

Reducing Energy Consumption In Network On Chip Using Data Encoding Techniques

EPPAKAYALA RANJITH⁽¹⁾, MR. M. RAVI KUMAR⁽²⁾
M.Tech, DECS, ACE ENGINEERING COLLEGE⁽¹⁾
Professor, Dept. of ECE, ACE ENGINEERING COLLEGE⁽²⁾

Abstract— As technology improves, the power dissipated by the links of a network-on-chip (NoC) starts to compete with the power dissipated by the other elements of the communications subsystem, namely, the routers and the network interfaces (NIs). In this paper, we present a set of data encoding schemes to reduce the power dissipated by the links of an NoC. The proposed schemes are general and transparent with respect to the underlying NoC fabric (i.e., their application does not require any modification of the routers and link architecture). Experiments carried out on both synthetic and real traffic scenarios show the effectiveness of the proposed schemes, which can reduce no. of transitions from scheme-I to scheme-II & Scheme-II to Scheme-III which in terms reduces power consumption as well as area also.

Index Terms— Coupling switching activity, data encoding, interconnection on chip, low power, network-on-chip (NoC), power analysis.

I. INTRODUCTION

SHIFTING from a silicon technology node to the next technology results in faster and more power efficient gates but slower and more power hungry wires [1]. In fact, more than 50% of the total dynamic power is dissipated in interconnects in current processors, and this is expected to rise to 65%–80% over the next several years [2]. Global interconnect length does not scale with smaller transistors and local wires. Chip size remains relatively constant

because the function of the continues to increase and RC delay increases exponentially. At 32/28 nm, for instance, the RC delay in a 1-mm global wire at the minimum pitch is 25× higher than the intrinsic delay of a two-input NAND fanout of 5 [3]. If the raw computation horsepower seems to be unlimited, thanks to the ability of instancing more and more cores in a single silicon die, scalability issues.

II. OVERVIEW OF THE PROPOSAL

The basic idea of the proposed approach is encoding the flits before they are injected into the network with the goal of minimizing the self-switching activity and the coupling switching activity in the links traversed by the flits. In fact, self-switching activity and coupling switching activity are responsible for link power dissipation. In this paper, we refer to the end-to-end scheme. This end-to-end encoding technique takes advantage of the pipeline nature of the wormhole switching technique. Note that since the same sequence of flits passes through all the links of the routing path, the encoding decision taken at the NI may provide the same power saving for all the links. For the proposed scheme, an encoder and a decoder block are added to the NI. Except for the header flit, the encoder encodes the outgoing flits of the packet such that the power dissipated by the inter-router point-to-point link is minimized. This end-to-end encoding technique takes advantage of the pipeline nature of the wormhole switching technique.

TABLE I

Change of transition types on effect of odd inversion

Time	Normal			Odd Inverted		
	Type I			Types II, III, and IV		
$t-1$	00, 11	00, 11, 01, 10	01, 10	00, 11	00, 11, 01, 10	01, 10
t	10, 01	01, 10, 00, 11	11, 00	11, 00	00, 11, 01, 10	10, 01
	T1*	T1**	T1***	Type III	Type IV	Type II
$t-1$	Type II			Type I		
t	01, 10 10, 01			01, 10 11, 00		
$t-1$	Type III			Type I		
t	00, 11 11, 00			00, 11 10, 01		
$t-1$	Type IV			Type I		
t	00, 11, 01, 10 00, 11, 01, 10			00, 11, 01, 10 01, 10, 00, 11		

In addition, the scheme was based on the hop-by-hop technique, and Hence, encoding/decoding is performed in each node. The scheme presented indealed with reducing the coupling switching. In this method, a complex encoder counts the number of Type I (Table I) transitions with a weighting coefficient of one and the number of Type II transitions with the weighting coefficient of two. If the number is larger than half of the to the complex encoder, the technique only works on the patterns whose full inversion leads to the link power reduction while not considering the patterns whose full inversions may lead to higher link power consumption. Therefore, the link power reduction achieved through this technique is not as large as it could be. This scheme was also based on the hop-by-hop technique. In another coding technique presented in, groups of four bits each are encoded with five bits. The encoded bits wereisolated using shielding wires such that the occurrence of the patterns “101” and “010” were prevented. This way, no simultaneous Type II transitions in two adjacent pair bits are induced. This technique effectivelyreduces the couplingswitching activity. Although the technique reduces the power consumption considerably, it increases the data transfer time,

and hence, the link energy consumption. This is due to the fact that for each four bits, six bits are transmitted which increases the communication traffic. This technique was also based on the hop-by-hop technique. A coding technique that reduces the coupling switching activity by taking the advent age of end-to-end encoding for wormhole switching has been presented in. It is based on lowering the coupling switching activity by eliminating only Type II transitions. In this paper, we present three encoding schemes. In Scheme I, we focus on reducing Type I transitions while in Scheme II, both Types I and II transitions are taken into account for deciding between half and full invert, depending the amount of switching reduction. Finally, in Scheme III, we consider the fact that Type I transitions show different behaviors in the case of odd and even inverts and make the inversion which leads to the higher power saving.

III. PROPOSED ENCODING SCHEMES

In this section, we present the proposed encoding scheme whose goal is to reduce power dissipation by minimizing the coupling transition activities on the links of the interconnection network.

A. Scheme I

In scheme I, we focus on reducing the numbers of Type I transitions (by converting them to Types III and IV transitions) and Type II transitions (by converting them to Type I transition). The scheme compares the current data with the previous one to decide whether odd inversion or no inversion of the current data can lead to the link power reduction. *1) Power Model:* If the flit is odd inverted before being transmitted, the dynamic power on the link is

$$P' \propto T'_{0 \rightarrow 1} + (K1T'1 + K2T'2 + K3T'3 + K4T'4)Cc(1)$$

where $T'_{0 \rightarrow 1}$, $T'1$, $T'2$, $T'3$, and $T'4$, are the self-transition activity, and the coupling transition activity of Types I, II, III, and IV, respectively. Table I reports, for each transition, the relationship between the coupling transition activities of the flit when transmitted as is and when its bits are odd inverted.

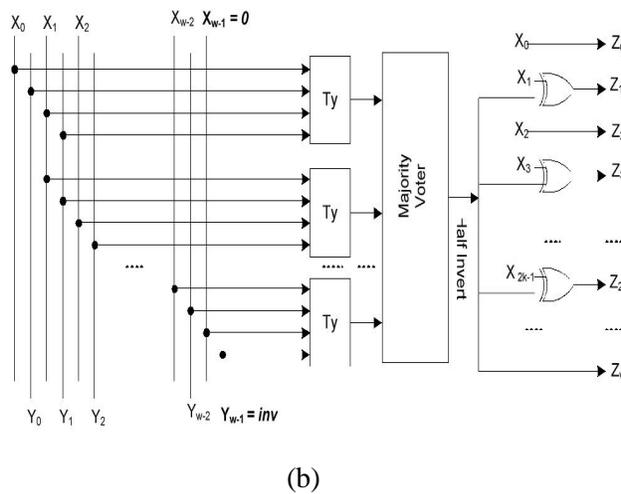
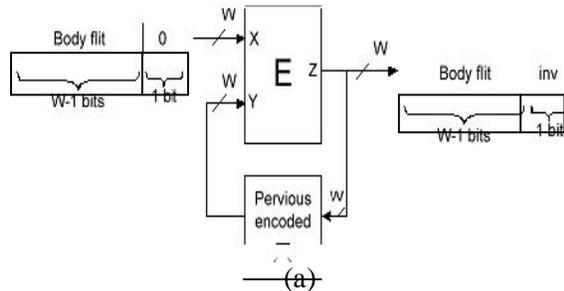


Fig.1 Encoder architecture scheme I. (a) Circuit diagram. (b) Internal view of the encoder block

This presents the condition used to determine whether the odd inversion has to be performed or not.

2) *Proposed Encoding Architecture:* The proposed encoding architecture, which is based on the odd invert condition defined is shown in Fig. 1. We consider a link width of w bits. If no encoding is used, the body flits are grouped in w bits by the NI and are transmitted via the link. In our approach, one bit of the link is used for the

inversion bit, which indicates if the flit traversing the link has been inverted or not. More specifically, the NI packs the body flits in $w - 1$ bits. The encoding logic E , which is integrated into the NI, is responsible for deciding if the inversion should take place and performing the inversion if needed. The decoder circuit simply inverts the received flit when the inversion bit is high.

B. Scheme II

In the proposed encoding scheme II, we make use of both odd (as discussed previously) and full inversion. The full inversion operation converts Type II transitions to Type IV transitions. The scheme compares the current data with the previous one to decide whether the odd, full, or no inversion of the current data can give rise to the link power reduction.

1) *Power Model:* Let us indicate with P , P' , and P'' the power dissipated by the link when the flit is transmitted with no inversion, odd inversion, and full inversion, respectively. The odd inversion leads to power reduction when $P' < P''$ and $P' < P$. The power P'' is given by

$$P'' \propto T1 + 2T4^{**} \quad (2)$$

Neglecting the self-switching activity, we obtain the condition $P' < P''$ as

$$T2 + T3 + T4 + 2T1^{***} < T1 + 2T4^{**} \quad (3)$$

Therefore, using (2) and (3), we can write

$$2(T2 - T4^{**}) < 2Ty - w + 1 \quad (4)$$

Based on the odd inversion condition is obtained as

$$2(T2 - T4^{**}) < 2Ty - w + 1Ty > (w-1)/2(5)$$

Similarly, the condition for the full inversion is obtained from $P'' < P$ and $P''' < P'$. The inequality $P''' < P$ is satisfied.

$$T2 > T4^{**} \quad (6)$$

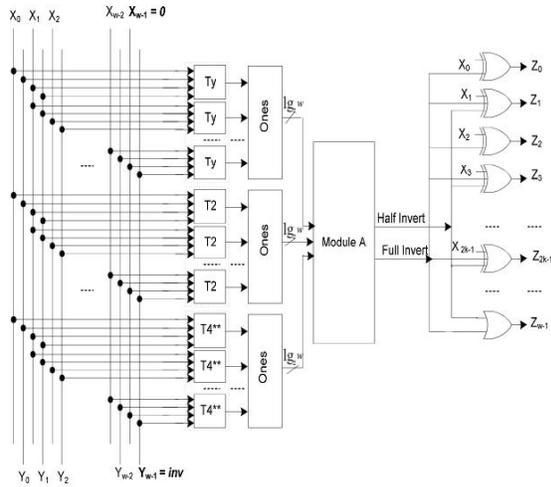


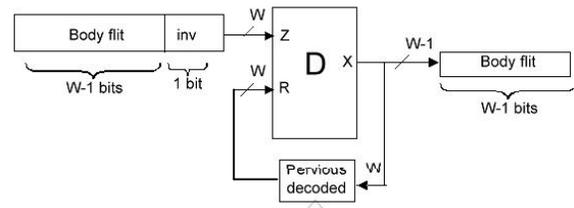
Fig. 2 Encoder architecture of scheme II

Therefore the full inversion condition is obtained as

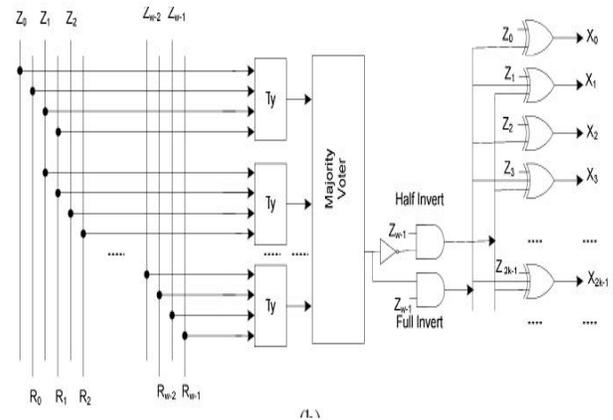
$$2(T2 - T4^{**}) > 2Ty - w + 1 \quad T2 > T4^{**} \quad (7)$$

When none of (6) or (7) is satisfied, no inversion will be performed.

2) *Proposed Encoding Architecture:* The operating principles of this encoder are similar to those of the encoder implementing Scheme I. The proposed encoding architecture, which is based on the odd invert condition of (6) and the full invert condition of (7), is shown in Fig.2 Here again, the w th bit of the previously and the full invert condition of (7) is shown in Fig. 2 Here again, the w th bit of the previously encoded body flit is indicated with inv which defines if it was odd or full inverted ($inv = 1$) or left as it was ($inv = 0$).



(a) Circuit Diagram



(b) Internal Part of Decoder block

Fig. 3 Decoder architecture for Scheme II.

C. Scheme III

In the proposed encoding Scheme III, we add even inversion to Scheme II. The reason is that odd inversion converts some of Type I ($T1^{***}$) transitions to Type II transitions. As can be observed from Table II, if the flit is even inverted, the transitions indicated as $T^{**} 1 / T1^{***}$ in the table are converted to Type IV/Type III transitions. Therefore, the even inversion may reduce the link power dissipation as well.

1) *Power Model:* Let us indicate with P' , P'' , and P''' the power dissipated by the link when the flit is transmitted with no inversion, odd inversion, full inversion, and even inversion, respectively.

The even inversion leads to power reduction when $P''' < P, P''' < P'$, and $P''' < P''$. Based on equations we obtain

$$Te > (w-1)/2, Te > Ty, 2T2 - T4^{**} < 2Te - w + 1 \quad (8)$$

TABLE 2

Change in Transition types on effect of even inversion.

Time	Normal			Even Inverted		
	Type I			Types II, III, and IV		
$t-1$	01, 10	00, 11, 01, 10	00, 11	01, 10	00, 11, 01, 10	00, 11
t	00, 11	10, 01, 11, 00	01, 10	10, 01	00, 11, 01, 10	11, 00
	T1*	T1**	T1***	Type II	Type IV	Type III
$t-1$	Type II			Type I		
t	01, 10			01, 10		
	10, 01			00, 11		
$t-1$	Type III			Type I		
t	00, 11			00, 11		
	11, 00			01, 10		
$t-1$	Type IV			Type I		
t	00, 11, 01, 10			00, 11, 01, 10		
	00, 11, 01, 10			10, 01, 11, 00		

The full inversion leads to power reduction when $P''' < P, P''' < P'$, and $P''' < P''$. Therefore, using (8), the full inversion condition is obtained as

$$2(T2 - T4^{**}) > 2Ty - w + 1, \quad (T2 > T4^{**})$$

$$2(T2 - T4^{**}) > 2Te - w + 1. \quad (9)$$

Similarly, the condition for the odd inversion is obtained from $p' < P, P' < P''$ and $p' < p'''$. The odd inversion condition is satisfied when

$$2(T2 - T4^{**}) < 2Ty - w + 1, Ty > (w-1)/2$$

$$Te < Ty \quad (10)$$

When none of the equations is satisfied, no inversion will be performed.

2) *Proposed Encoding Architecture:* The operating principles of this encoder are similar to those of the encoders implementing Schemes I and II. The proposed encoding architecture,

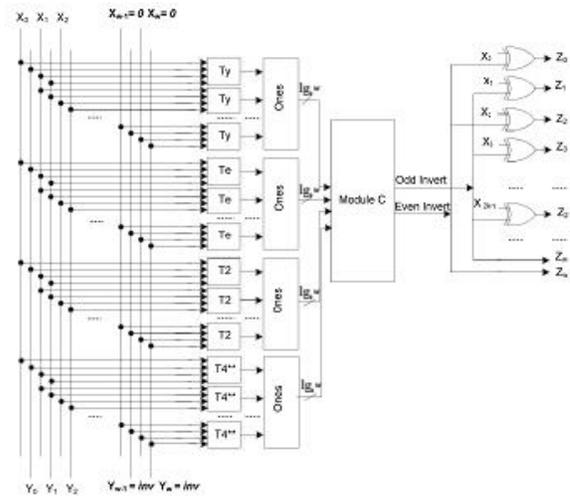


Fig. 4 Encoder architecture for Scheme III.

which is based on the even invert condition of (8), the full invert condition of (9), and the odd invert condition of (10), is shown in Fig. 4. The w th bit of the previously encoded body flit is indicated by inv which shows if it was even, odd, or full inverted ($inv = 1$) or left as it was ($inv = 0$). Similar to the procedure used to design the decoder for scheme II, the decoder for scheme III may be designed.

IV. RESULTS AND DISCUSSION

The proposed data encoding schemes have been assessed by means of a cycle-accurate NoC simulator based on Noxim. The power estimation models of Noxim include NIs, routers, and links. The link power dissipation

was computed using (3) where the terms $T_{0 \rightarrow 1}$, T_1 , and T_2 were computed based on the information obtained from the cycle accurate simulation. The following parameters were used in the simulations. The NoC was clocked at 700 MHz while the baseline NI with minimum buffering and supporting open core protocol 2 and advanced high-performance bus protocols dissipated 5.3 mW. The average power dissipated by the wormhole-based router was 5.7 mW. Based on a 65-nm UMC technology, a total capacitance of 592 fF/mm was assumed for an inter-router wire. About 80% of this capacitance was due to the crosstalk. We assumed 2-mm 32-bit links and a packet size of 16 bytes (eight flits). Using the detailed simulations, when the flits traversed the NoC links, the corresponding self and coupling switching activities were calculated and used along with the self- and coupling capacitance of 0.237 and 0.947 nf, respectively, to calculate the power ($V_{dd} = 0.9$ V and $F_{ck} = 700$ MHz).



Fig.7 Simulation results of scheme III

A. Overheads Due to the Encoder/Decoder Logic

The encoder and the decoder were designed in Verilog HDL described at the RTL level, synthesized with synopsys design compiler and mapped onto an UMC 65-nm technology library.

B. Energy Analysis

To analyze the efficacy of the proposed data encoding schemes in reducing the energy consumption, we consider an 8X8 mesh-based NoC. We only report results for the bit-reversal traffic as for the other synthetic traffics we found similar trends. That is, 0.016 when no data encoding is used, 0.010 for the FPC, and 0.013 for the remaining data encoding schemes. Random data patterns were considered. All the three proposed schemes show energy savings for all the data streams considered in this paper. For this encoding scheme, the maximum of energy and power more than 20% and 60%, respectively, was achieved for the picture workload. Finally, it should be pointed out, in general, that the efficiency of any encoding schemes depends on workload data patterns which are transmitted via the bus.

C. Power Versus Performance

The tradeoff between the reduction of the average power dissipation of the communication system with the completion time (i.e., the amount of the time needed to drain a given amount of traffic volume) is an important characteristic of the system. The percentage increase of completion time is defined as the percentage increase of the time needed to drain

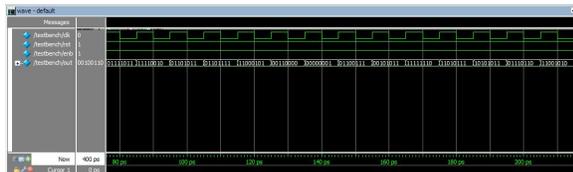


Fig. 5 Simulation results of Scheme I

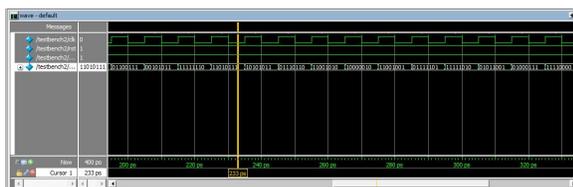


Fig. 6 Simulation results of scheme II

Thus, in the worst case (eight partitions), one additional flit is required to transfer the original four-flit payload. When the FPC is used, additional 11 bits are needed for each encoded flit. Thus, for a four flit payload, we would have 44 additional bits, which require two additional flits. Note that, in the case of the baseline implementation, the network saturation point occurs at a higher ρ value as compared to the implementations which use data encoding. This is because, for a given ρ , when a data encoding technique is used, other than the normal traffic injected into the network, there is also a traffic component related to the control information (in our case $\ln v$ information) which increases the congestion level in the network.

D. Multimedia SoC Case Study

In this section, we analyze the efficacy of the proposed data encoding schemes on two complex heterogeneous systems. The first one, which is mapped to an 8×8 mesh, consisted of a triple video object plane decoder which has 38 cores (D 38 tvopd) and multimedia and wireless communication which has 26 cores (D 26 media). We assumed a minimum of two-flit and maximum eight-flit packets, deterministic XY routing, and input FIFO buffers of four flits. The time distribution of the traffic followed Poisson's distribution while random data sets were used as workloads. This lowers the effectiveness of the proposed data encoding techniques.

V. CONCLUSION

In this paper, we have presented a set of new data encoding schemes aimed at reducing the power dissipated by the links of a NoC. In fact, links are responsible for a significant fraction of the overall power dissipated by the communication system. In addition, their contribution is expected to increase in future technology nodes. As compared to the previous encoding schemes proposed in the literature, the

rationale behind the proposed schemes is to minimize not only the switching activity, but also (and in particular) the coupling switching activity which is mainly responsible for link power dissipation in the deep submicrometer technology regime.

The proposed encoding schemes are agnostic with respect to the underlying NoC architecture in the sense that their application does not require any modification neither in the routers nor in the links. An extensive evaluation has been carried out to assess the impact of the encoder and decoder logic in the NI. The encoders implementing the proposed schemes have been assessed in terms of power dissipation and silicon area. The impacts on the performance, power, and energy metrics have been studied using a cycle- and bit accurate NoC simulator under both synthetic and real traffic scenarios. Overall, the application of the proposed encodingschemes allows reduced no. of transitions from scheme-I to scheme-II & Scheme-II to Scheme-III which reduces power consumption as well as area also.

REFERENCES

- [1] *International Technology Roadmap for Semiconductors*. (2011) [Online]. Available: <http://www.itrs.net>
- [2] M. S. Rahaman and M. H. Chowdhury, "Crosstalk avoidance and error correction coding for coupled RLC interconnects," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2009, pp. 141–144.
- [3] W. Wolf, A. A. Jerraya, and G. Martin, "Multiprocessor system-on-chip MPSoC technology," *IEEE Trans. Comput.-Aided Design Integr. CircuitsSyst.*, vol. 27, no. 10, pp. 1701–1713, Oct. 2008.
- [4] L. Benini and G. De Micheli, "Networks on chips: A new SoC paradigm," *Computer*, vol. 35, no. 1, pp. 70–78, Jan. 2002.
- [5] S. E. Lee and N. Bagherzadeh, "A variable frequency link for a power aware network-on-chip (NoC)," *Integr. VLSI J.*, vol. 42, no. 4, pp. 479–485, Sep. 2009.