. Swaminathan, "Impact of power-supply noise on timing in high-frequency microprocessors," *IEEE Transactions on Advanced Packaging*, vol. 27, no. 1, pp. 135–144, 2004.

[2] S. R. Nassif, "Power grid analysis benchmarks," in *Design Automation Conference. ASPDAC'08. Asia and South Pacific*, 2008, pp. 376–381.

[3] A. H. Ajami, K. Banerjee, and M. Pedram, "Scaling analysis of on-chip power grid voltage variations in nanometer scale ulsi," *Analog Integrated Circuits and Signal Processing*, vol. 42, no. 3, pp. 277–290, 2005.

[4] H. H. Chen and D. D. Ling, "Power supply noise analysis methodology for deep-submicron vlsi chip design," in *Design Automation Conference. Proceedings of the 34th*, 1997, pp. 638–643.

[5] H. H. Chen and J. S. Neely, "Interconnect and circuit modeling techniques for full-chip power supply noise analysis," *Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging, IEEE Transactions on*, vol. 21, no. 3, pp. 209–215, 1998.

[6] Y. Wang, J. Xu, Y. Xu, W. Liu, and H. Yang, "Power gating aware task scheduling in mpsoc," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, no. 99, pp. 1–12, 2010.

[7] J. Hwisung and M. Pedram, "Optimizing the power delivery network in dynamically voltage scaled systems with uncertain power mode transition times," in *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, 2010, pp. 351–356.

[8] A. Todri, M. Marek-Sadowska, and J. Kozhaya, "Power supply noise aware workload assignment for multi-core systems," in *Computer-Aided Design ICCAD '08. IEEE/ACM International Conference on*, 2008, pp. 330–337.

[9] W. J. Dally and B. Towles, "Route packets, not wires: On-chip interconnection networks," in *Design Automation Conference. Proceedings*. IEEE, 2001, pp. 684–689, design Automation Conference, 2001. Proceedings.

[10] H. Wang, L.-S. Peh, and S. Malik, "Power-driven design of router microarchitectures in on-chip networks," in *Microarchitecture, 2003. MICRO-36. Proceedings. 36th Annual IEEE/ACM International Symposium on*, dec. 2003, pp. 105 – 116.

[11] S. Vangal, J. Howard, G. Ruhl, S. Dighe, H. Wilson, J. Tschanz, D. Finan, P. Iyer, A. Singh, and T. Jacob, "An 80-tile 1.28 tflops network-on-chip in 65nm cmos," in *Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International*. IEEE, 2007, pp. 98–589, solid-State Circuits Conference. ISSCC '07. Digest of Technical Papers. IEEE International.

[12] P. Vellanki, N. Banerjee, and K. S. Chatha, "Quality-of-service and error control techniques for mesh-based network-on-chip architectures," *Integr. VLSI J.*, vol. 38, no. 3, pp. 353–382, Jan. 2005. [Online]. Available: http://dx.doi.org/10.1016/j.vlsi.2004.07.009

[13] H. Zimmer and A. Jantsch, "A fault model notation and error-control scheme for switch-to-switch buses in a network-on-chip," in *Hardware/Software Codesign and System Synthesis, 2003. First IEEE/ACM/IFIP International Conference on*, oct. 2003, pp. 188 –193.

[14] C. Grecu, L. Anghel, P. Pande, A. Ivanov, and R. Saleh, "Essential fault-tolerance metrics for NoC infrastructures," in *On-Line Testing Symposium IOLTS 07. 13th IEEE International*, july 2007, pp. 37 –42.

[15] W. Hang-Sheng, Z. Xinping, P. Li-Shiuan, and S. Malik, "Orion: a power-performance simulator for interconnection networks," in *Microarchitecture. (MICRO-35). Proceedings. 35th Annual IEEE/ACM International Symposium on*, 2002, pp. 294–305.

[16] C. Seiculescu, S. Murali, L. Benini, and G. De Micheli, "Sunfloor 3d: A tool for networks on chip topology synthesis for 3-d systems on chips," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 29, no. 12, pp. 1987–2000, 2010.

[17] A. B. Kahng, B. Li, L. S. Peh, and K. Samadi, "Orion 2.0: A power-area simulator for interconnection networks," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–5, 2011.

[18] A. Tapani, A. S. David, T. enza, B. Hong, and N. Jari, "Topology optimization for application-specific networks-on-chip," 2004, 966758 53-60.

[19] X. Jiang, W. Wayne, H. Joerg, and C. Srimat, "A design methodology for application-specific networks-on-chip," *ACM Trans. Embed. Comput. Syst.*, vol. 5, no. 2, pp. 263–280, 2006, 1151076.

[20] J. Hu and R. Marculescu, "Energy- and performance-aware mapping for regular NoC architectures," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 24, no. 4, pp. 551 – 562, 2005.

[21] S. Murali and G. De Micheli, "Bandwidth-constrained mapping of cores onto NoC architectures," in *Proceedings of the conference on Design, automation and test in Europe - Volume 2*, ser. DATE '04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 896 – 901 Vol.2.

[22] S. Bodapati and F. N. Najm, "High-level current macro-model for power-grid analysis," in *39th Design Automation Conference Proceedings.*, 2002, pp. 385–390.

[23] S. Kvatinsky, E. G. Friedman, A. Kolodny, Scha, x, and L. chter, "Power grid analysis based on a macro circuit model," in *26th Convention of IEEE in Israel*, 2010, pp. 708–712.

[24] J. N. Kozhaya, S. R. Nassif, and F. N. Najm, "A multigrid-like technique for power grid analysis," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 21, no. 10, pp. 1148–1160, 2002.

[25] A. Goyal and F. N. Najm, "Efficient rc power grid verification using node elimination," in *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, 2011, pp. 1–4.

[26] M. Zhao, R. V. Panda, S. S. Sapatnekar, and D. Blaauw, "Hierarchical analysis of power distribution networks," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 21, no. 2, pp. 159–168, 2002.
[27] J. M. S. Silva, J. R. Phillips, and L. M. Silveira, "Efficient representation and analysis of power grids," in *Design, Automation and Test in Europe*, 2008, pp. 420–425.
[28] L. Hang, J. Fan, Q. Zhenyu, S. X. D. Tan, W. Lifeng, Y. Cai, and X. Hong, "Partitioning-based approach to fast on-chip decoupling capacitor budgeting and minimization," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 25, no. 11, pp. 2402–2412, 2006.
[29] Y. Boyuan, S. X. D. Tan, C. Gengsheng, and W. Lifeng, "Modeling and simulation for on-chip power grid networks by locally dominant krylov subspace method," in *Computer-Aided Design ICCAD. IEEE/ACM International Conference on*, 2008, pp. 744–749.
[30] B. Boghrati and S. Sapatnekar, "A scaled random walk solver for fast power grid analysis," in *Design, Automation and Test in Europe Conference and Exhibition (DATE), pages = 1-6, keywords = algebra integrated circuit design linear algebraic equations on-chip power grids power grid analysis random walk solver, year = 2011*.
[31] L.-R. Zheng and H. Tenhunen, "Fast modeling of core switching noise on distributed lrc power grid in ulsi circuits," *Advanced Packaging, IEEE Transactions on*, vol. 24, no. 3, pp. 245 –254, aug 2001.
[32] S. Tuuna, L. R. Zheng, J. Isoaho, and H. Tenhunen, "Modeling of on-chip bus switching current and its impact on noise in power supply grid," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 16, no. 6, pp. 766–770, 2008.
[33] P. P. Sotiriadis and A. P. Chandrakasan, "A bus energy model for deep submicron technology," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 10, no. 3, pp. 341–350, 2002.
[34] J. Chen and L. He, "A decoupling method for analysis of coupled rlc interconnects," in *Proc. IEEE/ACMInt. Great Lakes Symp. VLSI,*. IEEE, 2002, pp. 41–46, proceedings of the 12th ACM Great Lakes symposium on VLSI.
[35] C. Po-Hao and C. Jia-Ming, "A decoupling technique on switch factor based analysis of rlc interconnects," in *Electro/Information Technology IEEE International Conference on*, 2007, pp. 73–78.
[36] F. Fazzino, M. Palesi, and D. Patti, "Noxim: Network-on-chip simulator," *URL: http://sourceforge.net/projects/noxim*, 2008.
[37] F. X. Nizar Dahir, Terrence Mak and A. Yakovlev, "NoC power supply analysis tool," 2012. [Online]. Available: http://www.staff.ncl.ac.uk/terrence.mak/
[38] A. Agarwal, K. Chopra, D. Blaauw, and V. Zolotov, "Circuit optimization using statistical static timing analysis," in *Design Automation Conference '05. Proceedings. 42nd*, june 2005, pp. 321 – 324.
[39] P. Bogdan and R. Marculescu, "Non-stationary traffic analysis and its implications on multicore platform design," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 30, no. 4, pp. 508 –519, april 2011.
[40] ——, "Statistical physics approaches for network-on-chip traffic characterization," in *Proceedings of the 7th IEEE/ACM international conference on Hardware/software codesign and system synthesis*, ser. CODES+ISSS '09. New York, NY, USA: ACM, 2009, pp. 461–470. [Online]. Available: http://doi.acm.org/10.1145/1629435.1629498
[41] M. S. Gupta, J. L. Oatley, R. Joseph, W. Gu-Yeon, and D. M. Brooks, "Understanding voltage variations in chip multiprocessors using a distributed power-delivery network," in *Design, Automation and Test in Europe Conference and Exhibition. DATE '07*, 2007, pp. 1–6.
[42] F. Mohamood, M. B. Healy, L. Sung Kyu, and H. H. S. Lee, "Noise-direct: A technique for power supply noise aware floorplanning using microarchitecture profiling," in *Design Automation Conference. ASP-DAC '07. Asia and South Pacific*, 2007, pp. 786–791.
[43] D. S. R. Nassif, "IBM power grid benchmarks," 2008.
[44] N. K. Joseph, "Fast power grid simulation," in *37th Conference on Design Automation (DAC'00)*, R. N. Sani, Ed., vol. 0, 2000, pp. 156–161.
[45] N. Khan, S. Alam, and S. Hassoun, "Power delivery design for 3-d ics using different through-silicon via (tsv) technologies," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 19, no. 4, pp. 647 –658, april 2011.
[46] Y. Cao, T. Sato, D. Sylvester, M. Orshansky, and C. Hu, "Predictive technology model," *Nanoscale integration and modeling group, Arizona State Univerity, http://ptm.asu.edu/*, 2006.
[47] H. Jingcao and R. Marculescu, "Energy-aware mapping for tile-based NoC architectures under performance constraints," in *Design Automation Conference, 2003. Proceedings of the ASP-DAC 2003. Asia and South Pacific*, 2003, pp. 233–239.
[48] S. Kirolos, Y. Massoud, and Y. Ismail, "Accurate analytical delay modeling of cmos clock buffers considering power supply variations," in *Circuits and SystemsISCAS. IEEE International Symposium on*, 2008, pp. 3394–3397.
[49] J. Choi, M. Swaminathan, N. Do, and R. Master, "Modeling of power supply noise in large chips using the circuit-based finite-difference time-domain method," *Electromagnetic Compatibility, IEEE Transactions on*, vol. 47, no. 3, pp. 424 – 439, aug. 2005.
[50] S. Kirolos, Y. Massoud, and Y. Ismail, "Power-supply-variation-aware timing analysis of synchronous systems," in *Circuits and Systems ISCAS. IEEE International Symposium on*, 2008, pp. 2418–2421.
[51] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, *Digital integrated circuits : a design perspective*, ser. Prentice Hall electronics and VLSI series. Pearson Education, Jan. 2003.

Nizar S. Dahir (S'12) received both B.Sc. and M.Sc. degrees in Computer Engineering from AL-Nahrain University Baghdad/Iraq in 1997 and 2000, respectively. He is currently working toward the Ph.D. degree with the School of Electrical, Electronic and Computer Engineering, Newcastle University, UK. His research interests include variations tolerant VLSI design and Networks-on-Chip.

Terrence Mak (S'05-M'09) received both B.Eng. and M.Phil. degrees in Systems Engineering from the Chinese University of Hong Kong in 2003 and 2005, respectively, and the Ph.D. degree from Imperial College London in 2009. He joined the School of Electrical, Electronic and Computer Engineering at Newcastle University as a lecturer from 2010 until 2012. He is currently with the Department of Computer Science and Engineering, Chinese University of Hong Kong. During his Ph.D., he worked as a Research Engineer Intern in the VLSI group at Sun Microsystems Laboratories in Menlo Park, California. He also worked as a Visiting Research Scientist in the Poon's Neuroengineering Laboratory at MIT. He was the recipient of both the Croucher Foundation Scholarship and the US Navel Research Excellence in Neuroengineering in 2005. In 2008, he served as the Co-Chair of the UK Asynchronous Forum, and in March 2008 he was the Local Arrangement Chair of the Fourth International Workshop on Applied Reconfigurable Computing. His research interests include FPGA architecture design, Network-on-Chip, reconfigurable computing and VLSI design for biomedical applications.

Fei Xia is a Senior Research Associate with the School of Electrical, Electronic, and Computer Engineering, Newcastle University, UK. His research interests include the design, modelling, and analysis of microelectronic systems focusing on asynchronous systems and data communication in such systems. Dr Xia holds a BEng in Automation from Tsinghua University, an MSc in Control Systems from the University of Alberta, and a PhD in Electronic Engineering from King's College London.

Alexandre (Alex) Yakovlev received the MSc and PhD degrees from St. Petersburg Electrical Engineering Institute in 1979 and 1982, respectively, where he worked in the area of asynchronous and concurrent systems since 1980 and the DSc degree from Newcastle University in 2006. In the period between 1982 and 1990, he held positions of assistant and associate professor at the Computing Science Department. Since 1991, he has been at the Newcastle University, where he worked as a lecturer, reader, and professor at the Computing Science Department until 2002, and is now heading the Microelectronic Systems Design Research Group (http://async.org.uk) at the School of Electrical, Electronic and Computer Engineering. His current interests and publications are in the field of modeling and design of asynchronous, concurrent, real-time, and dependable systems on a chip. He has published four monographs and more than 200 papers in academic journals and conferences, has managed over 25 research contracts. He has chaired program committees of several international conferences, including the IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC), Petri nets (ICATPN), Applications of Concurrency to Systems Design (ACSD), and is currently a chairman of the Steering committee of the Conference on Application of Concurrency to System Design. He is a senior member of the IEEE and a member of the IET. In April 2008, he was general chair of the 14th ASYNC Symp. and 2nd Int. Sympo. on NoCs, and tutorial chair at Design Automation and Test in Europe (DATE) in 2009.