LongLiveNoC: Wear Levelling, Write Reduction and Selective VC allocation for Long lasting Dark Silicon aware NoC Interconnects

Khushboo Rani
Thesis advisor: Prof. Hemangee K. Kapoor
Department of CSE, IIT Guwahati

Introduction

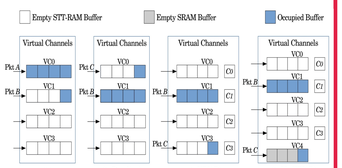
- With many cores on the chip, the communication across them are facilitated by Network-on-Chip (NoC).
- In the era of dark silicon, different components of the chip are powered off, whereas; the communication network is expected to be available.
- Among the different NoC component, the SRAM buffers are the most power hungry.
- In order to reduce the standby power of the NoC in such events, one looks for avenues in non-volatile memory (NVM) technologies.
- The benefits of NVMs are **low static power** and **high density**.
- However, NVM buffer suffer from **costly write operations** and **weak endurance**.
- Due to weak endurance, the lifetime of NVM buffers are limited.
- Further , write variation due to buffer allocation and working set size restrict the lifetime of NVM buffers.
- In the thesis, we proposed **wear-levelling** and **write reduction** techniques to enhance the **lifetime** and reduce the **effect of costly write operation** of NVM buffers in the dark silicon scenario.

Wear-Levelling Techniques

- To quantify the write variation we define terms:
  - Intra-VNet Write Variation: Write variation inside the VNet
  - Inter-VNet Write Variation: Write Variation across the VNet
- Static Buffer Assignment to VCs
  - Intra-VNet Wear Levelling: **WVAR, SET-RR**, and **SET-VNet-WC**
- Dynamic Buffer assignment to VCs
  - Inter- and Intra- VNet Wear Levelling: **Dy-WVAR**
- Processing Element status-based approaches:
  - The power status of associated PE is taken into considerations
  - **gBUF-DN, SRAM-FRQSCL, Hy-SEL-ON, Hy-NEIG-SEL-ON** and **Hy-CRIT-WRD**

WVAR:

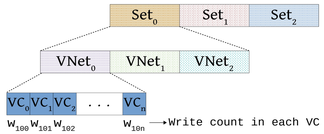
- There is a write counter associated with each VC
- This counter is used to track the number of writes entertained by VC.
- Based on the write counter value, the flit is allocated to the appropriate VC (least write count).



Working Example of WVAR and Hy-WVAR

SET-RR:

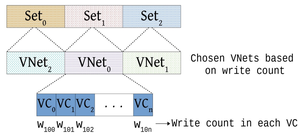
In **SET-RR**, among the different sets, only one set is used at a time for flit allocation in a round robin fashion.



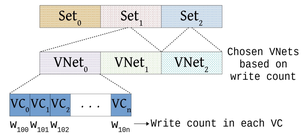
SET-RR: Distribution of iso-area resources into three sets

SET-VNet-WC:

- **SET-VNET-WC** is the combination of **WVAR** and **SET-RR**.
- In particular, the VNet sets are used in a round robin fashion and inside each set **WVAR** is applied.



SET-VNET-WC: Selection of VNet at **K**th interval

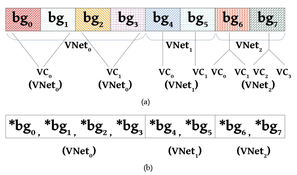


SET-VNET-WC: Selection of VNet at **(K+1)**th interval

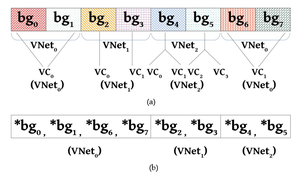
Policy	LI (times)	IntraV Red.(%)	Lat. Inc. (%)	Energy Sav. (%)
SET-RR	3.0	55	16.5	1
SET-VNet-WC	10.7	99.9	16.5	1
SET-VNET-WC-Hy-WVAR	366	97.5	5	-26

Dy-WVAR:

- Proposed technique partition the set of router buffer into buffer groups.
- With each buffer group, a write counter is associated.
- Based on the write counter value, different groups are assigned to different VNet for certain pre-defined interval (*I*).
- On each interval, the buffer group assignment to the different VNet is changed dynamically.
- Within the interval, the **WVAR** is applied inside the VNET.



Dy-WVAR buffer allocation at timestamp t_1



Dy-WVAR buffer allocation at timestamp t_2

- All the proposed approaches uses NVM buffers that increases the network latency.
- To mitigate, the hybrid variant: **Hy-WVAR**, **Hy-Dy-WVAR** and **SET-VNET-WC-Hy-WVAR** have been proposed.

Policy	LI (times)	IntraV / InterV Red. (%)	Lat. Inc.(%)	Energy Sav. (%)
WVAR	3.24	99.97 / -	23	56
Hy-WVAR	19.3	99.99 / -	16	46.2
Dy-WVAR	19.9	99.99 / 99	27	55.8
Hy-Dy-WVAR	55.5	99.99 / 99	22	45.65

gBUF-DN, SRAM-FRQSCL:

- These uses SRAM based buffers, and power gates VCs or reduces router frequency to save energy.

Hy-SEL-ON, Hy-NEIG-SEL-ON, Hy-CRIT-WRD:

- These policies use SRAM and STT-RAM-based VC buffers. The policy **Hy-SEL-ON** power gates SRAM and activates only STT VCs for routers associated PE is in the dark state
- **Hy-NEIG-SEL-ON** extends the idea to the neighbouring router's ports as well.
- In addition to STT VC activation, the policy **Hy-CRIT-WRD** allows critical words to go through SRAM VCs irrespective of associated PE status.

Policy	LI (times)	IntraV Red. (%)	Lat. Inc.(%)	Energy Sav. (%)
gBUF-DN	-	-	19	50
SRAM-FRQSCL	-	-	15	1
Hy-SEL-ON	6.8	98.6	10	52.4
Hy-NEIG-SEL-ON	5.7	98.3	11	61
Hy-CRIT-WRD	9.8	99.7	7	60

Write-ReductionTechniques

- The techniques reduce the number of flits in the network which improves the lifetime and reduces the effect of costly write operation of NVM

ZENCO:

- The policy encodes all continuous and non-continuous zero bytes present in the data packet at the network interface.

DidaSel:

- The proposal uses hybrid buffers to store incoming packets at the router.
- If the number of dirty flits in a packet is below the threshold value, then the packet is send using STT-RAM VC otherwise using SRAM VC.

Policy	LI (times)	Flit Red. (%)	Lat. Inc.(%)	Energy Sav. (%)
ZENCO	1.9	48	4	66
ZENCO-RR	5.4	47	4.5	66
DidaSel	5.25	27	-0.82	29.4

Publications

Journal:

(1) **K.Rani**, and H.K.Kapoor, "InvestigatingFrequencyScaling, Non-Volatile, and Hybrid Memory Technologies for On-Chip Routers to Support the Era of Dark Silicon," **In *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, doi: 10.1109/TCAD.2020.3007555.**

(2) **K. Rani**, and H. K. Kapoor, "Write Variation Aware Buffer Assignment for Improved Lifetime of Non-Volatile Buffers in On-Chip Interconnects," **In *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 27, no. 9 (2019): 2191-2204.**

(3) **K. Rani**, and H. K. Kapoor, "Write-Variation Aware Alternatives to Replace SRAM Buffers with Non-Volatile Buffers in On-Chip Interconnects," **In *IET Computers and Digital Techniques* 13, no. 6 (2019): 481-492.**

Conference:

(1) **K. Rani**, S. Agarwal, and H. K. Kapoor, "DidaSel: Dirty data based Selection of VC for effective utilization of NVM Buffers in On-Chip Interconnects." **In *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design*, pp. 151-156. 2020.**

(2) **K. Rani,** and H. K. Kapoor, "ZENCO: Zero-bytes based ENCOding for Non-Volatile Buffers in On-Chip Interconnects." **In *2020 57th ACM/IEEE Design Automation Conference (DAC)*, pp. 1-6. IEEE, 2020.**

(3) **K. Rani,** and H. K. Kapoor, "Write Variation Aware Non- Volatile Buffers for On-Chip Interconnects." **In *2019 32nd International Conference on VLSI Design and 18th International Conference on Embedded Systems (VLSID)*, pp. 7-12. IEEE, 2019. (Best Paper Candidate)**

(4) **K. Rani**, S. Agarwal, and H. K. Kapoor, "Non-blocking Gated Buffers for Energy Efficient on-chip Interconnects in the era of Dark Silicon." **In *2018 8th International Symposium on Embedded Computing and System Design (ISED)*, pp. 74-79. IEEE, 2018. (Best Paper Award)**