

MemSim Configurations of Specific Interest

G. Hazari

May 2, 2009

This document is meant to accompany [1] which contains sensitivity analysis in VLSI systems. The Augmint-MemSim simulator (which is a cycle accurate model for parallel applications operating on a multi-processor system with a distributed and shared non-uniform memory sub-system) is used as a basis for the analysis. A wide range of randomly generated memory configurations are studied using four multi-threaded applications (LU decomposition, matrix multiplication, bubble sort and matrix inversion). The sensitivities are measured by perturbing the throughput and latency of components in the memory sub-system. Sensitivity coefficients for various design parameters are then defined as $\beta = \frac{\partial P / \partial z}{P/z}$ where P is system performance and z is the particular parameter. Three sensitivity coefficients are developed and incorporated into the Augmint-MemSim simulator. These coefficients cover the following system parameters: (i) throughput of the memories, (ii) latency of access paths, and (iii) throughput of the queues at the end of interconnect wires between the input ports and memories. The coefficients are expected to lie between 0 – 1 for most configurations. However, values outside this range can be expected in certain situations where the perturbation leads to a subsequent change in the access pattern entering the memory sub-system.

In [1] two techniques are proposed to perturb memory throughput and path latency: (i) the memories are modeled as two-stage pipelines, and (ii) an extra delay is added at the output ports. The first technique fails when the memory output queues are small. Select configurations for the LU decomposition application are demonstrated in Table 1. The configurations are given in Figure 1 – Figure 4.

Config.	Two-stage Models		Models with Delay at Output Port	
	Memory Throughput Coeff.	Latency Coeff.	Memory Throughput Coeff.	Latency Coeff.
FC-1	0.20	-2.70	0.01	0.06
FC-2	0.21	-2.13	-0.04	0.09
FC-3	-0.12	4.55	0.00	0.49
FC-4	-0.24	4.73	0.01	0.24

Table 1: Examples of Faulty Measurements On the Two-stage Models [1]

NUM_INPORTS=2;	NUM_INPORTS=4;
NUM_MEMBANKS=2;	NUM_MEMBANKS=2;
NUM_OUTPORTS=8;	NUM_OUTPORTS=2;
INARBITERS_DELAY=2,10;	INARBITERS_DELAY=7,4,3,3;
INQUEUES_SIZE=5,1;	INQUEUES_SIZE=7,2,9,1;
INSCHEDULERS_DELAY=8,4;	INSCHEDULERS_DELAY=11,11,4,4;
NUM_INWIRES_STAGES=32,12,6,48;	NUM_INWIRES_STAGES=20,62,9,80,
MEMINQUEUES_SIZE=1,6,2,3;	25,26,31,5;
MEMINARBITERS_DELAY=2,17;	MEMINQUEUES_SIZE=1,3,1,1,2,3,
MEMBANKS_DELAY=7,6;	8,3;
ROWHIT_DELAY=7,6;	MEMINARBITERS_DELAY=7,6;
MEMOUTQUEUES_SIZE=2,4;	MEMBANKS_DELAY=37,10;
OUTSCHEDULERS_DELAY=16,4;	ROWHIT_DELAY=28,6;
NUM_OUTWIRES_STAGES=1,2,3,1,8,3,	MEMOUTQUEUES_SIZE=2,1;
2,4,3,2,4,2,1,2,8,2;	OUTSCHEDULERS_DELAY=15,2;
OUTQUEUES_SIZE=8,1,1,4,2,4,9,1,	NUM_OUTWIRES_STAGES=4,8,7,8;
7,1,1,5,5,4,5,4;	OUTQUEUES_SIZE=9,6,1,2;
OUTARBITERS_DELAY=5,2,3,11,7,4,	OUTARBITERS_DELAY=18,3;
14,6;	
Memory allocation: 0,1,1,0,0,1,	Memory allocation: 0,0,0,0,1,0,
1,0,0,1,0,1,1,1,0,1,0,1,1,1,	1,1,1,0,1,1,1,0,1,0,1,0,1,0,
1,0,1,0,0,1,0,0,0,0,0,0	1,1,1,1,1,1,1,1,1,1,1,1

Figure 1: Configuration with Faulty Sensitivity Measurement: FC-1

Figure 2: Configuration with Faulty Sensitivity Measurement: FC-2

NUM_INPORTS=8;	NUM_INPORTS=8;
NUM_MEMBANKS=2;	NUM_MEMBANKS=2;
NUM_OUTPORTS=4;	NUM_OUTPORTS=1;
INARBITERS_DELAY=19,2,2,2,2,3,9,18;	INARBITERS_DELAY=3,3,13,4,3,7,18,7;
INQUEUES_SIZE=5,6,5,5,2,1,7,6;	INQUEUES_SIZE=2,6,5,2,5,3,7,1;
INSCHEDULERS_DELAY=16,11,9,11,10,3,2,5;	INSCHEDULERS_DELAY=4,14,5,3,4,19,2,2;
NUM_INWIRES_STAGES=53,90,25,75,14,10,19,23,34,20,44,43,69,95,32,92;	NUM_INWIRES_STAGES=22,11,5,49,13,24,35,48,75,64,31,45,5,44,7,12;
MEMINQUEUES_SIZE=3,8,2,2,2,6,2,1,1,3,1,1,2,4,3,1;	MEMINQUEUES_SIZE=6,2,3,1,1,7,1,2,3,1,1,8,3,1,1,1;
MEMINARBITERS_DELAY=3,2;	MEMINARBITERS_DELAY=2,4;
MEMBANKS_DELAY=9,6;	MEMBANKS_DELAY=5,19;
ROWHIT_DELAY=9,6;	ROWHIT_DELAY=5,8;
MEMOUTQUEUES_SIZE=3,1;	MEMOUTQUEUES_SIZE=1,9;
OUTSCHEDULERS_DELAY=15,3;	OUTSCHEDULERS_DELAY=13,11;
NUM_OUTWIRES_STAGES=7,5,6,3,2,4,2,1;	NUM_OUTWIRES_STAGES=2,1;
OUTQUEUES_SIZE=8,5,1,6,1,4,8,8;	OUTQUEUES_SIZE=1,1;
OUTARBITERS_DELAY=16,2,12,3;	OUTARBITERS_DELAY=11;
Memory allocation: 0,1,1,1,1,1,0,1,0,1,1,1,0,1,0,1,0,0,0,0,1,0,0,0,1	Memory allocation: 1,1,1,1,1,0,1,0,0,1,1,1,0,1,1,0,0,0,0,0,0,0,1,0,1,0,1,0,1,1,0,0,1,0,1

Figure 3: Configuration with Faulty Sensitivity Measurement: FC-3

Figure 4: Configuration with Faulty Sensitivity Measurement: FC-4

Config. Id. (Application)	Hypo-sensitive Coefficient	Other Coefficients	$\partial P/P$	$\partial P/P$ (reduced model)
HC-1 (Inv)	PL: -1.47	MT:0.32, WQ:0.92	-1.2%	WQM: -1.2%
HC-1 (Mul)	PL: -1.05	MT:0.21, WQ:0.56	-0.9%	WQM: -0.8%
HC-2 (Inv)	PL: -0.30	MT:0.91, WQ:-0.25	-0.1%	OLM: -0.1%
HC-2 (Inv)	WQ: -0.25	MT:0.91, PL: -0.30	-0.2%	OLM: -0.2%
HC-3 (Mul)	MT: -0.60	PL: -0.02, WQ:0.15	+1.7%	WQM: +0.7%
HC-3 (LU)	MT: -0.26	PL:0.35, WQ:0.11	+0.7%	PLM: +0.02%
HC-4 (BS)	MT: -0.21	PL:0.36, WQ:1.26	+0.5%	WQM: +0.5%
HC-5 (BS)	MT: -0.20	PL:0.07, WQ:0.22	+0.5%	WQM: +0.7%
HC-6 (Mul)	PL: -0.26	MT:0.04, WQ: -0.12	-0.2%	WQM: -1.6%

Table 2: Hypo-sensitive Configurations [1]

When a particular coefficient is close to 1, the corresponding parameter is the system bottleneck. Thus the sensitivity coefficients classify the design space into three distinct regions: (i) the memory throughput limited region, (ii) the latency limited region, and (iii) the wire-queue limited region. There are also large sections of the design space which do not lie in any of these three bottleneck regions. The following trace-driven reduced performance models (which are far simpler than the Augmint-MemSim model) are established (in [1]) for the respective bottleneck regions: (i) an *open loop memory model*, (ii) a *path latency model*, and (iii) a *wire-queue model*.

Besides evaluating performance in the bottleneck regions, the reduced models are useful in understanding performance trends when the sensitivity coefficients lie outside the 0 - 1 range [1]. A substantial number of hypo-sensitive configurations (that is, having negative coefficient values) are observed and the extreme values are substantially less than 0. Such situation arises when the perturbation causes a change in access pattern which have an adverse effect on the impact of a more critical parameter. This effect is quantified by running the traces from the original and perturbed configurations on the reduced models, as demonstrated in Table 2¹. The configurations shown in this table are given in Figure 5 – Figure 10.

¹The applications are labeled as *LU* for LU decomposition, *Mul* for matrix multiplication, *BS* for bubble sort and *Inv* for matrix inversion in the following tables. The sensitivity coefficients are labeled as *MT* for memory throughput, *PL* for path latency and *WQ* for the wire-queue coefficient. The reduced models are labeled as *OLM* for the open loop memory model, *PLM* for the path latency model and *WQM* for the wire-queue model.

NUM_INPORTS=2;	NUM_INPORTS=4;
NUM_MEMBANKS=2;	NUM_MEMBANKS=8;
NUM_OUTPORTS=8;	NUM_OUTPORTS=4;
INARBITERS_DELAY=7,2;	INARBITERS_DELAY=4,4,4,11;
INQUEUES_SIZE=1,1;	INQUEUES_SIZE=2,4,1,1;
INSCHEDULERS_DELAY=10,2;	INSCHEDULERS_DELAY=9,11,4,4;
NUM_INWIRES_STAGES=16,37,57,51;	NUM_INWIRES_STAGES=56,85,9,5,9, 73,13,39,88,9,52,5,46,41,23, 15,97,88,5,8,99,34,76,17,51, 93,87,46,11,9,33,6;
MEMINQUEUES_SIZE=6,3,1,2;	MEMINQUEUES_SIZE=1,3,2,5,1,5,1, 1,3,4,2,3,5,3,4,5,8,1,3,8,1, 4,6,4,2,4,8,2,4,1,6,1;
MEMINARBITERS_DELAY=9,2;	MEMINARBITERS_DELAY=17,5,17,19, 16,17,14,5;
MEMBANKS_DELAY=17,7;	MEMBANKS_DELAY=10,8,49,7,12,40, 38,7;
ROWHIT_DELAY=8,7;	ROWHIT_DELAY=9,6,8,7,10,4,3,7;
MEMOUTQUEUES_SIZE=1,1;	MEMOUTQUEUES_SIZE=8,2,2,6,1,7, 8,5;
OUTSCHEDULERS_DELAY=10,7;	OUTSCHEDULERS_DELAY=2,2,2,2,8, 12,7,12;
NUM_OUTWIRES_STAGES=1,4,6,1,1,9, 2,2,5,4,3,1,7,4,4,9;	NUM_OUTWIRES_STAGES=6,8,1,2,2,2, 3,3,4,3,2,2,2,1,1,1,1,4,7,9, 1,9,7,3,2,4,5,3,2,1,1,1;
OUTQUEUES_SIZE=1,2,1,3,7,1,2,5, 2,5,2,2,1,7,2,2;	OUTQUEUES_SIZE=1,7,1,1,7,9,3,2, 2,5,1,1,1,1,3,4,1,5,4,2,5,2, 9,6,1,1,5,4,1,1,1,1;
OUTARBITERS_DELAY=8,16,4,3,3,8, 2,2;	OUTARBITERS_DELAY=18,4,19,18;
Memory allocation: 1,1,1,0,1,1, 0,1,0,1,0,0,1,1,0,1,0,0,0,1, 0,0,1,1,1,1,0,0,0,1,0,0	Memory allocation: 6,2,1,2,3,2, 5,5,5,5,2,7,6,7,2,5,4,2,1,2, 3,1,3,5,0,3,2,2,2,4,2,2

Figure 5: Hypo-sensitive Configuration: HC-1

Figure 6: Hypo-sensitive Configuration: HC-2

NUM_INPORTS=4;	NUM_INPORTS=1;
NUM_MEMBANKS=8;	NUM_MEMBANKS=8;
NUM_OUTPORTS=4;	NUM_OUTPORTS=2;
INARBITERS_DELAY=8,13,2,10;	INARBITERS_DELAY=2;
INQUEUES_SIZE=5,3,4,1;	INQUEUES_SIZE=1;
INSCHEDULERS_DELAY=6,11,6,6;	
NUM_INWIRES_STAGES=39,32,33,87,	INSCHEDULERS_DELAY=2;
7,96,7,55,31,50,43,6,12,5,	NUM_INWIRES_STAGES=34,15,40,27,
13,22,24,32,13,7,68,22,30,	61,18,57,44;
12,20,26,73,7,92,20,58,78;	
MEMINQUEUES_SIZE=2,3,1,1,6,1,1,	MEMINQUEUES_SIZE=1,1,8,3,1,1,
2,1,2,8,2,1,2,2,1,4,6,4,1,1,	2,2;
1,4,2,1,6,7,7,1,9,9,3;	MEMINARBITERS_DELAY=12,9,4,19,
MEMINARBITERS_DELAY=14,3,11,16,	17,12,2,11;
6,3,2,8;	
MEMBANKS_DELAY=37,12,17,12,5,33,	MEMBANKS_DELAY=6,36,48,29,6,8,
36,37;	36,9;
ROWHIT_DELAY=10,12,6,7,5,16,	ROWHIT_DELAY=6,21,25,14,6,4,
5,14;	17,9;
MEMOUTQUEUES_SIZE=1,4,1,1,1,8,	
5,2;	MEMOUTQUEUES_SIZE=8,1,6,1,3,4,
OUTSCHEDULERS_DELAY=2,7,2,7,9,5,	8,1;
5,2;	OUTSCHEDULERS_DELAY=4,5,12,9,3,
NUM_OUTWIRES_STAGES=8,4,1,6,9,4,	2,9,16;
4,1,2,2,5,2,8,7,9,8,3,2,9,1,	
4,1,2,2,4,4,6,2,6,1,6,5;	NUM_OUTWIRES_STAGES=4,1,3,9,2,8,
OUTQUEUES_SIZE=3,1,4,9,1,7,1,3,	3,1,1,1,1,3,1,2,5,2;
2,2,2,4,1,4,1,1,1,3,2,2,4,1,	OUTQUEUES_SIZE=1,1,4,8,6,6,4,5,
5,5,2,2,2,5,2,6,7,2;	2,9,1,9,9,1,5,2;
OUTARBITERS_DELAY=3,11,3,7;	OUTARBITERS_DELAY=9,6;
Memory allocation: 3,6,6,5,0,0,	Memory allocation: 1,3,6,5,0,7,
6,2,7,5,3,5,5,0,5,7,5,5,0,5,	7,1,5,5,0,2,3,1,0,4,1,7,4,5,
4,0,7,7,6,2,4,7,4,1,5,0	3,7,6,3,1,7,6,7,5,4,4,4

Figure 7: Hypo-sensitive Configuration: HC-3

Figure 8: Hypo-sensitive Configuration: HC-4

```

NUM_INPORTS=1;
NUM_MEMBANKS=8;
NUM_OUTPORTS=2;

INARBITERS_DELAY=3;
INQUEUES_SIZE=1;
INSCHEDULERS_DELAY=19;
NUM_INWIRES_STAGES=29,35,5,29,5,
    57,49,50;
MEMINQUEUES_SIZE=5,1,1,7,4,1,
    2,8;
MEMINARBITERS_DELAY=2,16,7,3,9,
    5,11,17;
MEMBANKS_DELAY=5,41,28,9,23,20,
    7,8;
ROWHIT_DELAY=5,29,3,6,4,4,7,8;
MEMOUTQUEUES_SIZE=1,2,5,7,4,2,
    7,2;
OUTSCHEDULERS_DELAY=3,7,2,10,5,
    2,12,5;
NUM_OUTWIRES_STAGES=1,2,3,5,2,2,
    8,6,1,1,3,2,1,3,4,1;
OUTQUEUES_SIZE=1,1,9,8,2,5,5,7,
    1,3,1,9,7,1,3,3;
OUTARBITERS_DELAY=8,2;

Memory allocation: 3,5,7,5,1,6,
    5,4,3,1,0,4,6,4,4,6,2,5,4,3,
    5,7,0,3,3,0,5,1,5,7,7,4

```

Figure 9: Hypo-sensitive Configuration: HC-5

```

NUM_INPORTS=1;
NUM_MEMBANKS=8;
NUM_OUTPORTS=1;

INARBITERS_DELAY=12;
INQUEUES_SIZE=5;
INSCHEDULERS_DELAY=15;
NUM_INWIRES_STAGES=45,6,24,43,
    10,92,52,9;
MEMINQUEUES_SIZE=6,5,4,1,1,7,
    2,3;
MEMINARBITERS_DELAY=3,2,4,9,2,6,
    6,2;
MEMBANKS_DELAY=47,12,5,21,15,18,
    25,7;
ROWHIT_DELAY=13,3,5,5,4,6,8,7;
MEMOUTQUEUES_SIZE=8,3,6,2,4,4,
    5,6;
OUTSCHEDULERS_DELAY=9,7,3,6,4,7,
    4,2;
NUM_OUTWIRES_STAGES=4,4,1,5,2,3,
    9,5;
OUTQUEUES_SIZE=7,9,5,4,4,1,1,1;
OUTARBITERS_DELAY=3;

Memory allocation: 1,5,6,2,5,1,
    6,7,6,2,7,0,7,7,2,3,1,4,6,6,
    5,4,1,4,5,6,6,4,7,6,5,3

```

Figure 10: Hypo-sensitive Configuration: HC-6

Configuration Id. (Application)	Hyper-sensitive Coefficient	$\partial P/P$	$\partial z/z$	$\partial P/P$ (reduced model)
HC-7 (BS)	MT:1.12	-2.3%	-2.1%	OLM:-0.3%
HC-8 (Mul)	PL:1.16	+0.9%	+0.8%	PLM:+0.1%
HC-4 (BS)	WQ:1.26	+1.6%	+1.2%	WQM:+0.4%

Table 3: Hyper-sensitive Configurations

At the other extreme are hyper-sensitive configurations in which the coefficient values are greater than 1. These are fewer in number than the hypo-sensitive ones. The extreme values are also much closer to the 0 – 1 range. Hyper-sensitive situations arise when a parameter is already critical and the effect of the change in access pattern adds to the primary impact of the perturbation. The reduced models demonstrate this effect as shown in Table 3. The configurations shown in this table are given in Figure 11 and Figure 12. (One of the hyper-sensitive configurations is also hypo-sensitive and has been presented earlier.)

NUM_INPORTS=8;	NUM_INPORTS=4;
NUM_MEMBANKS=2;	NUM_MEMBANKS=8;
NUM_OUTPORTS=8;	NUM_OUTPORTS=4;
INARBITERS_DELAY=3,3,3,5,5,4, 9,15;	INARBITERS_DELAY=3,19,3,9;
INQUEUES_SIZE=7,5,2,1,1,8,9,1;	INQUEUES_SIZE=4,12,7,10;
INSCHEDULERS_DELAY=2,2,4,10,6,6, 10,5;	INSCHEDULERS_DELAY=9,10,2,13;
NUM_INWIRES_STAGES=47,7,8,51,15, 9,50,7,6,26,10,23,74,48, 32,83;	NUM_INWIRES_STAGES=13,10,10,16, 15,3,6,3,11,13,20,3,24,3,20, 9,20,21,15,7,9,3,8,12,7,25, 4,23,5,10,24,26;
MEMINQUEUES_SIZE=4,2,7,9,3,6,1, 5,2,1,4,3,2,2,3,1;	MEMINQUEUES_SIZE=4,17,5,12,4,11, 10,19,12,19,12,7,8,7,4,4,12, 4,11,12,7,13,6,7,8,4,6,5,9, 16,16,4;
MEMINARBITERS_DELAY=2,18;	MEMINARBITERS_DELAY=3,3,8,2,3,6, 15,2;
MEMBANKS_DELAY=47,44;	MEMBANKS_DELAY=26,9,20,27,6,15, 29,15;
ROWHIT_DELAY=12,13;	ROWHIT_DELAY=6,9,7,8,6,3,5,3;
MEMOUTQUEUES_SIZE=1,1;	MEMOUTQUEUES_SIZE=13,5,14,9,4,4, 16,8;
OUTSCHEDULERS_DELAY=6,14;	OUTSCHEDULERS_DELAY=12,5,4,11,2, 4,5,2;
NUM_OUTWIRES_STAGES=4,5,5,5,8,6, 6,1,2,1,1,1,5,3,1,6;	NUM_OUTWIRES_STAGES=3,14,4,27,6, 14,20,16,13,5,26,9,15,12,27, 18,3,23,6,3,4,29,11,5,3,3, 26,3,7,4,7,17;
OUTQUEUES_SIZE=1,9,2,1,3,5,3,1, 1,6,3,1,7,2,7,4;	OUTQUEUES_SIZE=9,4,8,9,11,4,9,8, 4,8,9,4,14,7,5,14,7,15,18,9, 9,11,7,9,8,10,10,12,8,4,5,13;
OUTARBITERS_DELAY=8,5,4,18,10, 13,3,12;	OUTARBITERS_DELAY=14,10,8,7;
Memory allocation: 0,1,0,0,0,1, 0,0,0,1,0,1,1,0,0,0,0,1,0,0, 0,1,1,1,0,0,0,1,1,0,0,1	Memory allocation: 6,7,1,1,3,0, 6,0,1,0,6,7,5,5,7,3,6,6,3,2, 3,1,5,4,5,3,5,3,4,3,6,7

Figure 11: Hyper-sensitive Configuration: HC-7

Figure 12: Hyper-sensitive Configuration: HC-8

In this document we have given the configuration details for interesting situations encountered during the empirical studies on randomized memory configurations in [1].

References

- [1] G. Hazari (under the guidance of Prof. M. P. Desai and Prof. H. Narayanan), “A Spectrum of Performance Models for VLSI Memory Sub-systems”, *PhD Thesis, Department of Electrical Engineering, I.I.T. Bombay*, August 2009 (expected submission date).