Joint Consideration of Fault-Tolerance, Energy-Efficiency and Performance in On-Chip Networks

Alireza Ejlali[*] ejlali@sharif.edu
Bashir M. Al-Hashimi[+] bmah@ecs.soton.ac.uk
Paul Rosinger[+] pmr@ecs.soton.ac.uk
Seyed Ghassem Miremadi[*] miremadi@sharif.edu

[*]Department of Computer Engineering
Sharif University of Technology,
Azadi Ave., Tehran, Iran

[+]School of Electronics and Computer Science
University of Southampton,
Southampton, SO17 1BJ, UK

Abstract

High reliability against noise, low energy consumption and high performance are key objectives in the design of on-chip networks. Recently some researchers have considered the various trade-offs between two of these objectives. However, as we will argue later, the three design objectives should be considered jointly and simultaneously. The first aim of this paper is to analyze the impact of various error-control schemes on the simultaneous trade-off between reliability, performance and energy when voltage swing varies. We provide a detailed comparative analysis of the error-control schemes using analytical models and SPICE simulations. The second aim of this paper is to analyze the impact of noise power and time constraint on the effectiveness of error-control schemes, which have not been addressed in previous studies.

1. Introduction

On-chip networks have been proposed to cope with the ever-increasing complexity and communication requirements of SoCs [1]. Energy consumption is one of the most prominent issues in on-chip networks. It has been shown that on-chip interconnects account for a significant fraction of the total on-chip energy consumption [3]. On the other hand, the required reliability of on-chip interconnects is becoming harder to achieve due to shrinking feature-sizes and supply voltage scaling which makes on-chip interconnects more sensitive to noise [2]. To address the energy consumption issue, reduced voltage swing [3,4] is often used. However, reduced voltage swing leads to decreased noise margin; making interconnects less immune to noise. Reduced voltage swing also necessitates the reduction of interconnect operational frequency [3] which leads to performance reduction. To address the reliability issue, error-control schemes such as Automatic Repeat Request (ARQ), and Forward Error Control (FEC) can be used [2,3]. However, these mechanisms increase the energy consumption and can degrade the performance of the on-chip networks. Based on the above, high performance, high reliability and low energy consumption are conflicting objectives that require to be considered jointly when designing an on-chip network.

In the context of on-chip communication, the energy efficiency of the FEC and ARQ error-control schemes has been studied in [2]. This research has reported that, for the same constraint on system reliability, ARQ consumes less energy than FEC. However, this research has not considered the performance. Indeed, it has been assumed that timing penalties can be tolerated. Furthermore, this research has not considered the hybrid ARQ/FEC (HARQ) scheme. A dynamic voltage swing approach has been proposed in [3] to optimize the energy consumption of the ARQ scheme without degrading the performance and the reliability. However, this research has not considered the FEC and HARQ schemes. Recently [8] has compared the ARQ and HARQ schemes. This work provides useful information to select an appropriate error control scheme for a given application. However, it addresses energy/reliability and performance/reliability trade-offs separately and does not consider the impact of swing voltage on the simultaneous trade-off between reliability, performance, and energy consumption.

We believe that it is necessary to consider the three design objectives jointly and simultaneously. It is insufficient to study energy/reliability, energy/performance, and reliability/performance trade-offs separately, since in each study one may find some design candidates as optimum solutions while these design candidates may be ineffective to achieve the third objective (Section 3.2.2). Also, two other important issues which have not been addressed in all previous works [2,3,8] are the impacts of (i) time constraints and (ii) noise power on the effectiveness of the error-control schemes.

In this paper, we aim: (i) to analyze the impact of voltage swing and different error control schemes on the simultaneous trade-off between reliability, energy and performance, and (ii) to answer the following question: "If a message transmission has to be finished in a given time interval (time constraint) and in the presence of noise with a given power, which error control scheme and what voltage swing must be used to perform the transmission with the minimum energy and highest reliability?". It should be noted that the aim of the paper is not to propose any new error control scheme, rather to identify an appropriate error control scheme (among the existing ones) and to select a proper voltage swing for given time constraint and noise power. Also note that we concentrate on the physical (e.g. voltage swing) and data-link (e.g. error control) layers.

Based on the concept of performability metric [5], in this paper, we provide a definition of "interconnect performability" to measure the reliability and performance

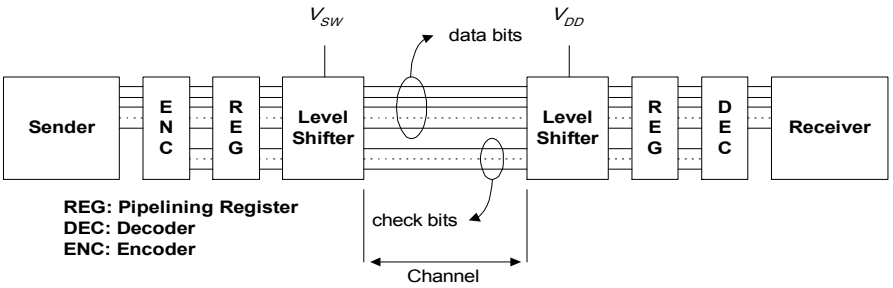


Figure 1: A possible architecture for an on-chip interconnect

of an on-chip network interconnect in a composite way. Using this metric, the tripartite trade-off between the reliability, performance, and energy is transformed to a bipartite trade-off between performability and energy, thereby making the trade-off analysis easier. Furthermore the performability metric helps to analyze the impact of time constraints on the effectiveness of the error control schemes. To consider the energy/performability trade-off, analytical models of performability and energy are developed for three error-control schemes (ARQ, FEC, and HARQ) and the simple non-fault-tolerant communication (SNFT). In the energy analysis, the energy overhead of the error-control circuits, estimated by SPICE simulations, is also considered. We have chosen SNFT to demonstrate why error-control schemes are necessary.

The rest of the paper is organized as follows. Section 2 provides the performability and energy models for communication schemes. Based on the models provided in Section 2, Section 3 analyzes and compares the different communication schemes. Finally, Section 4 concludes the paper.

2. Analytical models of error-control schemes

In on-chip networks, each message to be transmitted is partitioned into packets. Packets in turn are broken into *flits*. A flit can usually be transferred across a physical channel in a single cycle while multiple cycles are used to transfer a single packet [2]. Error control schemes can be implemented at different levels of granularity [2,8]. In packet-level error control, check bits are associated to an entire packet, but in flit-level error control, each flit contains its own check bits and additional wiring resources are used for the check bits [2]. Most of the related works [3,7] use flit-level error control. This is because flit-level error control has relatively lower packet latency and requires less buffer memory at error control circuits [2,8]. Similarly, in this paper we consider flit-level error control. Fig. 1 shows a possible architecture for an on-chip interconnect which uses flit-level error control. In this figure, the pipelining registers let the channel, the encoder, and the decoder units operate in an overlapped manner. Also, the level shifter units are used to change the voltage swing. In the rest of this section, we first introduce three error control schemes (ARQ, FEC, and HARQ), and then we develop the analytical models of performability and energy for the schemes.

2.1. Error control schemes

In this section, the terms "sender" and "receiver" are used to refer to cores or switches that transmit and receive data. The three error-control schemes for on-chip networks, considered in this work, are:

a) ARQ: In this scheme [2], the sender includes an encoder which encodes flits using an error detection code (e.g. CRC-8 code [3]). The receiver includes a decoder which can detect errors. When the receiver detects an error in a flit, it requests the sender to resend the flit. This process is repeated until the flit is error free.

b) FEC: In this scheme [2], the sender includes an encoder that encodes flits using an error correction code which can be used for single-bit error correction (e.g. overlapping parity bits [6]). The receiver includes a decoder which can correct single-bit errors. When the receiver detects a single-bit error in a flit, it corrects the error without any retransmission request. However, on the occasions that there is a multiple-bit error in a flit, it cannot be corrected and the transmission fails.

c) Hybrid FEC/ARQ (HARQ): In this scheme, the sender includes an encoder that encodes flits using an error correction code (e.g. overlapping parity bits [6]). The receiver includes a decoder which can correct single-bit errors and detect multiple-bit errors. When the receiver detects a single-bit error in a flit, it corrects the error without any retransmission request. However, on the occasions that the receiver detects a multiple-bit error in a single flit, it cannot correct the error and hence requests the sender to resend the flit. This process is repeated until the flit is error free.

2.2. Performability model

In this paper, we develop a *performability* metric to measure the performance and reliability of communication schemes in a composite way. There are different performability measures which are used to illustrate the ability of a system to complete useful work in a finite time interval [5]. Since in an on-chip interconnect the useful work is to transmit useful bits (i.e. original data bits excluding check bits), in this paper we define the performability $P(L,T)$ of an on-chip interconnect as the probability to transmit L useful bits during the time interval T in the presence of noise. To see how this definition can be used to combine reliability and performance analysis, consider the ARQ scheme. The presence of erroneous flits (i.e. low reliability) necessitates a more frequent retransmission of flits which requires more time and reduces the probability to finish the transmission of a fixed number of useful bits during a fixed time interval (i.e. performability). Also, reducing the bit rate (i.e. low performance) increases the time required for sending the flits. This time increase reduces the probability to finish the transmission during the time interval (i.e. performability).

Note that for different applications different levels of

performability might be required. For example, in safety-critical applications [6] a system is required to operate correctly with a probability greater than $1\text{-}10^{-9}$ [6]. Hence, the performability of an interconnect which is used for a safety-critical application must be greater than $1\text{-}10^{-9}$.

The analytical performability models for the communication schemes are presented next.

2.2.1. Flit error rate. For on-chip communications the Bit Error Rate (BER) is affected by the voltage swing. This is due to the fact that noise margins decrease as the voltage swing decreases [2,3]. In the context of on-chip interconnects, the relevant literature mostly uses Gaussian noise model [3,7]. In this model, the BER is given by:

$$BER(V_{SW}) = \frac{1}{\sqrt{2\pi}} \int_{\frac{V_{SW}}{2\sigma_N}}^{\infty} e^{\frac{-u^2}{2}} du \tag{1}$$

where V_{SW} is the voltage swing and σ_N^2 is the noise variance. For simple non-fault-tolerant communication (SNFT), when a flit has one erroneous bit, the whole flit is considered corrupted. Therefore the flit error rate, defined as the probability of having a corrupted flit is:

$$FER_{SNFT}(V_{SW}) = 1 - [1 - BER(V_{SW})]^{L_{SNFT}} \tag{2}$$

where L_{SNFT} is the flit size (in bits). Similarly, for ARQ, when a flit has one erroneous bit, the whole flit is considered corrupted and a retransmission is required. Therefore the flit error rate is:

$$FER_{ARQ}(V_{SW}) = 1 - [1 - BER(V_{SW})]^{L_{ARQ}} \tag{3}$$

where L_{ARQ} is the flit size (in bits) in ARQ. For FEC, a flit is considered corrupted when it has more than one erroneous bit. Those flits which have only one erroneous bit are not considered as corrupted flits, since they are recoverable. Therefore the flit error rate is:

$$\begin{aligned} FER_{FEC}(V_{SW}) &= 1 - [1 - BER(V_{SW})]^{L_{FEC}} \\ &- L_{FEC} \cdot BER(V_{SW}) \cdot [1 - BER(V_{SW})]^{L_{FEC}-1} \end{aligned} \tag{4}$$

where L_{FEC} is the flit size (in bits) in FEC. For HARQ, a flit is considered corrupted when it has more than one erroneous bit and a retransmission is required. Those flits which have only one erroneous bit are recoverable and require no retransmission. Therefore the flit error rate is:

$$\begin{aligned} FER_{HARQ}(V_{SW}) &= 1 - [1 - BER(V_{SW})]^{L_{HARQ}} \\ &- L_{HARQ} \cdot BER(V_{SW}) \cdot [1 - BER(V_{SW})]^{L_{HARQ}-1} \end{aligned} \tag{5}$$

where L_{HARQ} is the flit size (in bits) in HARQ.

2.2.2. Performability. Reducing the voltage swing of a channel necessitates the reduction of the operational frequency. When a channel is used at voltage swing V_{SW}, the operational frequency is [3]:

$$F(V_{SW}) = \frac{K_m}{C_L} \frac{(V_{SW} - V_{th})^2}{V_{SW}} \tag{6}$$

where K_m is the driver transistor transconductance, C_L is the channel capacitance, and V_{th} is the threshold voltage of the transistors. Suppose L bits are put into K flits of length L_F bits. Since each flit is transmitted in one cycle, the maximum number of flits which can be transmitted during the time interval T is:

$$N(V_{SW}) = \left\lfloor \frac{T}{1/F(V_{SW})} \right\rfloor = \lfloor T \cdot F(V_{SW}) \rfloor \tag{7}$$

In the schemes with the retransmission capability (ARQ and HARQ), in order to successfully finish the transmission of the L useful bits (i.e. K flits) within the time interval T, we need that at least K transmissions out of the $N(V_{SW})$ possible transmissions are successful. Hence, according to the binomial distribution, the performability $P(L, T)$ is:

$$P(L, T) = \sum_{i=K}^{N(V_{SW})} \binom{N(V_{SW})}{i} (1 - FER)^i \, FER^{N(V_{SW})-i} \tag{8}$$

where, FER is the flit error rate (see Section 2.2.1), and is given either by Eq. 3 or Eq. 5. In the schemes which do not have the retransmission capability (FEC and SNFT), when $K > N(V_{SW})$, this means that there is not enough time to transmit K flits, and therefore performability is 0. On the other hand, when $K \le N(V_{SW})$, there is enough time to transmit K flits, however each flit can only be transmitted once and there is no retransmission. Therefore, the transmission of the K flits will be successful if and only if the only transmission of each flit is successful, whose probability is $(1\text{-}FER)^K$. Therefore, the performability of the FEC and SNFT is:

$$P(L, T) = \begin{cases} 0 & K > N(V_{SW}) \\ (1 - FER)^K & K \le N(V_{SW}) \end{cases} \tag{9}$$

where, FER is given either by Eq. 2 or Eq. 4.

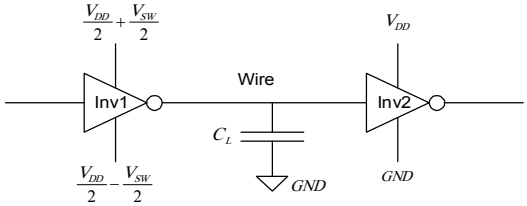


Figure 2. CMOS inverters used as level shifters

2.3. Energy consumption model

The dynamic energy consumption of an on-chip wire per bit is [4]:

$$E_{link}(V_{SW}) = \alpha \cdot C_L \cdot V_{DD} \cdot V_{SW} \tag{10}$$

where α is the switching activity (the probability that the logic value of a channel wire changes), C_L is the wire capacitance, and V_{DD} is the supply voltage. The driver inverter (denoted by Inv1 in Fig. 2) dissipates only a small amount of static energy because when there is no input transition, one of its transistors is always cutoff. This is, however, not true for the receiver inverter (denoted by Inv2 in Fig. 2), whose transistors may never be cutoff because of a low input voltage swing. When the wire shown in Fig. 2 is used at voltage swing V_{SW}, the voltage values $(V_{DD}/2)\text{-}(V_{SW}/2)$ and $(V_{DD}/2)+(V_{SW}/2)$ on the wire represent logic-0 and logic-1 respectively. Assuming that the receiver inverter is symmetric ($\beta_N = \beta_P$ and $V_{th\text{-}N} = -V_{th\text{-}P}$), for both of the wire voltages the current which flows through the receiver inverter is the same. Hence we consider only the case where a logic-0 is on the wire.

Table 1. Power, energy, and delay of error-control hardware*

Error control circuitry		Static Power$^{\Psi}$ (nW)	Dynamic Power (nW)	Total dynamic energy† (fJ)	Dynamic energy per flit† (fJ/flit)	Dynamic energy per useful bit† (fJ/ubit)	Circuit delay (ns)
CRC-8	Encoder	829	1734	793.6	12.4	0.3875	0.37
	Decoder	1107	1888	864	13.5	0.4219	0.37
Overlapping Parity	Encoder	3903	7846	3590.4	56.1	1.7531	1.17
	Decoder	4767	8797	4025.6	62.9	1.9656	2.14

* 2^{11} useful bits were put into 2^6 flits, each containing 32 useful bits
† Dynamic energy per useful bit has been calculated, since it should be inserted in Eq. 14
Ψ Static Power has been estimated, since it should be inserted in Eq. 13

When $(V_{DD}/2)-(V_{SW}/2) < V_{th\text{-}N}$, the N-transistor of the receiver inverter is cutoff. However, when $(V_{DD}/2)-(V_{SW}/2) > V_{th\text{-}N}$, the N-transistor and P-transistor of the receiver inverter are in the saturated and linear regions respectively; hence a considerable current flows through the inverter. This current can be calculated as:

$$I_{REC-Static}(V_{SW}) = \begin{cases} 0 & \frac{V_{DD}}{2} - \frac{V_{SW}}{2} \le V_{th} \\ \frac{\beta}{2}\left(\frac{V_{DD}}{2} - \frac{V_{SW}}{2} - V_{th}\right)^2 & \frac{V_{DD}}{2} - \frac{V_{SW}}{2} > V_{th} \end{cases} \quad (11)$$

where β is the transistor beta parameter. Therefore, the energy consumption per bit, dissipated by this current is:

$$E_{REC-Static}(V_{SW}) = \frac{V_{DD} \cdot I_{REC-Static}(V_{SW})}{F(V_{SW})} \quad (12)$$

Another important source of energy dissipation in on-chip interconnects is the error control circuit. The energy consumption of the error control circuit has two components: static and dynamic. Let P_S be the static power. Since each flit is transmitted in one cycle, the static energy consumption per flit is $P_S/F(V_{SW})$. Hence, the static energy per bit is:

$$E_{CIR-Stat}(V_{SW}) = \frac{P_S}{L_F \cdot F(V_{SW})} \quad (13)$$

where L_F is the flit size (in bits). Let $E_{CIR\text{-}Dyn}$ be the dynamic energy consumption per bit. The total energy per bit, consumed by the error-control circuit, is:

$$E_{CIR}(V_{SW}) = E_{CIR-Dyn} + E_{CIR-Stat}(V_{SW}) \quad (14)$$

Considering all the sources of energy dissipation (Eqs. 10, 12, 14), the total energy consumption per bit which is consumed by both the channel and error-control circuit is:

$$E_{tot}(V_{SW}) = E_{link}(V_{SW}) + E_{REC-Static}(V_{SW}) + E_{CIR}(V_{SW}) \quad (15)$$

In the schemes with the retransmission capability (ARQ and HARQ), if a flit becomes erroneous (with the probability *FER*), the first retransmission is required with the probability of *FER* and the second retransmission is required with the probability of FER^2. Similarly, the i_{th} retransmission will be required with the probability of FER^i. Thus, the expected number of transmissions is:

$$N_T = \sum_{i=0}^{+\infty} FER^i = \frac{1}{1 - FER} \quad (16)$$

Therefore, for the retransmission-based schemes, the expected energy consumption required to transmit a flit is:

$$E_{\text{RT-based}}(V_{SW}) = N_T L_F E_{tot}(V_{SW}) = \frac{L_F E_{tot}(V_{SW})}{1 - FER} \quad (17)$$

where L_F is the flit size, and is equal to either L_{ARQ} or L_{HARQ} and *FER* is given either by Eq. 3 or Eq. 5. In the retransmission-free schemes (FEC and SNFT), each flit is transmitted only once. Hence, in these schemes, the energy consumption required to transmit a flit is:

$$E_{\text{RT-free}}(V_{SW}) = L_F \cdot E_{tot}(V_{SW}) \quad (18)$$

where L_F is the flit size, and is equal to either L_{FEC} or L_{SNFT}.

3. Evaluation of the error-control schemes

In this section we will evaluate the different communication schemes for energy consumption and performability. We first estimate the energy overhead of the error-control circuits, using SPICE simulations. Then we use the analytical models, developed in Section 2 to analyze and compare the communication schemes.

3.1. Energy overhead of error-control circuits

To analyze the energy overhead of the error control circuits, we synthesized the circuits into 45nm SPICE models. The simulations were carried out using 45nm PTM technology [9] (V_{DD}=0.5V). A Cyclic Redundancy Code (CRC-8) with the generator polynomial x^8+1 [2] was synthesized for ARQ, while overlapping parity method [6] was used for FEC and HARQ. A Parallel Bit Code Generator [7] method was employed for carrying out CRC checking. The aim of these experiments is to obtain the energy and power values from the simulation to insert them in the analytical models obtained in Section 2.3, i.e., Eqs. 13 and 14. For Eq. 13, we needed to evaluate the static power P_S and for Eq. 14, we needed to evaluate the dynamic energy per bit $E_{CIR\text{-}Dyn}$. For the evaluation of dynamic energy per bit, some random data bits were encoded and decoded. In order to determine the interconnect operational frequency we also needed to evaluate the delay of the error control circuits (See Section 3.2). The values of energy consumption and circuit delays were obtained using TRANSIENT SPICE analysis. The simulation results are shown in Table 1.

3.2. Analysis of communication schemes

In this analysis, we make the following assumptions: the load capacitance is C_L=1pF (a few millimeters long wire in 45nm technology [10]). Threshold and supply voltage of the circuit are V_{th}=0.11V and V_{DD}=0.5V respectively; noise variance is σ_N=0.06V. Furthermore, we consider a switching activity of α=0.5 (all transmitted bits are independent and equally probable to be 0 or 1). The data that has to be transmitted consists of L=1120 useful bits, which have been split into K=35 flits, each containing 32 useful bits (since CRC-8 and overlapping parity methods have been used, the flit size in ARQ is L_{ARQ}=32+8 bits, while the flit size in HARQ and FEC is

L_{HARQ}=L_{FEC}=32+7 bits). It is also assumed that the data needs to be transferred during the time interval T=700 nS. In Section 3.2.1 and 3.2.2, we will examine the impact of the noise level σ_N, and T parameter (time constraint) on the communication schemes. Note that the minimum delay of a channel with a load capacitance of 1pF (the delay at V_{SW}= V_{DD} =0.5V) is 7.15nS (Eq. 6) which is much higher than the delays of the error control circuit (Table 1). Since the operational frequency is determined by the slowest pipeline stage (Fig. 1), the channel determines the operational frequency of the case study interconnect.

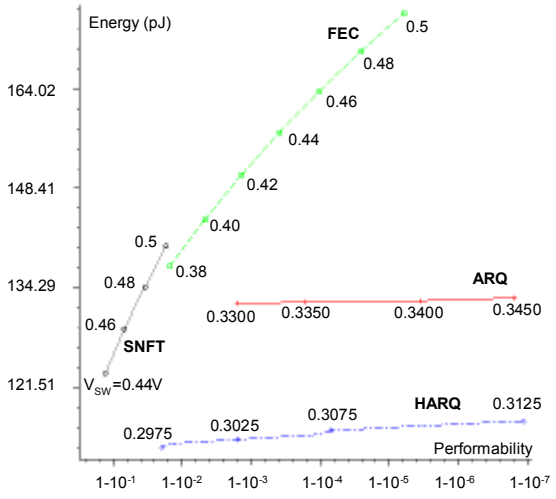


Figure 3. Energy/performability trade-off

Using the analytical models developed in Section 2 (i.e., Eqs. 8,17 for ARQ and HARQ and Eqs. 9,18 for FEC and SNFT), Fig. 3 shows the energy/performability trade-off for the communication schemes. Two main observations are made from Fig. 3:

I) The maximum achievable performability (at V_{SW}=0.5V) from SNFT is less than 1-10^{-2} =0.99, while error-control schemes can provide much better performabilities. Therefore, the usage of error-control schemes is essential in noisy environments to achieve a highly reliable communication. This observation is in line with previous works [2,3,8].

II) For a given performability constraint, HARQ consumes the least energy when compared with the other schemes. For example, if we require a performability more than 1-10^{-3}, we can use ARQ with V_{SW}=0.3350V. However, if we use HARQ with V_{SW}=0.3075V, we will achieve the required performability but with more than 12% energy saving. Note that none of the previous works [2,3,8] has reached to the same conclusion.

3.2.1. Influence of noise power. Fig. 4 shows the energy/performability trade-off of the communication schemes with different values for σ_N. Two key observations are made from Fig. 4:

I) When σ_N=0.04V (Fig. 4a), ARQ and HARQ provide almost the same performability, consume almost the same energy, and are more effective than FEC. However as the link becomes more noisy (Fig. 4b and 4c), HARQ becomes more advantageous than ARQ. We clarify this by means of the following example: Suppose we require a performability more than 1-10^{-4} (i.e. 0.9999). To achieve this level of performability:

- When σ_N=0.04V (Fig. 4a), we can use ARQ with V_{SW}=0.3000V and HARQ with V_{SW}=0.2925. In this case, ARQ offers about 2.1% energy saving over HARQ.
- When σ_N=0.06V (Fig. 4b), we can use ARQ with V_{SW}=0.3400V and HARQ with V_{SW}=0.3075. In this case, HARQ offers about 12.2% energy saving over ARQ.
- When σ_N=0.08V (Fig. 4c), we can use ARQ with V_{SW}=0.3925V and HARQ with V_{SW}=0.3450. In this case, HARQ offers about 22.2% energy saving over ARQ.

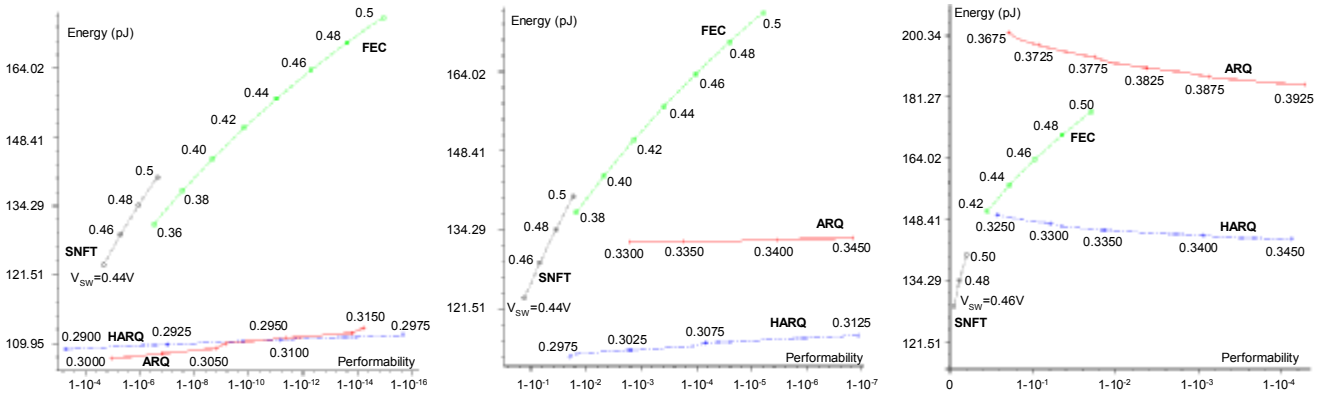
In short, with the performability constraint of 0.9999, as σ_N increases from 0.04V to 0.08V, the energy saving of HARQ over ARQ increases from -2.1% to 22.2%. This is because a strong noise can repeatedly affect the retransmitted flits. Therefore a simple retransmission scheme (i.e. ARQ) is not suitable for a very noisy channel and HARQ, which uses both retransmission and error correction, is more suitable.

II) As shown in Fig. 4c, when the voltage swing of ARQ increases from 0.3675V to 0.3925V, despite the fact that energy consumption per bit increases (See Eq. 10), the energy consumption decreases. This is because as the voltage swing increases, the noise margin increases so that the probability of having faulty flits decreases, therefore the number of flit retransmissions decreases which leads to a reduction in the energy consumption. A similar observation holds for HARQ.

3.2.2 Influence of time constraints. So far, we have analyzed the performability *P(L*=35*32*, T*=700nS*)*. Assuming that L is constant, for the applications which do not have tight time constraints, we can analyze the performability for relatively large T values. However, for the applications with tight time constraints, smaller T values have to be considered. In order to study the impact of the time constraints on the efficiency of the error-control schemes, Fig. 5 shows the energy/performability trade-off of the communication schemes when T=255nS, i.e., in Fig. 5, we consider the performability *P(L*=35*32*, T*=255nS*)*. Two key observations are made from Fig. 5:

I) When we compare Fig. 3 (T=700nS) with Fig. 5 (T=255nS), it can be seen that when T=700nS (relaxed time constraint), ARQ is more effective than FEC. However, when T=255nS (tight time constraint), ARQ is not effective at all and it is even worse than SNFT. This is because, when tight time constraints are imposed, ARQ does not have enough time to retransmit the packets, and hence it is not capable to provide a better performability than SNFT. But, due to the error control circuit and redundant coding bits, it still consumes more energy than SNFT. Note that this observation shows why it is necessary to consider the three design objectives (energy, reliability and performance) jointly and simultaneously. [2] has studied energy/reliability trade-off and reported that for the same constraint on system reliability, ARQ consumes less energy than FEC. This is true and our observation is in agreement with it (Fig. 3) but only when we do not require high performance (relaxed time constraints). It can be seen from Fig. 5 that when we require high performance FEC is better than ARQ.

II) When we compare Fig. 3 (T=700nS) with Fig. 5 (T=255nS), it can be seen that when T=700nS (relaxed



a) σ_N=0.04V b) σ_N=0.06V c) σ_N=0.08V

Figure 4: Influence of noise power

time constraint), HARQ is more effective than FEC. However, when T=255nS (tight time constraint), the HARQ and FEC curves are superimposed which indicates that they provide the same performability and consume the same energy. This is because, when tight time constraints are imposed, HARQ does not have enough time to retransmit the packets, and hence it is not capable to provide a better performability than FEC. Also, since both FEC and HARQ use the same coding, they consume the same energy and provide the same performability.

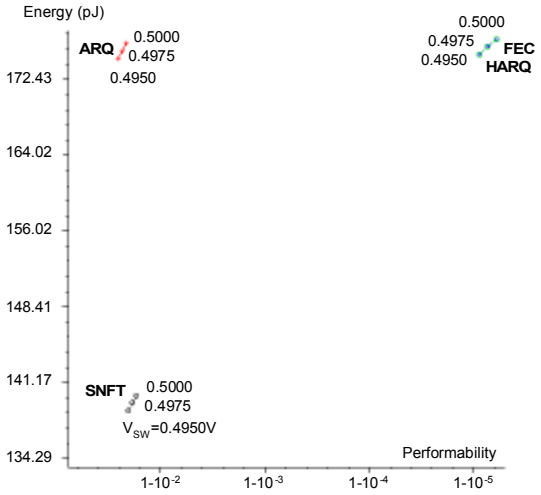


Figure 5: Energy/performability trade-off for an application with tight time constraints

4. Conclusions

We have analyzed the impact of three error-control schemes on the simultaneous trade-offs between three design objectives of on-chip networks (reliability, performance and energy), when voltage swing varies. Since noise power and time constraint vary for different applications and environments, the impacts of these two factors have also been analyzed. This analysis shows that for the environments with relatively low noise power, ARQ is the most preferable choice however as the noise power increases, HARQ proves more advantageous than all the other schemes (Fig. 4). Also, it has been shown that for relaxed time constraints ARQ is better than FEC (Fig. 3) however when tight time constraints are required (Fig. 5) ARQ is not effective at all and FEC is the most preferable choice (in this case FEC and HARQ provide the same performability and consume the same energy but since FEC has simpler implementation than HARQ, FEC is more preferable).

5. Acknowledgements

P. Rosinger, B.M. Al-Hashimi, and A. Ejlali acknowledge the Engineering and Physical Sciences Research Council (EPSRC) for funding this work under grant no. EP/C512804. Also, A. Ejlali, and S.G. Miremadi acknowledge Research Vice-Presidency of Sharif University of Technology for partially funding this work.

References

[1] L.Benini, G.De Micheli, "Networks on chips: a new SoC paradigm," *Computer*, 35(1), pp. 70-78, 2002.

[2] D.Bertozzi *et al.*, "Error Control Schemes for On-Chip Communication Links: The Energy–Reliability Tradeoff", *IEEE Trans. on CAD*, 24(6), pp. 818-831, 2005.

[3] F.Worm *et al.*, "A Robust Self-Calibrating Transmission Scheme for On-Chip Networks," *IEEE Trans. VLSI*, 13(1), pp. 126-139, 2005.

[4] C.Svensson, "Optimum voltage swing on on-chip and off-chip interconnect," *IEEE JSSC*, 36(7), pp. 1108-1112, 2001.

[5] R.M.Smith *et al.*, "Performability Analysis: Measures, an Algorithm, - and a Case Study," *IEEE Trans. Computers*, 37(4), pp. 406-417, 1988.

[6] D.K.Pradhan, *Fault-tolerant computing: theory and techniques*, Prentice Hall, 1986.

[7] D.Bertozzi *et al.*, "Error Control Schemes for On-chip Interconnection Networks: Reliability versus Energy Efficiency," *Networks on Chip*. A.Jantsch, H.Tenhunen, Eds. Kluwer, 2003.

[8] S. Murali *et al.*, "Analysis of Error Recovery Schemes for Networks on Chips", *IEEE Design & Test of Computers*, 22(5), pp. 434-442, 2005.

[9] W. Zhao, Y. Cao, "New generation of Predictive Technology Model for sub-45nm design exploration," Proc. ISQED 2006, pp. 585-590, 2006.

[10] C. Grecu *et al.*, "A scalable communication-centric SoC interconnect architecture", Proc. ISQED 2004, pp. 343-348, 2004.