N^{3}ASIC-BASED NANOWIRE VOLATILE RAM

Mostafizur Rahman*, Pritish Narayanan and Csaba Andras Moritz

***Abstract*—As CMOS technology advances into the nanoscale, the continuous push for low power, high performance, and dense volatile memory is reaching its limit. Moreover, in the nanometer regime complex design rules and manufacturing costs are escalating as it is getting increasingly difficult to control manufacturing process parameters. In this paper, we propose a novel 10 transistor based volatile Nanowire Random Access Memory (10T-NWRAM) which is highly scalable and manufacturing friendly since it is based on the very regular N^{3}ASIC fabric. Besides, it has the potential to be significantly faster and low leakage alternative to SRAM since high performance nanowire FETs and dynamic logic is used for memory architecture.**

I. INTRODUCTION

CMOS technology scaling has enabled on-chip caches to be more densely packed in smaller areas and to operate at higher speeds while consuming lower power. However, as feature sizes are getting smaller, secondary effects such as gate tunneling currents, sub-threshold conduction, and band to band leakage current are becoming more prominent. As a result, passive power dissipation in on-chip caches has become a major source of concern.

Numerous optimizations of conventional 6T SRAM cell have been proposed over the years to minimize leakage current. QPG-SRAM [1], the 9-transistor SRAM cell [2], IWL-VC SRAM and PP-SRAM [3] to name a few. At the same time, irregular 2-dimensional layout of SRAM is proving to be more and more difficult to manufacture as the lithography process has to deal with line end shortening, corner rounding, various hot-spots, and find the best compromise in illumination and dose settings to resolve all structures and pitches simultaneously [4][8][22]. These complex challenges have ultimately led to escalating numbers of design rules, complex optical proximity corrections, and iterations of yield simulation using design rule check tools. Moreover, according to lithography simulations proximity effects of complex 2-dimensional designs can cause 1-15% critical dimension variation [4], thus resulting in higher leakage current. So is there any alternative solution to SRAM for manufacturing friendly, low leakage power, high performance and dense volatile memory?

We acknowledge support from the Focus Center Research Program (FCRP) - Center on Functional Engineered Nano Architectonics (FENA). This work was also supported by the Center for Hierarchical Manufacturing (CHM) University of Massachusetts Amherst and NSF award number 0541066. M. Rahman, P.Narayanan and C. A. Moritz are with University of Massachusetts, Amherst. (*corresponding author, phone: 413-330-8257; e-mail: rahman@ ecs.umass.edu)

We propose a novel volatile 10 transistor based Nanowire Random Access Memory (10T-NWRAM) which has the potential to address leakage power and manufacturing concerns without sacrificing on density or performance. Key features of 10T-NWRAM are dynamic logic for storage, use of external read buffer for read operation, control mechanism using non overlapping clock phases to alleviate stability concerns and very regular grid-like layout for better manufacturability.

The 10T-NWRAM is implemented on the N^{3}ASICs nanowire-based computing fabric [7]. 10T-NWRAM is highly scalable, more variation tolerant and manufacturing friendly since it follows a very regular and uniform grid-based design.

The rest of the paper is organized as follows: Section II presents the underlying N^{3}ASICs fabric. Section III introduces 10T-NWRAM architecture, functionality and layout, Section IV provides insight into methodology, comparison metrics and area evaluation. In Section V we discuss architectural benefits and manufacturing aspects. Finally in section VI conclusions are drawn.

II. N^{3}ASIC FABRIC

10T-NWRAM is built on Nanoscale 3-dimensional application specific integrated circuits (N^{3}ASIC) fabric [7]. This is a hybrid nanowire-CMOS fabric. All logic/memory implementation is on uniform parallel semiconductor nanowire arrays. Active devices in N^{3}ASICs are single type, doped dual channel crossed nanowire transistors (2C-xnwFETs). Area-distributed interfaces or vias are used to connect outputs of nanowire stages to a standard CMOS metal stack. Reliable CMOS control circuitry is used for dynamic control. Some highlights of N^{3}ASIC are hybrid integration of CMOS and nanowires for logic and memory, combination of manufacturing friendly low cost unconventional techniques with conventional lithographic manufacturing flow for ease of integration and the use of scalable high performance nanowire FETs for fast dynamic circuit style.

Fig.1 shows overview of assembly sequence for N^{3}ASICs. Firstly, nanowires are directly patterned on ultra-thin silicon-on insulator substrates using unconventional techniques such as nano-imprint [5] or SNAP [6] for dense sub-lithographic features. Then the transistors are formed at certain places using lithographic masking and deposition.

Finally, area distributed interfaces (contact pins) are laid down for interconnection between logic stages using 3-D CMOS metal stack.

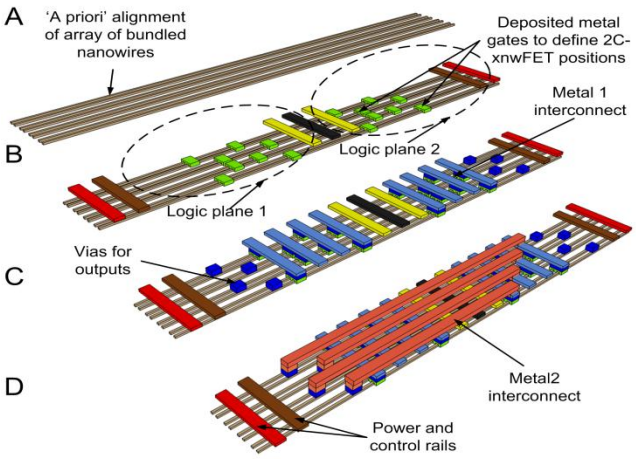

Fig.1. Simplified Assembly sequence of N^3ASIC

2C-xnwFETs (Fig.2) are used for N^3ASIC fabric. The omega gated structure along with dual nanowire channels give 2C-xnwFET more advantage over NMOSFETs as shown in Table I.

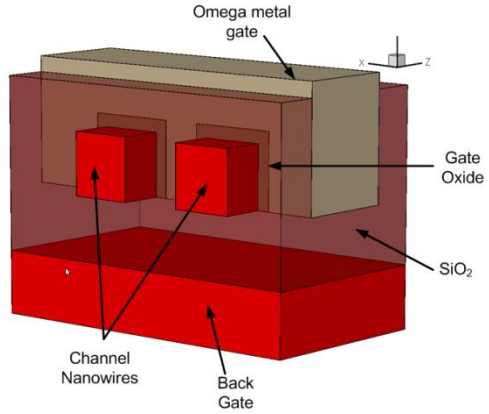

Fig.2. 3D structure of 2C-xnwFET

TABLE I
COMPARISON OF DEVICE MODELS

	2C-xnwFET	PTM_HP(NMOS)	PTM_LP (NMOS)
Ion	4.08^{-05}	3.68^{-05}	1.47^{-05}
Ioff	1.56^{-09}	1.05^{-08}	1.99^{-12}
Vdd(nominal)	0.8	0.7	0.9
Vth	0.27	0.47	0.68
Length/Width	16/16	16/32	16/32

Superior benefits of 2C-xnwFETs in terms of high on current (I_{on}), large I_{on} /I_{off} ratio and lower V_{th} in comparison to PTM [10] high performance and low power device models make N^3ASIC based logics more promising in terms of power and performance compared to designs based in CMOS FETs.

III. 10T-NWRAM ARCHITECTURE

10T-NWRAM utilizes two dynamic NAND gates connected in cross-coupled manner to retain the stored value. This is similar to the conventional SRAM where cross coupled inverters are used for storage. However, in contrast to SRAM where complementary storage values are changed simultaneously in case of write, in 10T-NWRAM a completely different approach is used. During 'Write' operation complementary values change in non-overlapping clock phases thus eliminating similar stability concerns of cross coupled inverted logic for future scaled technology nodes. To further take advantage of non-overlapping clocking scheme, 'Read' operation is done using a 2-input dynamic AND gate where the inputs are stored bit (out) and the read signal ($read_0$).

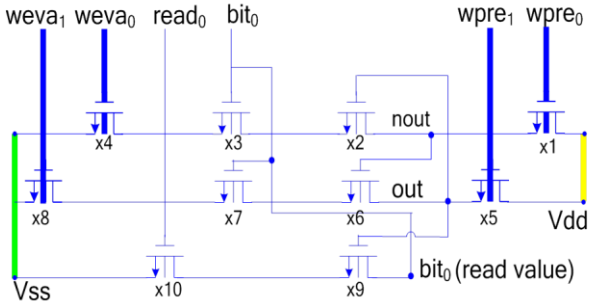

Fig.3. Schematic of 10T-NWRAM

A. Write Operation

To write a 1 in storage node 'out' the input signal 'bit_0' is kept low during precharge and evaluate phase ($wpre_1$,$weva_1$) as depicted in Fig.3. As a result during evaluate phase ($weva_0$) of 'nout', the 'nout' storage node goes to 0, thus storing the complementary value of 'out'. To write a 0 in storage node 'out' the input signal 'bit_0' is kept low during precharge and evaluate phase ($wpre_0$, $weva_0$) allowing the 'nout' storage node to go high which ultimately pulls '1' to '0' during $weva_1$ as depicted in Figure 3. If the 'bit0' signal is kept unchanged during subsequent precharge and evaluate phases ($wpre_1$,$weva_1$,$wpre_0$,$weva_0$) the complementary values ('out', 'nout') keep restoring each other.

This is in direct contrast to SRAM, where in order to change a value on the state node the pull down current discharging internal state node has to overcome the pull up current of that node which is achieved by the sizing of transistors.

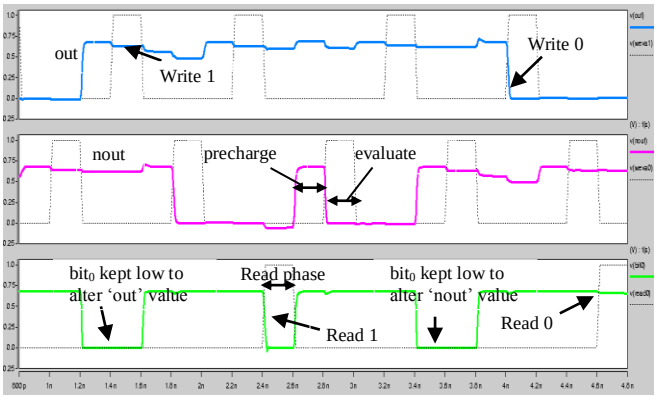

Fig.4. Simulated waveform of 10T-NWRAM operations.

B. Read Operation

During read operation the 'bit_0' signal is initially precharged to 1, consequently when the '$read_0$' signal is turned on the 'bit_0' signal goes to 0 or remains at 1 depending on the value stored in 'out'. Fig. 3 shows both reading 1 and reading 0 mechanisms. This read mechanism is very similar to read buffer concept [25] used in SRAM to maintain adequate read stability.

C. Restore Operation

10T-NWRAM exploits the behavior that during a fixed period of time the activity in a cache is only centered on a small subset of word lines. Therefore all the other inactive input signals are kept low to put the storage nodes on state-preserving mode. However, due to leakage in nanowire transistors the stored charge starts to decay in the long run. In order to restore the charges back to storage conditions the clock signals associated with each bit (e.g $wpre_1$, $weva_1$, $wpre_0$, $weva_0$) are turned on and due to self-restoring mechanism of 10T-NWRAM the original stored values are retained without the need of read-out and write-back mechanism as in DRAM.

Fig. 5 shows detail steps of 10T-NWRAM layout-

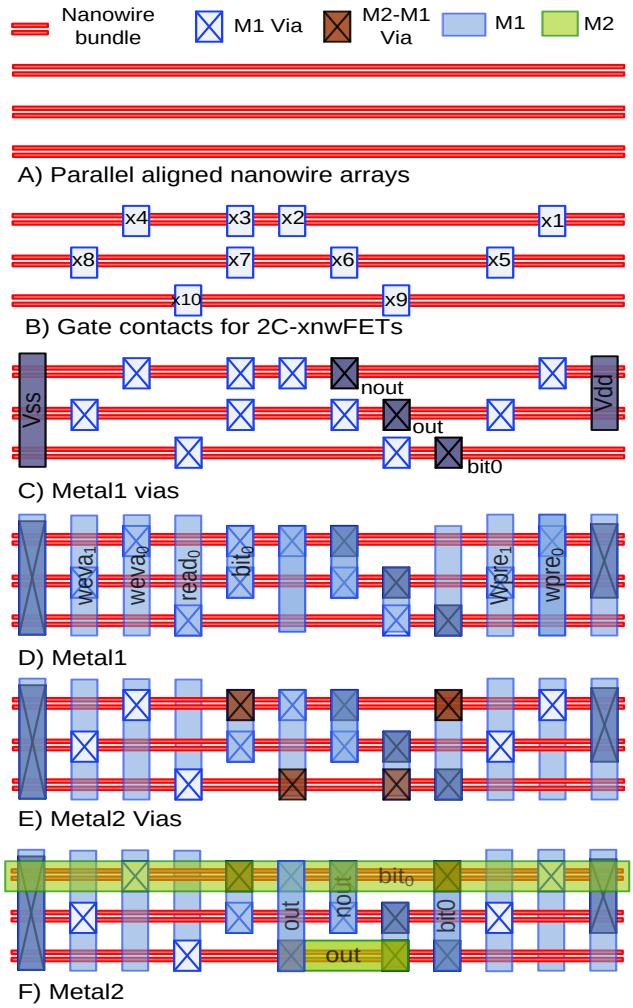

Fig.5. Step by Step Layout of 10T-NWRAM

After patterning nanowires in step A, transistors are formed in step B, then contacts are laid out for power rail and transistor gates in step C. In step D and Step E metal lines are formed and metal2-metal1 vias are placed for subsequent interconnection. Finally in step F metal2 lines are drawn to complete local and global routing for the cell. Overall, 10T-NWRAM requires only two metal layers (metal2 and metal1) to achieve signal routing within the cell and for global signals such as bitlines and wordline clock signals. Whereas in an SRAM cell 4 metal layers are required for local and global routing requirements.

IV. Methodology & Evaluation

10T-NWRAM is based on very regular one dimensional grid. Therefore, design rules [8][9] for 1D gridded design (Table II) is used for area calculation.

TABLE II
1D Gridded Design Rules

Pitch (16nm Tech)	M1,M2 interconnect	Contact
1D-Gridded Design	60 ~ 40 nm	50 nm

For SRAM scaling down to 16nm technology node a wide range of factors were considered rather than individual parameters since scaling rules are specific to process of particular manufacturer. We have collected design rules from experimental data published by the industry [11-21] and examined the scaling factors across technology node. For example, according to [13] and [14] SRAM area scaled by a factor of 2.02 from 45nm to 32nm, so we used the same factor 2.02 to scale the area from 32nm. Table III shows scaled SRAM cell area for different factors.

TABLE III
16 nm Scaled 6T-SRAM Area

Scaling factors	**2.45**	**2.02**	**1.75**	**1.64**
Area in μm^2	0.028	0.042	0.056	0.064
	0.026	0.038	0.051	0.058
	0.025	0.037	0.049	0.056

12 different design rules based on different scaling factors were extracted. Table IV shows the design rules corresponding to the cell area in Table III.

TABLE IV
Design Rules for 6T-SRAM

Scaling factors	**0.76**	**0.7**	**0.72**	**0.76**
M1x half pitch (nm)	32.49	27.5625	29.16	32.49
	28.88	24.5	25.92	28.88
	28.88	24.5	25.92	28.88
N+/P+ spacing (nm)	43.32	36.75	38.88	43.32
	33.7896	28.665	30.3264	33.7896
	37.544	31.85	33.696	37.544
Via spacing (nm)	32.49	27.5625	29.16	32.49
	28.88	24.5	25.92	28.88
	28.88	24.5	25.92	28.88

Based on the area in Table III and corresponding design rules in Table IV we will calculate interconnect dimensions and extract RC data for accurate simulation of 6T-SRAM for power and performance as a part of future work.

Additionally, we will also investigate power and performance results of 8 transistor based gridded SRAM [22] which uses regular 1 dimensional grid and uniformly sized transistors. For area evaluation of 8T-SRAM we used the design rules outlined in Table II.

Detail area comparison of 10T-NWRAM, 8T-SRAM and 6T-SRAM cells is shown here. Fig.6 shows upper bound and lower bound in area for 10T-NWRAM and 8T-SRAM corresponding to the range of values for metal and contact pitch as shown in Table II. Also the scaled SRAM area from Table III is plotted for comparison.

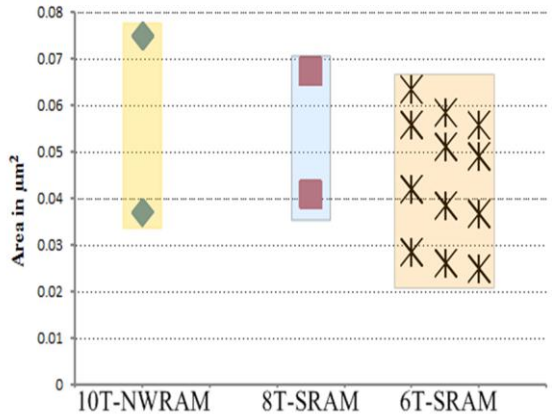

Fig.6. Area comparison

While the upper bound shows slightly larger area for a single cell 10T-NWRAM in comparison to 6T-SRAM, this is a pessimistic assumption for the 10T-NWRAM design. Design rules are expected to be closer to or better than the lower bound with 10T-NWRAM since it is based on highly regular and defect tolerant N^3ASICs fabric that uses only two metal layers for interconnect and global routing, Further, this implies that scaling could be more aggressive since design rules depend on parameters such as transistor dimensions, cell routing, metal layers, overlay, delay, reliability and overall process technology [23].

V. Discussion

From Fig.6 it is clear that the lower bound of 10T-NWRAM is comparable to very aggressively scaled SRAM in terms of density. We anticipate similar or better results for power and performance since 2c-xnwFETs are used in 10T-NWRAM design which has higher I_{on}, I_{off} ratio as shown in Table I. Moreover, we expect significant improvement in terms of leakage power compared to SRAMs since 10T-NWRAM has built-in leakage control mechanism using natural stack of dynamic NAND gates. Furthermore, innovative clocking scheme with non-overlapping clock phases mitigates noise concerns in 10T-NWRAM which is very common in cross coupled logics.

10T-NWRAM is based on N^3ASIC fabric where nanowires are patterned with unconventional techniques such as low cost nano-imprint lithography which provides high throughput, high resolution, high aspect ratio features. The uniform grid structure of the bottom nanowire layer implies that an initial process may be offset with no loss of functionality, thus reducing the registration requirements [24][26]. Less number of process steps, CMOS integration for control signals, uniform devices are the key features of N^3ASIC that gives it the manufacturing benefits. In addition, 10T-NWRAM architecture with same type and uniformly doped transistors, separate read logic allows more manufacturing defect tolerance.

VI. Conclusion

A new 10T-NWRAM design was described and thoroughly analyzed. Based on the layout and methodology presented in section II and IV the area of 10T-NWRAM was calculated to be $0.037\mu m^2$ which is better than 8T-gridded SRAM that uses same design rules and slightly larger than 6T SRAM where more aggressive design rules are used. Moreover, we expect 10T-NWRAM to have substantial advantages over SRAMs in terms of performance and leakage power mainly due to 2C-xnwFETs, dynamic circuit style and drowsy characteristics. Low cost assembly of parallel aligned nanowire arrays, integration of metal layers for interconnect, reliable CMOS clocking scheme and combination of architectural level innovation for highly stable fast read/write operation makes low leakage 10T-NWRAM a promising candidate for next generation volatile memory.

References

[1] P. Nair, S. Eratne, E. John,"A quasi-power-gated low-leakage stable SRAM cell," *in Proceedings of International Midwest Symposium on Circuits and Systems* (*MWSCAS*), 2010 pp.761-764.

[2] Z. Liu and V. Kursun, "High Read Stability and Low Leakage Cache Memory Cell," *in Proceedings of International Symposium on Circuits and Systems* (*ISCAS)* 2007, pp. 2774-2777.

[3] Razavipour, G. et al.; "Design and Analysis of Two Low-Power SRAM Cell Structures," *VLSI Technology Symposium*, vol.17, no.10, Oct. 2009

[4] M. Smayling, "Gridded Design: Rules1-D Approach Enables Scaling of CMOS Logic," *Nanochip Technology Journal*, vol. 6, issue 2, 2008

[5] D. Wang et al, "Development of ultra-high density silicon nanowire arrays for electronics applications," *Nano Research*, vol. 1, no. 1, pp. 9-21, 2008.

[6] J. R. Heath, "Superlattice Nanowire Pattern Transfer (SNAP)," *Accounts of Chemical Research*, vol. 41, no. 12, pp. 1609-1617, Dec. 2008.

[7] P.Panchapakeshan, P.Narayanan, C.A.Moritz, "N^3ASICs: 3-D Nanoscale Application Specific Integrated Circuits," *to be published in Proceedings of International Symposium on Nanoscale Architectures (NANOARCH) 2011*.

[8] C. Bencher, H. Dai, and Y. Chen. "Gridded design rule scaling: taking the CPU toward the 16nm node," *in Proceedings of SPIE* 7274, 2009

[9] L. Burn, "Lithography Candidates for the 16nm, 11nm, and 8nm Logic Nodes," *Semicon* 2010

[10] Predictive Technology Model(PTM) website, Available :- http://www.eas.asu.edu/~ptm/

[11] Bai, P. et al , "A 65nm logic technology featuring 35nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57

μm2 SRAM cell," *International Electron Devices Meeting (IEDM) Technical Digest* , pp. 657- 660, Dec. 2004

[12] Jan, C.-H. , "A 65nm ultra low power logic platform technology using uni-axial strained silicon transistors," *International Electron Devices Meeting (IEDM) Technical Digest*, pp.60-63, Dec. 2005

[13] Mistry, K. et al, "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," *International Electron Devices Meeting (IEDM) Technical Digest*, pp.247-250, Dec. 2007

[14] Natarajan, S., "A 32nm logic technology featuring 2nd-generation high-k + metal-gate transistors, enhanced channel strain and 0.171μm2 SRAM cell size in a 291Mb array," *International Electron Devices Meeting (IEDM) Technical Digest* , , pp.1-3, Dec. 2008

[15] Greene, B.et al, "High performance 32nm SOI CMOS with high-k/metal gate and 0.149μm^2 SRAM and ultra low-k back end with eleven levels of copper," *VLSI Technology Symposium* , pp.140-141, June 2009

[16] Chen, X. et al, "A cost effective 32nm high-K/ metal gate CMOS technology for low power applications with single-metal/gate-first process," *VLSI Technology Symposium*, pp.88-89, June 2008

[17] Narasimha, S. et al, "High Performance 45-nm SOI Technology with Enhanced Strain, Porous Low-k BEOL, and Immersion Lithography," *International Electron Devices Meeting (IEDM) Technical Digest* , pp.1-4, Dec. 2006

[18] Steegen, A.et al , "65nm cmos technology for low power applications," *International Electron Devices Meeting (IEDM) Technical Digest*, pp.64-67, Dec. 2005

[19] Leobandung, E.et al, "High performance 65 nm SOI technology with dual stress liner and low capacitance SRAM cell," *VLSI Technology, Symposium*, pp. 126- 127, June 2005

[20] K.-L. Cheng et al., "A Highly Scaled, High Performance 45nm Bulk Logic CMOS Technology with 0.242 μm^2 SRAM Cell," *IEEE International Electron Devices Meeting(IEDM) Technical Digest*, pp. 243-246, December 2007.

[21] Diaz, C.H., et al "32nm gate-first high-k/metal-gate technology for high performance low power applications," *International Electron Devices Meeting (IEDM) Technical Digest*, pp.1-4, 15-17 Dec. 2008

[22] R. T. Greenway et al., "Interference assisted lithography for patterning of 1D gridded design," *in Proceedings of SPIE*, vol. 7271, 2009

[23] Wilson, Syd R., Clarence J. Tracy, and John L. Freeman. Handbook of Multilevel Metallization for Integrated Circuits: Materials, Technology, and Applications. Park Ridge, N.J., U.S.A.: Noyes, 1993. 25. Print.

[24] Priyamvada V. et al "3D Integration requirements of Nanoscale fabrics and CMOS". in press *International Conference on Nanotechnology* (*IEEENANO)* 2011.

[25] Teman A., Pergament L., Cohen O., Fish A. "A Minimum Leakage Quasi-Static RAM Bitcell," *Journal of Low Power Electronics and Applications*. 2011; 1(1):204-218.

[26] Priyamvada V,"Impact of Nanomanufacturing Flow on Systematic Yield Losses in Nanoscale Fabrics", *in Proceedings of International Symposium on Nanoscale Architectures (NANOARCH)* 2011