Optimizing the Location of ECC Protection in Network-on-Chip

Junshi Wang[1,2], Letian Huang[1], Qiang Li[1], Guangjun Li[1], Axel Jantsch[2]
[1]University of Electronic Science and Technology of China, Chengdu, China, 611731
[2]TU Wien, Vienna, Austria, 1040
wangjsh@std.uestc.edu.cn, {huanglt, gjl}@uestc.edu.cn, axel.jantsch@tuwien.ac.at

ABSTRACT

The communication in Network-on-Chips (NoCs) may be subject to errors. Error Correcting Codes (ECCs) can be used to tolerate the transient faults in flits caused by Single Event Upsets (SEU). ECC can improve the reliability of a NoC significantly at the cost of extra area and power consumption. However, ECC units (encoders and decoders) may also suffer from SEU faults and thus may lead to *over-protection*, meaning that providing more ECC units does not further improve reliability.

This work analyzes reliability in NoCs, i.e. fractions of correctly received flits, considering the SEU errors introduced by both protected circuits and ECC units. The results show the potential for over-protection. Based on this analysis, we maximize the protection by optimizing the location of the ECC units. We study the reliability of an 8×8 Mesh NoC with six ECC protection strategies, and we conclude that one protection strategy called SLOPE achieves the best trade-off among the six examined strategies by considering reliability, latency, area, energy consumption and design space comprehensively.

CCS Concepts

•**Hardware** → **Network on chip; Fault tolerance;** *Transient errors and upsets;*

Keywords

Error Correcting Code; Single Event Upset; Protection Strategy; Network-on-Chip

1. INTRODUCTION

Network-on-Chips (NoCs) should be able to deliver packets efficiently and correctly, even when facing physical failure phenomena, like crosstalk, Single Event Upsets (SEUs) and wear-out effects [1]. Crosstalk and SEUs usually cause transient faults while aging problems lead to intermittent or permanent faults [2]. Not only memory units but also logic circuits can suffer from SEUs caused by radiation. Moreover, the SEUs in logic circuits increase steadily and become more critical as technology scales [3, 4, 5]. With decreasing transistor size the contribution of the combinational logic to the chip-level Soft Error Rate (SER) may be as high as 50% at 28nm technology [6].

One router in NoCs consists of the data path (links, buffers, crossbars) and the control path (arbiters, allocators, routing calculation circuits). Because the data path occupies more area than the control path, the probability of faults is higher for the data path. As the most common fault detection, diagnosis, and fault-tolerance methods, Error Correcting Codes (ECCs) can tolerate a limited number of transient faults and trigger higher level fault-tolerance methods by introducing information redundancy [7, 8]. Even without any other fault tolerance method, ECC is necessary to provide basic protection and detection [2].

Different kinds of ECC designs in NoCs can be identified by the *Coding Method* and the *Protection Strategy.* The selected coding method determines the number of correctable and detectable error bits. Most popular codes in NoCs are Single-Error-Correcting and Double-Error-Detecting Codes (SEC-DED) [9, 10, 11]. Naturally, reliability can be increased by using more powerful ECC methods.

One decoder can correct the errors introduced by the circuit before it, which is called *protected domain.* Protection strategies dominate the locations of ECC units, and they can achieve different levels of fault-tolerance by configuring the protected domains. Two typical strategies are Hop-to-Hop (H2H) [7, 12] and End-to-End (E2E) [11]. Under the E2E strategy, only Network Interfaces (NIs) provide encoders and decoders. The protected domain contains all the components from source to destination. The H2H strategy places ECC units in each router and the protected domains only include the components from one router to the next.

One common shortage of published researches is that the ECC units (encoders and decoders) are often assumed to be error free. Under this assumption, the reliability can be improved steadily by increasing the capacity and the number of ECC units in the network up to H2H. H2H results in better reliability, larger area, and higher power consumption. However, this assumption is an idealization. Because the SER affecting combinational logic increases in advanced technologies, and due to the simple architecture of the data path, the number of errors introduced by ECC units is comparable to the errors introduced in data paths and should not be ignored [4, 5, 6]. Therefore, it is conceivable that as the protection domains are reduced, the reliability does not

CODES/ISSS '16, October 01-07, 2016, Pittsburgh, PA, USA

DOI: http://dx.doi.org/10.1145/2968456.2968460

increase unlimited. After a certain point providing more ECC units and reducing the protection domain does not lead to a further increase in reliability, which is defined as *over-protection*. Over-protection does not mean that ECC cannot protect the data and improve the correction, but it does suggest the existence of an optimal protection strategy.

Considering the errors introduced by ECC units, we study the reliability of the data path protected by ECCs. Moreover, this work achieves the best reliability by optimizing the protection strategy. Our contributions include:

1. This work describes a simple reliability model of one flit based on the fault parameters of ECC units, links, and routers. The analysis results confirm the existence of over-protection. The analysis results also indicate the protection domain of each decoder should be as similar as possible to achieve the lowest packet error rate.

2. Six ECC protection strategies for mesh networks are described and the protection strategies with the best balance of reliability, latency, power consumption and area are examined under different ECC methods, router architectures, and fault parameters. The simulation results show that the SLOPE strategy can achieve the best trade-off under the conditions considered and simulated.

The paper is organized as follows. Published works on ECCs in NoCs and protection strategies are reviewed in Section 2. Then, the simple reliability model is described and studied in Section 3. Section 4 describes the ECC protection strategies and Section 5 illustrates and discusses the simulation results. Finally, a summary concludes the paper.

2. RELATED WORKS

SEC-DED codes can correct one and detect two erroneous bits. Two typical SEC-DEC methods are Hamming codes and Hasio codes [9, 10, 11]. Except SEC-DED, Other kind of codes, for example, S2SC [12], BCH codes [13], RS codes [14] and 2G4L codes [15] have been considered as well. The advantage of Hamming and Hasio codes is their simple implementations, low hardware area, and low power consumption. The disadvantage is the weak fault-tolerance capacity [16].

SEC-DED can be combined with interleaving to increase the fault-tolerant capacity by avoiding locality effects. In [17], the encoded words based on Hamming(39,32) and Hamming(38,32) codes are interleaved together to build up a triple-error-correction and quadruple-error-detection code. [18] introduces a more complex coding method which combines hamming code, forbidden pattern code (FPC) and interleaving to tolerate multiple continues errors.

As the control paths need the information in the head parts of flits to guide the arbiter and switch, [19] introduces one Unequal Error Protect (UEP) Code to protect the head part of the flit better. In the work, the code can tolerate all single-bit errors and all double adjacent bit errors in the header.

To improve the power efficiency of ECC, [9] proposes to use alternatively one of three coding methods, Strong ECC, Weak ECC, and Power-efficient ECC depending on the number of error bits in the traffic.

Both E2E and H2H have been proposed. In H2H protection, data can be corrected in each router [7, 12] and even with smaller granularity [20]. In E2E protection [13, 15], data are not corrected until they reach the destination. Considering the same ECC method and without considering the errors of ECC units, it is obvious that, H2H can tolerate more faults than E2E and costs more in terms of area, power, and packet delay.

To combine advantages of H2H and E2E, researchers have proposed configurable architectures to switch between these two strategies. In [12], three different protection strategies are selected, including E2E, low area H2H and high-performance H2H. In [21], the protection strategy can switch between H2H and E2E online as each router, and network interfaces contain the codecs.

In [22], although every router has the ECC units, flits are not checked and corrected at each hop. Instead, the packets are checked after passing a specified number of hops. As one counter field is added to the flits to count the number of hops without ECC, this strategy is called COUNTER in the following sections. The packets which cannot be recovered will be retransmitted.

[23] proposes another strategy, which we call SQUARE. The network is divided into several sub-mesh networks, and the ECC units are placed between these sub-mesh networks. We evaluate COUNTER, SQUARE and other protection strategies in our experiments.

To summarize, high fault-tolerance capacity comes with the penalty of area and power consumption. Therefore, the goal of this work is to achieve the best balance between the performance and the cost.

3. RELIABILITY MODEL OF DATA PATH

We use an abstract fault model which approximately captures the occurrence of faults in a particular design of routers and fault parameters of SEU and allows us to study the relevant trade-offs.

3.1 Fault Model and Fault Injection Points

SEUs are described by the changes of signals which are called *Fault Injection Points* (FIPs). Let $e(t)$ be the state of one FIP at the t-th cycle. The value of $e(t)$ can switch between Living (L) and Faulty (F). If one FIP is faulty, the value of this signal is flipped [2]. Otherwise, the value is not changed. We define the living probability of one FIP as

$$P = P\left(e\left(t\right) = L\right). \tag{1}$$

A vector of FIPs $E\left(t\right) = \left\{e_1\left(t\right), e_2\left(t\right), \dots\right\}$ can describe the fault model of multiple signals with the spatial relationship, e.g. the outputs of one module.

FIP abstracts the SEU failures in a circuit with its probabilities determined by the fault model and geometry. The derivation of these probabilities from fault models, while interesting and valuable, is beyond the scope of this paper.

3.2 Fault Model of the Data Path

Figure 1 shows the abstract model of the data path. One flit passes through H routers from the source to the destination. On this trip, the flit traverses the encoder (*enc*), the inter-decoders (*int*) and the finial-decoder (*dec*). The outputs of inter-decoders are still coded words while the outputs of the finial decoder are original information. The inter-decoders split the data path into D segments. Each segment contains one ECC unit (an encoder for the first segment and inter-decoders for the rest), H_i routers and H_i

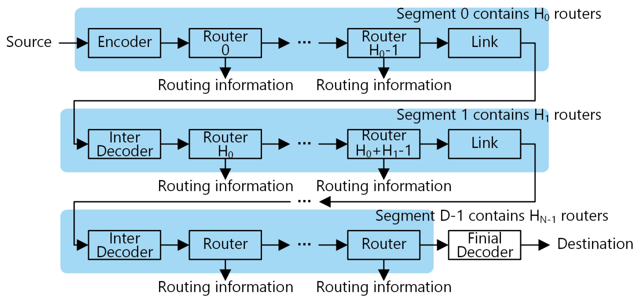


Figure 1: The model of Data path Protected by ECC.

or $H_i - 1$ links. The last segment ends by one router and lacks the last link. The protection strategy for a given path is denoted by the vector $\mathbf{H} = [H_0, H_1, \ldots, H_{D-1}]$. For example, for a path with 8 hops, the E2E strategy is $\mathbf{H} = [8]$ and the H2H strategy is $\mathbf{H} = [1, 1, 1, 1, 1, 1, 1, 1]$.

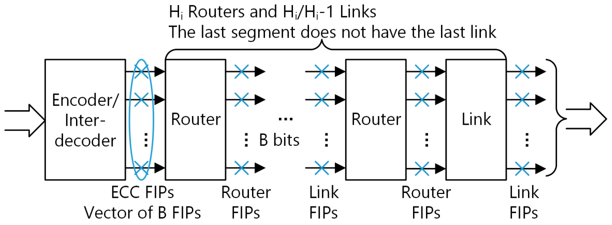


Figure 2: The model of One Segment and Location of FIPs.

ECC methods divide the data path into several subgroups. The model of one segment of one sub-group is shown in Figure 2. The following FIPs are considered (blue crosses in Figure 2):

1. link FIPs: 1-bit FIP located at each bit of links between routers presenting the faults injected by global links;
2. router FIPs: 1-bit FIP located at each bit of each output port presenting the faults injected by input buffers, switches and output buffers;
3. ECC FIPs: 1 vector of FIPs located at the output of ECC units (encoders, inter- or final-decoders) representing the faults injected by the ECC unit.

Fault probabilities of FIPs are calculated by analyzing the netlist and subjecting it to our fault model.

An SEC-DED code is considered in this paper. Each ECC code can correct a one-bit error. For each subgroup, there should be only one error bit at the end of each segment before inter/final decoder. The number of error bits introduced by ECC FIPs must be no more than one bit. Moreover, if there is one error bit at the ECC FIPs (e.g. bit 0), only the link FIPs and router FIPs of the error bit (bit 0) can be faulty. Therefore, one flit must address two conditions to provide correct delivery:

1. All segments of all subgroups must be correct, which includes two cases:
   (a) ECC FIPs introduce no error and no more than 1 bit of the protected circuit is wrong, or
   (b) ECC FIPs introduce one error and the other bits of the protected circuits are all living.
2. The final decoder must be correct.

3.3 Reliability of One Flit

Let P_L and P_R represent the living probability of link FIPs and router FIPs. We assume a homogeneous datapath with equal fault probabilities for all bits. $P_{unit,E}$ presents the probability that outputs of ECC units show errors on bits in set E. The position of *unit* should be the short name of ECC unit, *enc*, *int* or *dec*. Specially, the probability of the ECC unit with one error bit at k-th is $P_{unit,\{k\}}$, and the probability of the ECC unit without any error bit is $P_{unit,\emptyset}$. The fault model of the ECC unit in segment d is as follows.

$$P_{ecc,d,E} = \begin{cases} P_{enc,E} & d = 0 \\ P_{int,E} & 1 \le d \le D-1 \end{cases}. \tag{2}$$

As shown in Figure 2, the reliability model of a protected signal at every bit of d-th segment shows:

$$P_{pro,d} = \begin{cases} P_R^{H_d} P_L^{H_d} & 0 \le d \le D-2 \\ P_R^{H_d} P_L^{H_d - 1} & d = D-1 \end{cases} \tag{3}$$

After that, the correction probability of one segment d is

$$\begin{aligned} P_d = {} & P_{ecc,d,\emptyset} \left(P_{pro,d}^B + B\left(1 - P_{pro,d}\right) P_{pro,d}^{B-1} \right) \\ & + \sum_{k=0}^{B-1} \left(P_{ecc,d,\{k\}} P_{pro,d}^{B-1} \right) \end{aligned} \tag{4}$$

where B is the width of the code word. The first term and the second term are the probability of correction condition 1.a and 1.b, respectively. This equation can be written as:

$$\begin{aligned} P_d = {} & \left(P_{ecc,d,\emptyset} + \sum_{k=0}^{B-1} P_{ecc,d,\{k\}} \right) \\ & \times \left(P_{pro,d}^B + \underline{\underline{B\left(1 - P_{pro,d}\right) P_{pro,d}^{B-1}}} \right) \\ & - \sum_{k=0}^{B-1} P_{ecc,d,\{k\}} \left(B - 1\right) \left(1 - P_{pro,d}\right) P_{pro,d}^{B-1} \end{aligned} \tag{5}$$

Finally, the living probability of one flit is

$$P_{flit} = \left(P_{dec,\emptyset} \prod_{d=0}^{D-1} P_d \right)^G, \tag{6}$$

where G is the number of subgroups.

In summary, the reliability of one flit is dependent on the circuit architecture and the fault models of FIPs. Also, the coding method, the protection strategy and the routing algorithm, that determine the value of B, D and G, can also lead to differences in reliability.

The correction of one flit without ECC protection is

$$P_{flit-no-ecc} = \left(\prod_{d=0}^{D-1} P_{pro,d}^A \right)^G \tag{7}$$

where A is the width before coding.

Comparing Equations (5)-(6) and Equation (7), ECC improves the reliability by increasing the term of the one-bit error in the protected circuits in the equation (double underline part in Equation (5)). On the other hand, as the living probabilities are lower than 1, introducing the living probability of ECC units reduces the value of all terms and the correction of flits, which is the source of the negative effect. So, the relationship between the living probability of ECC units, links, and routers influence the efficiency of ECC designs.

3.4 Analysis of Protection Strategies

The fault model of each FIP is calculated by analyzing the netlist under a given level of faults. In this work, living probability unit (LPU) ρ is defined as the probability of a circuit of 1 μm^2 being alive in 1 cycle. Thus, the living probability of one FIP is defined as

$$P = P\left(e\left(t\right) = L\right) = \rho^{a}. \tag{8}$$

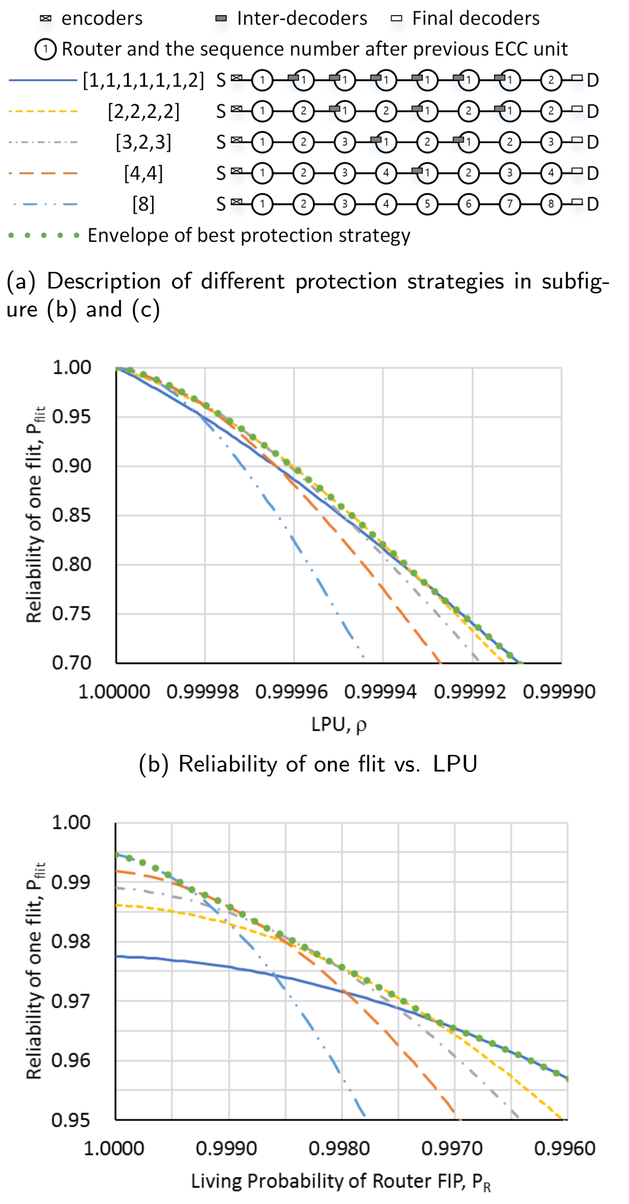
where a is the total area of the FIP in μm^2. Because FIPs are placed at the outputs of a circuit block, a is the sum of gates computing the corresponding output signal, which we derive based on the netlist (giving the number and types of gates) and the technology file (giving the area for each gate). LPU is a constant determined by the physical parameters of the IC. Router architecture and ECC circuits influence the parameters of router FIPs and ECC FIPs, respectively. Therefore, the LPU and the router architecture are two major factors to determine the performance of ECC methods.

A Hamming (7,4) code, one typical SEC-DED method, is examined in this section. Assume one flit with 32 bits before encoding goes through 8 routers from source to destination (8 routers is a common distance in an 8 × 8 network under uniform traffic). The data path with 8 routers has $2^7 = 128$ different possible protection strategies, and all strategies are examined under different LPU and router architectures (presented by different values of the probability of router FIPs) in this section.

Figure 3 gives an intuition for how LPU and router architecture influences the reliability of one flit. The maximum value of reliability among all protection strategies under one specified ρ or P_R is drawn as dotted lines in Figure 3(b) and Figure 3(c), which is also the envelope of all the curves. The curves which match the envelope are drawn in the figures as well. The description of strategies in the figures are provided in Figure 3(a). For a given LPU and router architecture, the protection strategy with the best reliability is called *Best Protection Strategy* (BPS).

Figure 3(b) shows the impact of the LPU on the reliability. LPU ρ reduces from 1.0 to 0.9999. FIPs of routers, links, and ECC units share the same LPU, so the living probability of FIPs drops as well. Figure 3(c) shows the impact of the router architecture on the reliability. The living probability of ECC FIPs is calculated under the LPU $\rho = 0.99999$. The P_R reduces from 1.0 to 0.994 which represents the increasing complexity of one router. Generally, as ρ and P_R reduce, the reliability of flit drops as well. But the rate of flit reliability decrease is different leading to cross-over points.

These two figures illustrate over-protection. For example, in Figure 3(b), when the ρ is 0.99998, the reliability of [1,1,1,1,1,1,2] (blue solid line) is lower than the reliability of [2,2,2,2] (yellow dash line) although [1,1,1,1,1,1,2] provides more ECC units. In fact, [1,1,1,1,1,1,2] does not become the BPS until the data path introduces enough errors ($\rho > 0.99992$ or $P_R > 0.997$). As another example, from $P_R = 0.99868$ to $P_R = 0.99814$ in Figure 3(c), the BPS is [3,2,3]. The protection strategies with smaller protection domains lead to over-protection. On the other hand, the protection strategies with larger protection domains do not provide enough error correction capacity.



(a) Description of different protection strategies in subfigure (b) and (c)

(b) Reliability of one flit vs. LPU

(c) Reliability of one flit vs. router architectures

Figure 3: Reliability of one flit vs. LPU and router architectures

In these two figures, the BPS varies as LPU and router architecture. As the increase of errors on the data path, the BPS shifts from E2E to [1,1,1,1,1,1,2]. Moreover, the protection domain **H** of the BPSs has the lowest variation among all the protection strategies with the same number of segments. For example, if the data path is divided into 2 segments and the BPS is [4,4] with a variation of 0. Also, the BPS with 3 segments is [3,2,3] with a variation of 0.2222,

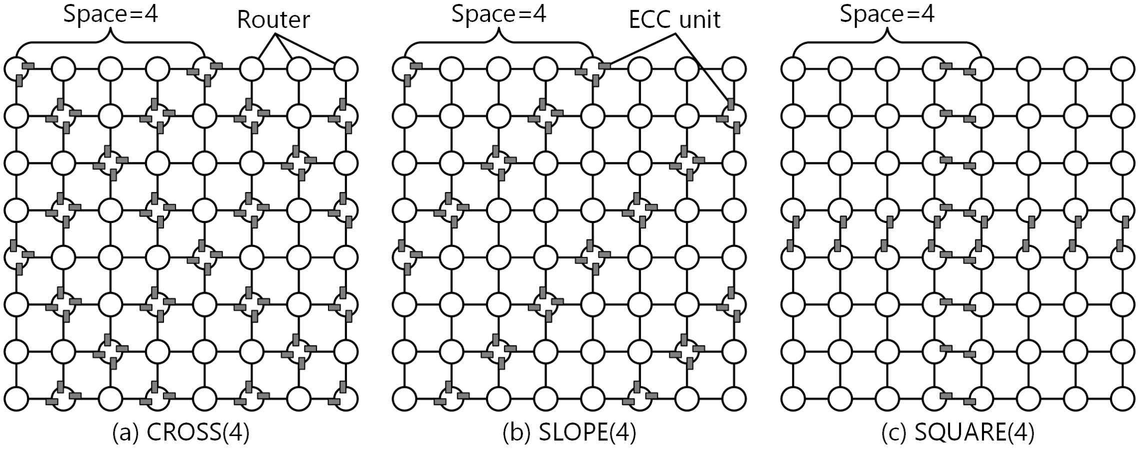


Figure 4: Example of ECC protection strategies with $space = 4$.

which is the lowest as well. This feature can be proved mathematically but which is out of scope of this paper. We draw two main conclusions about the BPF:

1. More errors in the data path need the protection strategy with smaller protection domains. In other words, the BPS has a smaller average of $\mathbf{H}$.
2. Among the protection strategies with the same number of segments, the protection strategy with the lowest variation of protection domains (lowest variation of $\mathbf{H}$) achieves the highest reliability.

Indeed, the specific choice of protection strategy depends on the details of the fault models. However, the general trade-off is driven by the relative ratio of fault occurrences between protected circuits and protection circuits. Hence, we believe our study gives valuable results independent of the specifics of fault models.

4. ECC PROTECTION STRATEGIES

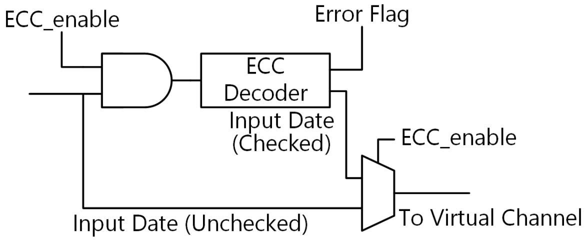


Figure 5: ECC decoder for COUNTER strategy [22]

The encoders and final-decoders are placed at the network interfaces. Protection strategies determine the location of inter-decoders. Inter-decoders can choose every port except the ports connected to the network interface. Using the triple $(x, y, dir = N, S, E, W)$ to denote the port on dir direction of router (x, y).The router at the northwest corner is (0,0). The local ports are not included.

Hop-to-Hop and End-to-End are very common strategies and widely used. The COUNTER strategy is proposed in [22], and the SQUARE strategy is proposed in [23]. Following the conclusions in the previous section, CROSS and SLOPE strategies are proposed in this work. Except H2H and E2E, the location of ECC units and the scale of protection domain can be adjusted by parameter *space*. The definitions of them are as follows.

1. Hop-to-Hop (H2H): Every port has inter-decoders.
2. End-to-End (E2E): No port has inter-decoders.
3. SQUARE($0 < space \leq n$) Strategy (Figure 4(c)): Ports addressing one of following conditions have inter-decoders.

$$dir = N \ \wedge \ y \mod space = 0 \tag{9a}$$
$$dir = S \ \wedge \ (y + 1) \mod space = 0 \tag{9b}$$
$$dir = W \ \wedge \ x \mod space = 0 \tag{9c}$$
$$dir = E \ \wedge \ (x + 1) \mod space = 0 \tag{9d}$$

4. COUNTER($0 < space < 2n$) Strategy (Figure 5): Every port has inter-decoders. Each packet has a counter in the head flit. If the counter is lower than *space*, the ECC unit does not work but increases the counter. If the counter is equal to *space*, the ECC unit corrects the errors in the flit and resets the counter to 0.
5. CROSS($0 < space < 2n-1$, *space* is even or 1) Strategy (Figure 4(a)): All ports of routers (x, y) addressing one of following equations have inter-decoders, except local port.

$$|y - x| \mod space = 0 \tag{10a}$$
$$|space - y - x| \mod space = 0 \tag{10b}$$

6. SLOPE($0 < space < 2n$) Strategy (Figure 4(b)): All ports of routers (x, y) addressing following equation have inter-decoders, except local port.

$$|space - y - x| \mod space = 0 \tag{11}$$

As defined, *space* cannot be lower than 1. If $space = 1$ we have H2H. The protection domains increase as the increase of *space* until *space* is higher than the upper bound. If *space* takes the upper bound, SQUARE and COUNTER strategies are equal to E2E, but CROSS and SLOPE strategies are not. At least, CROSS puts inter-decoders on the routers on diagonal $x = y$ and SLOPE has inter-decoders in the router at the corner $x = 0 \wedge y = 0$.

The **H** vector of a network is the combination of **H**s for all paths. Every pair of source-destination pair is counted so that every protection domain on every path has the same weight. Figure 6 shows the averages and variations of the sizes of protection domains (**H**) for all variants of proposed strategies. Variants belonging to one protection strategy are connected in one line. With the increase of *space*, the average of **H** increases. Because the errors in the data paths are unknown, it is impossible to predict the BPS only according to the average and variation of **H**. However, it is predictable that COUNTER can achieve the best reliability among the described protection strategy because COUNTER has the lowest variation of **H**. On the second and third place are S-LOPE and SQUARE, that achieve good performance in situations with the average of lower than 4. SLOPE has more choices with the average of higher than 4 while SQUARE only has two selectable configurations. CROSS strategy has the lowest reliability and smallest region on average.

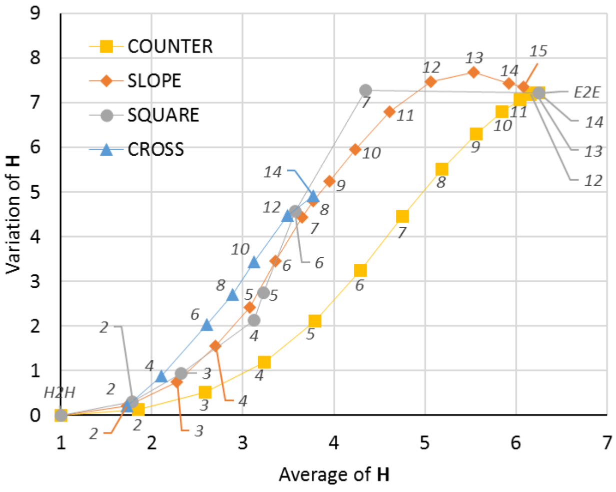


Figure 6: Average and variation of the size of protected domains for variants of different strategies. The number marked at each point in *italic* is the *space* value of variants.

5. SIMULATIONS AND DISCUSSION

In this section, we discuss the effects of LPU, router architectures and ECC methods on the reliability of flits. In each aspect, the performance and cost of these protection strategies are evaluated.

5.1 Simulation Setup

Simulations run on an 8 × 8 Mesh NoCs with 5 physical ports and no virtual channel for each router. The routing algorithm is XY routing algorithm. Packets with 5 flits are injected into the network under uniform traffic profile and a packet injection rate of 0.05 packets/cycle/router.

Our behavior level simulator [24] implements the FIPs, ECC methods, and different protection strategies. Our simulator can execute four Hamming coding methods, which are Hamming(7,4), Hamming(12,8), Hamming(21,16) and Hamming(38,32). Furthermore, we assume the ECC units increase the pipeline delay by one cycle.

During simulation, bits are corrupted at FIPs in links, routers and ECC units. Different from theoretical analysis, the FIPs in the simulator follow the transmission matrix of a Markov Chain:

$$FM = \begin{pmatrix} P_{LL} & P_{LF} \\ P_{FL} & P_{FF} \end{pmatrix}, \qquad (12)$$

where $P_{ij} = P\{e(t+1) = j | e(t) = i\}\ (i, j = L, F)$. As two parameters P_{LL} and P_{FL} can uniquely determine the transmission matrix, (P_{LL}, P_{FL}) is used to present the fault model of one FIP. For the FIPs of ECC units, a big transmission matrix of all combination states is considered. The matrix is calculated from the reliability analysis of the netlist generated by Synopsis Design Compiler (DC) which gives the estimation area, power and netlist of ECC units with 28nm technology.

The reliability is measured by the proportion of packets received without errors, called *delivery rate*. During each simulation, 10^6 packets are injected. Because the status of FIPs changes according to the probability, the simulation results are subject to stochastic variations and therefore the curves are not as smooth as expected.

To examine the delivery rate of the network under different situations, we have run simulations for four different architectures (A_1, A_2, A_3, A_4) and three different LPUs (L_1, L_2, L_3). In the following we present the results of only some of the combinations due to space limits which are listed in Table 1.

Because the relationship between FIPs and the area of circuits, the average area of routers and links for one bit is used to define the different router architectures. Because the scale of one tile determines the length of global links,the area of links for one bit is fixed as 10 μm^2. The router area for one bit takes 30, 100, 200, and 300 μm^2 in four different router architectures respectively. Thus, we abstract the router architectures to only one number, its area, because only the area matters for us under our fault model. The ECC FIPs, Link FIPs, and Router FIPs are calculated under the same LPU, and the results are listed in Table 1 as well. The calculation methods are not within the discussion in this work.

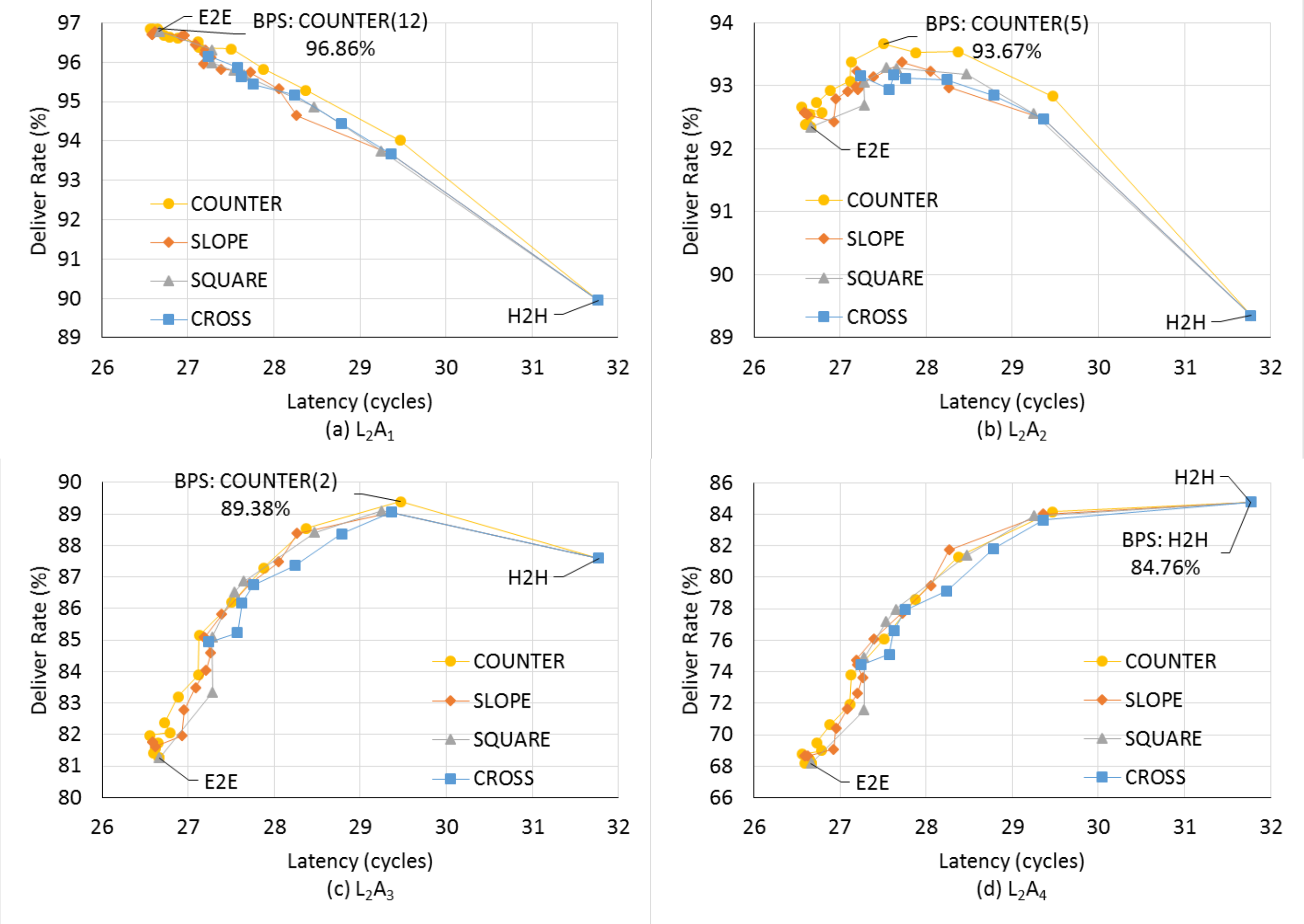
Note, that Table 1 is an abstracted view of the design space that we consider, as the fault model is only determined by LPU and circuit area. The router architecture is represented by the area of a router, which still is a useful abstraction for the purpose of our study.

5.2 Delivery Rate, Latency, Area and Energy for Different Strategies

The best protection strategies should be able to achieve the best balance among reliability, latency, area overhead and power consumption. Moreover, they should have many possible selectable variants to fit different situations. In simulations, we study the different strategies with four router architectures (A_1, A_2, A_3, A_4) in combination with 3 LPUs (L_1, L_2, L_3). However, we report only 6 of the possible 12 combinations in detail, as listed in Table 2, because there

Table 1: Simulation configurations and FIPs

LPU	Router Arch.	LPU (P_{LL}, P_{FL})	Link		Router	
			Area $(\mu m^2/bit)$	FIPs (P_{LL}, P_{FL})	Area $(\mu m^2/bit)$	FIPs (P_{LL}, P_{FL})
L_1	A_2	(0.9999,0.9)	10	(0.99900,0.89914)	100	(0.99005, 0.89064)
L_2	A_1	(0.99999,0.9)	10	(0.99990, 0.89991)	30	(0.99970, 0.89972)
	A_2				100	(0.99900, 0.89906)
	A_3				200	(0.99800, 0.89811)
	A_4				300	(0.99700, 0.89716)
L_3	A_2	(0.999999,0.9)	10	(0.99999,0.90001)	100	(0.99990, 0.89991)



(a) L_2A_1

(b) L_2A_2

(c) L_2A_3

(d) L_2A_4

Figure 7: The reliability of Network-on-Chip with different router architecture and the corresponding average latency.

the trade-offs are best visible.

In Figures 7 and 8, each point shows the delivery rate under one protection strategy variant and the latency and area for its implementation. From the left to right the points belonging to one protection strategy are connected in one line. Therefore, the leftmost points denote E2E, and the rightmost points denote H2H. Reliability does not increase continually with the decrease of protection domains. In L_2A_2 and L_2A_3, the reliability increases at first and drops after the BPS due to the over-protection. In L_2A_1, the reliability reduces directly because the protection domains are too small so that all the protection strategies are in the over-protection zone. The BPS among all strategies for each configuration is marked in Figure 7. Except for COUNTER, the point of maximum delivery rate of each protection strategy is marked in Figure 8.

Figure 7 shows the average latency for different protection strategies. As each ECC unit causes one extra cycle delay, packets going through more segments experience higher latency. Moreover, Figure 7 also illustrates the average power consumption for different protection strategies, because packets going through more segments consume more power. In L_2A_2 and L_2A_3 the protection strategy with the highest reliability also achieves a better trade-off between reliability and cost. Note that all protection strategies reach the highest value at similar latencies.

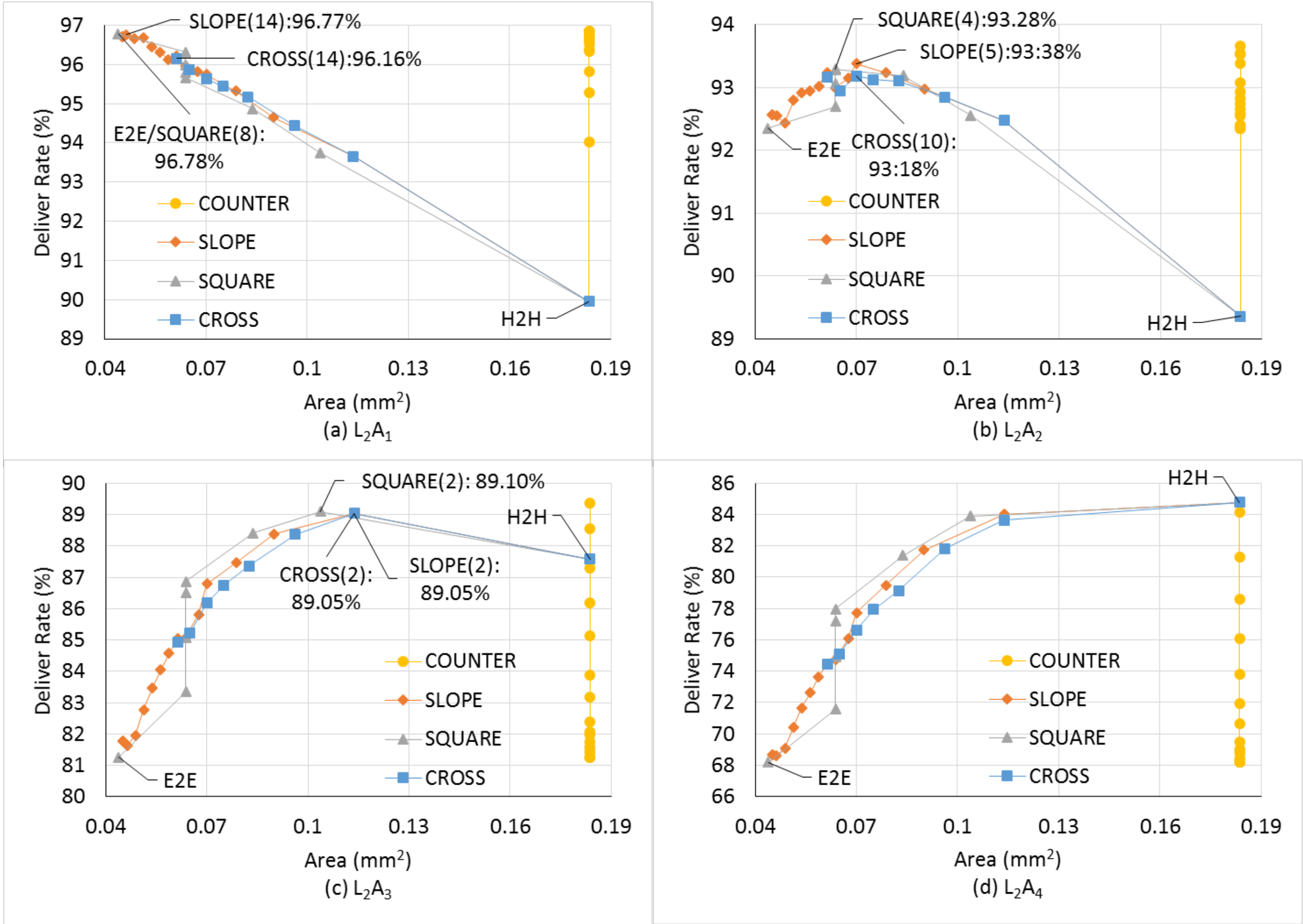


Figure 8: The reliability of Network-on-Chip with different router architecture and the corresponding area of ECC units.

Because these strategies share the same router architecture, Figure 8 only compares the additional area for all ECC units (encoders, inter-decoders, and final-decoders) in the network. The COUNTER strategy needs inter-decoders at each port, so the area of ECC units to implement COUNTER is constant and equal to H2H. Therefore, COUNTER is not every efficient in terms of area. For CROSS, SLOPE and SQUARE the area of ECC units increases with the reduction of *space*, because smaller protection domains need more inter-decoder units.

The delivery rate, latency, area overhead and power consumption of the variant with best reliability are listed in Table 2. As the fault probability of routers increases from L$_2$A$_1$ to L$_2$A$_4$, the BPS moves from E2E to H2H in all strategies.

Reliability. The BPSs in Figure 7 are COUNTER(12), COUNTER(5), COUNTER(2) and H2H, which all belong to the COUNTER strategy due to its lowest variation of the size of protection domains. In L$_2$A$_4$, four strategies have the same performance on delivery rate, latency, and area because the maximum occurs at H2H. Otherwise, CROSS shows more loss in delivery rate than SQUARE and SLOPE. For example, CROSS shows 0.7% loss in delivery rate compared to E2E in L$_2$A$_1$ while SQUARE and SLOPE show delivery rates close to the BPS with no more than 0.4% error.

Area Overhead. In L$_2$A$_1$-L$_2$A$_3$, comparing with BPS of COUNTER strategy, SLOPE, SQUARE, and CROSS can reduce a lot of area overhead. For example, the SLOPE strategy reduces 75.5%, 61.9% and 38.0% of the area of ECC units to implement COUNTER.

Latency and Energy consumption. The latency and energy consumption of the best variants of SLOPE and SQUARE also fluctuate around the latency and power consumption of the best COUNTER variant. The difference of average latency is lower than 1 cycles.

Usage Range. The number of selectable variants of SQUARE is only half of that of SLOPE. SQUARE(n/2-1) to SQUARE(n-1), as SQUARE(4) to SQUARE(7) in the figures, need the same area reducing the design space further. As the *space* number can only be even for CROSS strategy, the CROSS strategy contains the fewest selectable variants.

COUNTER can achieve the highest delivery rate among the tested strategies, but suffers from a large area. On the other hand, SLOPE leads to delivery rates close to COUNTER and reduces the area overhead of ECC units significantly. Moreover, SLOPE provides more selectable variants than CROSS. Therefore, considering delivery rate, latency, area, power consumption and the number of selectable variants, SLOPE is a reasonable choice among the studied protection

Table 2: The Delivery Rate, Latency and Area of BPS for different protection strategies

	COUNTER	SQUARE	SLOPE	CROSS
	Best Protection Strategy (*space*)			
L_2A_1	12	14	14	14
L_2A_2	5	4	5	10
L_2A_3	3	2	2	2
L_2A_4	1/H2H	1/H2H	1/H2H	2
	Delivery Rate (%)			
L_2A_1	96.86	96.78	96.77	96.16
L_2A_2	93.67	93.29	93.38	93.18
L_2A_3	89.38	89.10	89.05	89.05
L_2A_4	84.76	84.76	84.76	84.76
	Latency(cycle)			
L_2A_1	26.65	26.67	26.59	26.62
L_2A_2	27.51	27.53	27.72	27.24
L_2A_3	29.47	29.27	29.36	29.36
L_2A_4	31.77	31.77	31.77	31.77
	Area of ECC units (μm^2)			
L_2A_1	0.184	0.044	0.045	0.061
L_2A_2	0.184	0.064	0.070	0.070
L_2A_3	0.184	0.104	0.114	0.114
L_2A_4	0.184	0.184	0.184	0.184
	Energy of ECC units (nJ)			
L_2A_1	492.79	501.91	474.51	860.94
L_2A_2	1567.38	1116.29	1162.17	1101.12
L_2A_3	2250.98	2115.04	2185.82	2185.82
L_2A_4	3687.53	3687.53	3687.53	3687.53

strategies.

5.3 To Choose the Best SLOPE Configuration

This section studies the design space of SLOPE under different router architectures, LPU, and coding methods. Figure 9 shows the delivery rate for three different LPU. Figures 7 and 9 illustrate that the BPS moves from large protection domains to small protection domains to provide higher error correction capacity, as the errors on the data paths increase caused by the complexity of router architectures and the LPU. For example, the best protection strategy of Hamming(7,4) moves from SLOPE(14) (L_3A_2) to SLOPE(1) (L_1A_2) in Figure 9. Similarly, the best protection strategy moves from SLOPE(14) to SLOPE(1) in Figure 7.

Figure 9 shows the BPS for four coding methods. From the curve of Hamming(7,4) to the curve of Hamming(38,32), the correction capacity of ECC method reduces, and the best protection strategy moves to the strategies with larger protection domains. As shown in Figure 9(b), the BPS of Hamming(7,4) is SLOPE(5) with the lowest density, and the protection strategies of Hamming (38,32) is SLOPE(2) with the smallest protection domain. SLOPE(3) at the middle level is the BPS of Hamming(12,8) and Hamming(21,16). Similarly, in Figure 9(c), SLOPE(14), the BPS of Hamming(7,4), also uses smaller protection domain than SLOPE(11) and SLOPE(12).

6. CONCLUSION

Suffering from Single Event Upsets (SEUs), Network-on-Chips (NoCs) typically use Error Correcting Codes (ECCs) to protect flits on the data path. At the same time, ECC units increase energy consumption, area, delay, and may introduce errors of their own. In this work, the reliability of the data path protected by ECC is studied. The theoretical analysis shows the over-protection phenomenon meaning that the reliability does not increase continually with the increase of ECC units in the Network. Therefore, the protection strategy with the best reliability is determined by the fault parameters, router architecture, ECC coding method, and protection strategy.

To achieve the best trade-off between delivery rate, latency, area overhead, and power consumption, six protection strategies, namely H2H, E2E, CROSS, SLOPE, SQUARE, and COUNTER, are described and evaluated by simulation under different router architectures, fault parameters, and ECC methods. The simulation results illustrate that COUNTER can achieve the best reliability, but at the cost of the high area. We conclude that SLOPE is a good compromise as it exhibits up to 75% less area with only 0.4% loss in delivery rate comparing with COUNTER. Also, SLOPE shows advantages in power, latency and the number of selectable variants.

Indeed, the simulation setups in this paper only approximately model real situations, but designers can still evaluate their design based on the proposed method and actual parameters. Moreover, the router architecture and fault parameters only change the specific values of the results but not the general conclusion.

7. ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their constructive and helpful suggestions and comments. This paper was supported by the National Natural Science Foundation of China (NSFC) under grant No. 61176025, No. 61006027, No. 61534002, and the Oversea Academic Training Funds (OATF), UESTC.

8. REFERENCES

[1] R. Marculescu, U. Y. Ograst, L. Peh, N. E. Jergere, and Y. Hoskote. Outstanding research problems in noc design: system, microarchitecture, and circuit perspectives. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 28(1):3–21, 2009.

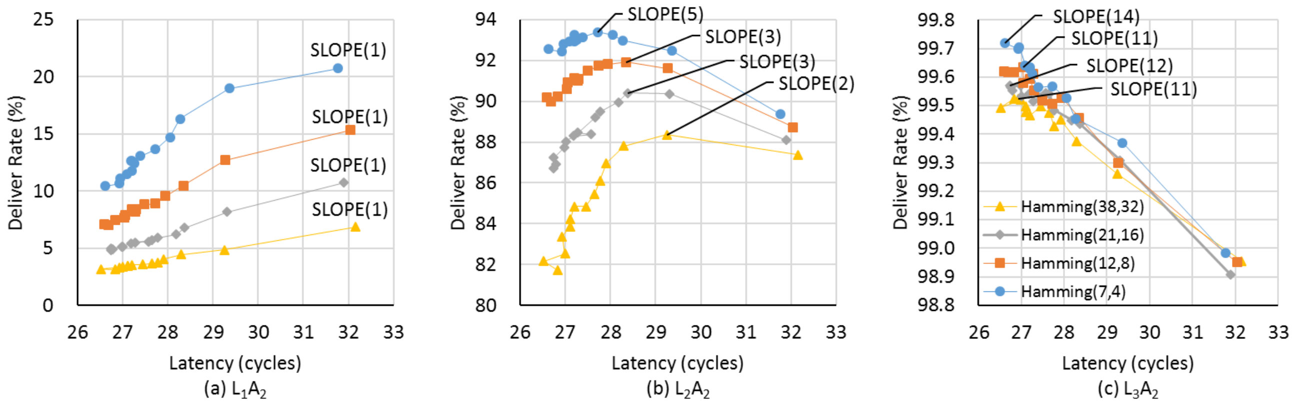
[2] M. Radetzki, C. Feng, X. Zhao, and A. Jantsch. Methods for fault tolerance in networks-on-chip. *ACM Computing Surveys (CSUR)*, 46(1):8, 2013.

[3] International technology roadmap for semiconductors. http://www.itrs2.net, 1012.

[4] P. Shivakumar, M. Kistler, S. W. Keckler, D. Burger, and L. Alvisi. Modeling the effect of technology trends on the soft error rate of combinational logic. In *International Conference on Dependable Systems and Networks*, pages 389–398. IEEE, 2002.

[5] R. Baumann. Soft errors in advanced computer systems. *IEEE Design & Test of Computers*, 22(3):258–266, 2005.

[6] N.N. Mahatme, N.J. Gaspard, S. Jagannathan, and H. Abdel-Aziz T.D. Loveless, B.L. Bhuva, L. W. Massengill, S. Wen, and R. Wong. Estimating the frequency threshold for logic soft errors. In *IEEE International Reliability Physics Symposium (IRPS)*, pages 3D–3. IEEE, 2013.



(a) L_1A_2

(b) L_2A_2

(c) L_3A_2

Figure 9: The reliability of Network-on-Chip with different coding methods and LPU and the corresponding average latency.

[7] D. Fick, A. DeOrio, J. Hu, V. Bertacco, D. Blaauw, and D. Sylvester. Vicis: a reliable network for unreliable silicon. In *Proceedings of the 46th Annual Design Automation Conference*, pages 812–817. ACM, 2009.

[8] J. Wang, M. Ebrahimi, L. Huang, A. Jantsch, and G. Li. Design of fault-tolerant and reliable networks-on-chip. In *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pages 545–550. IEEE, 2015.

[9] T. Boraten and A. Kodi. Energy-efficient runtime adaptive scrubbing in fault-tolerant network-on-chips (nocs) architectures. In *IEEE 31st International Conference on Computer Design (ICCD)*, pages 264–271. IEEE, 2013.

[10] V. Pasca, L. Anghel, C. Rusu, R. Locatelli, and M. Coppola. Error resilience of intra-die and inter-die communication with 3d spidergon stnoc. In *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 275–278. IEEE, 2010.

[11] D. Zamzam, M. Abd El Ghany, K. Hofmann, and M. Ismail. Highly reliable and power efficient noc interconnects. In *NORCHIP*, pages 1–4. IEEE, 2011.

[12] D. Rossi, P. Angelini, and C. Metra. Configurable error control scheme for noc signal integrity. In *13th IEEE International On-Line Testing Symposium (IOLTS)*, pages 43–48. IEEE, 2007.

[13] H. Bokhari, H. Javaid, M. Shafique, J. Henkel, and S. Parameswaran. Supernet: multimode interconnect architecture for manycore chips. In *Proceedings of the 52nd Annual Design Automation Conference*, page 85. ACM, 2015.

[14] I. Datta, D. Datta, and P.P. Pande. Design methodology for optical interconnect topologies in nocs with ber and transmit power constraints. *Journal of Lightwave Technology*, 32(1):163–175, 2014.

[15] S. Shamshiri, A. Ghofrani, and K. Cheng. End-to-end error correction and online diagnosis for on-chip networks. In *IEEE International Test Conference (ITC)*, pages 1–10. IEEE, 2011.

[16] T. Lehtonen, P. Liljeberg, and J. Plosila. Analysis of forward error correction methods for nanoscale networks-on-chip. In *Proceedings of the 2nd international conference on Nano-Networks*, page 3. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007.

[17] A. Ganguly, P.P. Pande, and B. Belzer. Crosstalk-aware channel coding schemes for energy efficient and reliable noc interconnects. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 17(11):1626–1639, 2009.

[18] B. Wang, J. Xie, Z. Mao, and Q. Wang. Multiple continuous error correct code for high performance network-on-chip. In *Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia)*, pages 98–101. IEEE, 2011.

[19] A. Dutta and N. A Touba. Reliable network-on-chip using a low cost unequal error protection code. In *22nd IEEE International Symposium on Defect and Fault-Tolerance in VLSI Systems (DFT)*, pages 3–11. IEEE, 2007.

[20] L. Xie, K. Mei, and Y. Li. Repair: A reliable partial-redundancy-based router in noc. In *IEEE Eighth International Conference on Networking, Architecture and Storage*, pages 173–177. IEEE, 2013.

[21] Q. Yu and P. Ampadu. Dual-layer adaptive error control for network-on-chip links. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 20(7):1304–1317, 2012.

[22] H. Zhao, M. Kandemir, and M.J. Irwin. Exploring performance-power tradeoffs in providing reliability for noc-based mpsocs. In *12th International Symposium on Quality Electronic Design (ISQED)*, pages 1–7. IEEE, 2011.

[23] C. Killian, C. Tanougast, and A. Dandache. Hybrid fault detection for adaptive noc. *IEEE Embedded Systems Letters*, 5(4):69–72, 2013.

[24] J. Wang, L. Huang, G. Li, and A. Jantsch. Visualnoc: Visualization network-on-chipp design framework. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), University Booth Proceedings*, page 13. IEEE, 2016.