Vol 05, Article 10479, November 2014
http://ijves.com

International Journal of VLSI and Embedded Systems-IJVES
ISSN: 2249 – 6556

# SURVEY ON MEMORY YIELD PREDICTION USING PHYSICS BASED 2-D DEVICES

NIKHIL RAJ[1], GNANA SHEELA K[2]
[1,2] Department of Electronics and Telecommunication, Toc H Institute of Science and Technology Kerala, India
Nikhilraj2810@gmail.com, sheelabijins@gmail.com

## ABSTRACT

*Accurate memory yield prediction using three dimensional methods with computer aided design tools poses a limitation that it involves computation of a large number of nodes in their mesh representation. The demerit can be overcome by using advanced physics-based 2-D devices with optimized meshes that are derived from 3-D Fin FET models with tuned device parasitic. I t has been shown that by using this method speed increases up to 200times and wall clock time increases around five times of magnitude compared to other traditional methods including montecarlo analysis. The proposed method can also be used for hardware measurements. This method enables physics-based simulation as well as physics-based variability input parameters. The proposed method is implemented on FinFET/tri-gate static random access memory (SRAM) design.*

***Keywords:*** *static random access memory (SRAM), Technology Computer Aided Design, FinFET*

## [1]INTRODUCTION

In both present and future technology generations where fast and accurate circuit simulations are required 3-D method poses a challenge to capture rapidly changing device features and intrinsic device fluctuations. Thus it makes difficult for compact models to catch up with process/technology changes. In this paper, it is proposed a methodology for improved Circuit design and manufacturability by providing a method for TFM of process-sensitive circuits. TFM approach can enable on the spot process/memory design yield optimizations even if real hardware is not available.

For intelligent systems such as autonomous mobile robots and intelligent vehicles, accurate three dimensional information is required for carrying out given tasks. DFM methods use software-based analysis tools for predicting. 3-D information can be obtained by reducing time intervals between frames. The proposed methodology is based on application of physical 2D devices into basic 3D design.

The proposed method is to build a robust and fast framework by combining TCAD and fast statistical sampling techniques [1]–[4] and the method is based on accurate models, fast algorithms and efficient simulations. The main advantages of this method is more accurate physics based 2-D mesh representations of the 3-D devices, advanced automated circuit parasitic extraction for generic FinFET circuits, efficient TCAD parametric representation for device variability including random dopant fluctuation and a fast statistical sampling-based simulation engine. The TCAD mixed-mode analysis enables more physical and accurate analysis of the effects of the statistical process parameter variation on the yield. Such effects may not be well abstracted or captured with traditional CMs, or existing table-based abstractions, which are designed to match only certain regions of the device characteristics.

The main advantages of this methodology includes estimating low fail probability with a reasonably low number of samples using fast statistical analysis, more physical and accurate analysis of the effects of the statistical process parameter variation on the yield using the TCAD mixed-mode analysis. With the increase in variability, process variations often stretch beyond the modeling limits, and traditional CM-based approaches often rely on extrapolations to account for such large variations, whereas a TFM approach captures the true physical device characteristics, which is critical for rare event estimation.

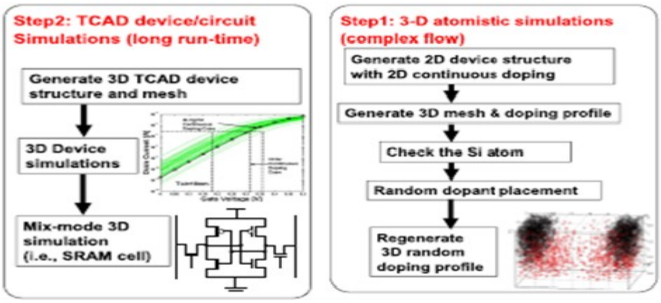
Step2: TCAD device/circuit Simulations (long run-time)
Generate 3D TCAD device structure and mesh
3D Device simulations
Mix-mode 3D simulation (i.e., SRAM cell)
Gate Voltage (V)
Step1: 3-D atomistic simulations (complex flow)
Generate 2D device structure with 2D continuous doping
Generate 3D mesh & doping profile
Check the Si atom
Random dopant placement
Regenerate 3D random doping profile

Fig. 1. Conventional 3D method and proposed 2-D method

The disadvantage of this method is that due to the structural difference in 3-D and 2-D, the discrepancy of parasitic capacitance in 3-D and 2-D must be modeled to achieve equal ac performance. Extraction is a major problem for non planar devices [5] therefore special rules are needed to recognize FinFETs. It should be noted that the lack of accurate parasitics modeling makes it difficult for the traditional CM simulation approach to predict proper functionality and enable design optimizations. The method can be effectively applied to early pre hardware stages as well as updated-process/device in consecutive design cycles. A key feature of the TFM methodology is the ability to predict yield trends for new processes who's CMs have not been developed.

## 2. LITERATURE SURVEY

**In 1989, Sodini c et al** observed a hierarchical framework which connects device and technology design parameters to specific circuit applications [5]. A comprehensive list of functional circuit blocks which are used in the Framework is defined. Technology-intensive circuit examples are also given which demonstrate the effect of technology enhancements on specific circuit applications. This framework and the device and technology data will help Circuit designers evaluate the impact of specific device and technology improvements on particular circuit and system requirements. it is still up to the circuit and system designer to decide which of those circuit functions must be optimized. This framework, Along with the device and technology data which have been presented, will help the circuit designer acquire a better understanding of the impact of device and technology improvements on his particular circuit and system requirements.

**In 1997, Seunghwan L et al** announced a three-dimensional(3D) instrumentation based on optical flow and is a promising instrumentation method for intelligent systems in which accurate 3-D information is required [6]. However, real-time instrumentation is difficult since much computation time and a large memory bandwidth are required. In this paper, a 3-D instrumentation VLSI processor with a concurrent memory-access scheme is proposed.

To reduce the access time, frequently used data are stored in a cache register array and are concurrently transferred to processing elements using simple interconnections to the 8-nearest neighbor registers. Based on a row and column memory access pattern, it proposes a diagonally interleaved frame memory by which pixel values of a row and column are stored across memory modules. Therefore, the pixel values are read in parallel from the frame memory. Based on the concurrent memory-access scheme, the performance of the processor is 2 million times faster than that of a 28.5 MIPS workstation.

**In 2000, Paliouras V et al** proposed a VLSI architecture for fast and accurate Floating-point sine/cosine evaluation, combining Floating-point and simple fixed-point arithmetic [7]. The algorithm Implemented by the architecture is based on second-order polynomial Interpolation within an approximation interval which is partitioned into regions of unequal length. The exploitation of certain properties of the trigonometric functions and of specific bit Patterns that appear in the involved computations, has led to reduced Memory size and low overall hardware complexity. In fact, a 40% memory size reduction is achieved by the introduced simplified Memory interleaving scheme, when compared to traditional interleaved memory architecture. The proposed architecture has been designed and simulated in a 0.7- m CMOS process technology, to prove its amenability for VLSI implementation. The time required to evaluate a sine is less than the time required for three single-precision floating-point multiply-accumulate (MAC) operations, while the computed values are guaranteed to be accurate to half a unit in last position.

**In 2003, Shimada Y et al** propose a new accurate yield prediction method for system-LSI embedded memories to improve the productivity of chips [8]. Their new method is based on the failure-related yield prediction method in which failure bits in memory are tested to see whether they are repairable or not by using built-in redundancies. The important concept of the new method is called "repairable matrix" (RM). In RM, $rm_{ij}$=1 means that i row redundancy sets and j column redundancy sets are needed for repair, where $rm_{ij}$ is an element of the matrix. Here, RM can indicate all the candidate combinations of the number of row and column redundancy sets for repair. The new yield prediction method using RM solves two problems, "asymmetric repair" and "link set." These have a significant effect on accurate yield prediction but have not yet been approached by conventional analytical methods. The calculation of yield by the new method is demonstrated in two kinds of advanced memory devices that have different design rules, failure situations, and redundancy designs. The calculated results are consistent with the actual yield. On average, the difference in accuracy between the new method and conventional analytical methods is about 5%.

**In 2003, Wang Y et al** proposed extensive transient simulations for on-chip Power delivery networks are required to analyze power delivery fluctuations caused by dynamic IR and Ldi/dt drops. Speed and memory has become a bottleneck for simulation of power distribution networks in modern VLSI design where clock frequency is of the order of Ghz [9]. The traditional SPICE based tools are very slow and consume a lot of memory during simulation. The problem is further aggravated for huge networks like power distribution Network within a stack of ic's inter-connected through tsvs. This type of 3D power distribution network may contain Billions of nodes at a time. It is proposed a faster transient simulation algorithm using visual C++. The Proposed algorithm is quite accurate with 1-2% error when compared with soft Nexxim4.1. This algorithm is several times faster than Ansoft Nexxim as well as consumes significantly less memory. The Proposed algorithm is quite accurate with 1-2% error as compared to Ansoft Nexxim4.1.

**In 2006, Joshi R et al** introduced a novel methodology for statistical SRAM design and analysis. It relies on an efficient form of importance sampling, mixture importance sampling [10]. The method is comprehensive, computationally efficient and the results are in excellent agreement with those obtained via standard Monte Carlo techniques. All this comes at significant gains in speed and accuracy, with speedup of more than 100X compared to regular Monte Carlo. To the best of their knowledge, this is the first time such a methodology is applied to the analysis of SRAM designs.
**In 2006, Sherwood T et al** proposed a system characterization of almost 60 real memory designs from the past 15 years [11]. Making good architectural decisions early in the design process requires a reasonably accurate model for the important structures. Dealing with continuously changing SRAM design practices and VLSI technologies make this a very difficult problem. Most hand-built memory models capture only a single parameterized design and fail to account for changes in design practice for different size memories or problems with wire scaling. Instead, in this paper it is presented a high level model that can be used to make simple analytical estimates.
**In 2011,Yang Y** et al proposed a new RAM/ROM module system with reconfigurable memory architecture for three-dimensional (3D) image processing VLSI system [12]. To enable flexible image data processing, suitable input/output data control is critical feature for high performance image processing system. The fast speed 3D VLSI system also requires efficient pipeline data operation. New RAM/ROM synthesis design system is realized by specific arrangement with RAM, ROM, pin and interconnection. The pipeline Flip- Flop control, clock buffer insertion and critical signal route have been improved to enhance whole system operation speed. The network-on-chip system is also proposed to enable fast signal transmission and correct control operation. The new 3D reconfigurable memory system can deal with inner data and control instruction signals directly for dependable VLSI chip. Further image robust methods, including self-repairable operation and re-healing system, are also used in proposed 3D image processing system.
**In 2011, Rosenfield P et al** proposed an algorithm for accurate memory simulator. many of the CPU Simulators overlook the need for accurate models of the memory System [13]. Many such simulators include simplistic models of memory which fail to account for the highly complex behavior of modern Memory systems. A typical DDR memory controller reorders and Schedules requests multiple times while keeping track of dozens of timing parameters.
**In 2014, Kanj R et al** proposed an efficient physics-based mixed mode statistical simulation methodology for nano scale devices and circuits [14]. Here, 3-D Technology Computer Aided Design Models pose a barrier for efficient simulation of variability as they generally involve millions of nodes in their mesh Representations. The proposed methodology, which has been implemented for finFET/tri-gate static random access memory (SRAM) design, overcomes this barrier by leveraging advanced physics-based 2-D (P2-D) devices with optimized Meshes that are derived from 3-D finFET models with tuned Device parasitics. This enables physics-based simulation as well as physics-based variability input parameters. The proposed physics based methodology is also shown to corroborate well with hardware measurements. The pragmatic physics based mixed-mode statistical simulation methodology for the first time. The runtime, which was impractical for statistical dynamic margin analysis before is made feasible via the newly developed P2-D meshes with enhanced parasitic modeling.
To improve Accuracy, an embedded automated flow enables extraction of all external nodal parasitics, directly from a 3-D finfet Circuit layout representation. The circuits consisting of advanced P2-D devices are then back annotated with the nodal parasitic to enable fast and accurate SRAM dynamic margin mixed-mode Simulations. Results demonstrate up to 200× speedup compared With traditional 3-D device simulations, and around five orders Of magnitude wall clock time improvement on account of fast Statistical methodologies, which are superior in comparison with Traditional Monte Carlo analysis. This makes it feasible to supplant often inaccurate compact model-based simulations by True mixed-mode device simulations in statistical engines.

## 3. COMPARITIVE ANALYSIS

Table.1. Comparative analysis on various design methodologies and prediction of memory

| Author | Year | Algorithm | Advantages | Disadvantages | Results |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| Sodini Yet al | 1989 | 1.High perform Bus-Dominated algorithm<br>2. Densely packed Static algorithm | 1.Improves power efficiency<br>2.. Analog Signal Processing Circuits<br>3. Analog interface circuits<br>4. Regular Structures. | 1. Speed depends on the device performance parameters<br>2.The intrinsic gate capacitance depends on the gate area And oxide thickness<br>3. Reduction of the oxide thickness is practically limited by yield and reliability Considerations.<br>4. Some advanced sub micrometer CMOS processes Have a smaller threshold voltage. | 1. Will help the circuit designer Acquire a better understanding of the impact of device and Technology improvements.<br>2. Significant benefit in Analog signal processing circuits.<br>3. Power, Speed, noise margin is good |
| Seunghwan G et al | 1997 | 1.Three-dimensional(3-D) instrumentation based algorithm<br>2. Diagonally interleaved frame memory algorithm<br>3. A cache register array and diagonal interleaving scheme | 1. Reduce the hardware overhead By interconnections.<br>2. We can reduce idle time of the subtracter. And the chip size.<br>3. A computer simulation of the algorithm shows That average error is about 5% | 1.Huge run time<br>2. Repeated use of memory<br>3. A large number of pixel values has to be stored | 1 the performance of the Processor is 2 million times faster than that of a 28.5 mips workstation..<br>2. Frequently used data are stored in a cache register array and are concurrently transferred To processing elements using simple interconnections to the 8-nearest neighbor registers |
| Paliouras V et al | 2000 | Floating-point and simple fixed-point arithmetic<br>2. Second-order polynomial Interpolation within an approximation interval<br>3. Hörner polynomial | 1. Computational complexity of determining the polynomial Coefficients has been reduced<br>2.the partitioning Of the approximation interval into regions of unequal lengths Reduces the memory space required to achieve the half an Ulp accuracy | 1. In this Case 2.5 Mbits of memory space and 3000 gates of additional Logic are required for the evaluation of sine.<br>2.high delay<br>3.high memory access | 1. The time required to evaluate a sine is less than the time required For three single-precision floating-point. Multiply-accumulate operations<br>2. For every benchmark worst slew does not exceed the slew limit of 100ps.<br>3. Run time is within the minutes.<br>4. All skews are within 3% of maximum latency. |
| Wang Y et al | 2003 | 1. Alternating-direction-implicit algorithm.<br>2.finite difference methods<br>3. Thermal simulation | 1. Highly accurate and efficient in memory Usage.<br>2.. Ensures that time step is not limited by any stability requirement<br>3.very fast algorithm | 1. Boundary conditions<br>2. High temperature not only causes timing failures for both Transistors and interconnects, but also degrades chip reliability | Efficient and accurate chip-level transient Thermal simulations |

| | | | | | |
|---|---|---|---|---|---|
| Shimada Y et al | 2003 | Repairable matrix | 1. Failure-related yield prediction method in which failure bits in memory are tested to see whether they are repairable or not | 1.asymmetric repair<br>2. Link set | The difference in accuracy between the new method and conventional analytical methods is about 5%. |
| Joshi R et al | 2006 | It relies on an efficient form of importance sampling, mixture importance sampling | Comprehensive, computationally efficient and the results are in excellent agreement with those obtained via standard Monte Carlo techniques | 1.hardware complexity<br>2.electronic design automation | 1.gains in speed and accuracy, with speedup of more than 100X compared to regular Monte Carlo<br>2. This is the first time such a methodology is applied to the analysis of SRAM designs. |
| Sherwood T et al | 2006 | Simple and intuitive functions algorithm | 1. Minimizes cost.<br>2. Faster performance and predictability of responses.<br>3. High level model that can Be used to make simple analytical estimates | 1.Focuses on power and uses.<br>2.methods are statistical<br>3.Computationally intensive | Can automatically capture the Most important scaling trends with underlying technology and Size. These data can also be used to guide the design and Recalibration of more detailed implementation-specific models. |
| Yang Y et al | 2011 | 1.Transient simulation algorithm<br>2. 3D power distribution Network<br>4.vertical stack algorithm | 1.high performance and low power<br>2. Interconnect size is reducing and integration density is Increasing | 1. Many sources of power fluctuation.<br>2. Coupling and interference effects have not been captured | 1.an efficient and accurate algorithm For dc solution of a 3d power distribution bus<br>2. The Algorithm is quite flexible and applicable to network of any Size |
| Rosenfield P et al | 2011 | 1. Dramsim2 algorithm<br>2. Dsim<br>3. Rascas algorithm<br>4. C++ STL | 1. A cycle accurate Memory system simulator<br>2. Effectively overcomes the negative influence of obstacles<br>3. Accurate And publicly available .<br>4. Used to perform full system simulations<br>5. Reduces the routing cost. | 1. Time consuming.<br>2. Not very Enlightening for simulations of real programs that execute for billions Of cycles<br>3. Power consumption of a dramsim2 simulation Of the PARSEC benchmark “fluid animate”.<br>4.. Higher burst power, Which implies higher bandwidth | 1. Dramsim2 has a strong Focus on being accurate and easy to integrate<br>2. Dramvis supports standard graph interactions such as zooming And panning as well as breaking down latency and bandwidth by Rank or by bank.<br>3. Dramsim2 is an invaluable tool for The growing use of simulation in the computer architecture |
| Kanj R et al | 2014 | Capacitance, fast statistical sampling, finfet, Physics-based models, static noise margin | 1.Fast Statistical Sampling Method<br>2.Fast and Accurate Mixed-Mode Simulation Flow<br>3.Efficient TCAD | 1.Due to the structural difference in 3-D and 2-D, the discrepancy of parasitic capacitance in 3-D and 2-D must | The runtime, Which was impractical for statistical dynamic margin Analysis before is made feasible via the newly developed |

| | | | | | |
|---|---|---|---|---|---|
| | | (SNM), static random Access memory (SRAM), TCAD. | Simulation 4.3-D TCAD Layout-Aware Capacitance Extraction | be modeled to achieve equal ac performance 2.four types of delays | P2-D meshes with enhanced parasitic modeling |

## 4. CONCLUSION

This paper presents mapping of physics-based mixed-mode statistical Simulation methodology into conventional 3-D method. By using the newly developed P2-D meshes with enhanced parasitic modeling the runtime is analyzed. In order to increase speeds and up dynamic margin simulations the P2-D and DFM is used. It will practicality reduce the simulation runtime from several Months to hours for an eight-transistor design. Cell capacitances, which are a must for accurate circuit and layout interaction representation, are generated using automated structure generation at the circuit Level. The methodology Using finFET process technology is evaluated by most modern techniques (as a vehicle and observed Orders of magnitude speed improvement) for 3-D finFET-based SRAM cell design. It is also showed that SPICE cms, (Which are derived from the 3-D finFET models) deviate under Statistical variation conditions. Overall it is proved that Overall, TCAD simulations cannot be directly inserted into current generation DfM flows. Most importantly, the resultant configuration enables dynamic margin simulations with matched accuracy and 200× speedup compared with full 3-D simulations. . These methods are therefore a preferred alternative when it comes to the yield analysis of large and repetitive array structures

## REFERENCES

[1] A. Singhee and R. Rutenbar "Statistical blockade: A novel method for very fast Monte Carlo simulation of rare circuit events, and its application," in Proc. IEEE Des. Autom. Test Conf., Mar. 2007, pp. 1–6.

[2] R. Kanj, R. Joshi, and S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare fail events," in Proc. IEEE Des. Autom. Conf., Jun. 2006, pp. 69–72.

[3] C. Dong and X. Li, "Efficient SRAM failure rate prediction via Gibbs sampling," in Proc. IEEE Des. Autom. Conf., Jun. 2011, pp. 200–205.

[4] D. E. Hocevar, M. R. Lightner, and T. N. Trick, "A study of variance reduction techniques for estimating circuit yields," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 2, no. 3, pp. 180–192, 1983.

[5] Charles G Sodini , S Simon Wong , Ping Keung co IEEE journal of solid-state circuits. vol.24, no.1, 1989

[6] Seunghwan Lee, Masanori Hariyama and Michitaka Kameyama Department of Computer and Mathematical Sciences Graduate School of Information Sciences, Tohoku University, SICE '97. Proceedings of the 36th SICE Annual Conference, International Session Papers.

[7] Vassilis Paliouras, Member, IEEE, Konstantina Karagianni, Member, IEEE, and Thanos Stouraitis, , IEEE transactions on circuits and systems,analog and digital signal processing, vol. 47, no. 5, may 2000

[8] Shimada, Y. ; Renesas Technol. Corp., Hyogo, Japan ; Sakurai, K. Semiconductor Manufacturing, IEEE transactions on (Volume:16 , Issue: 3 )

[9] Ting-Yuan Wang and Charlie Chung-Ping Chen IEEE transactions on very large scale integration (vlsi) systems, vol. 11, no. 4, august 2003

[10] IBM Austin Research Labs, Austin Rouwaida Kanj TX,Rajiv Joshi IBM tj Watson labs,Yorktown Heights,NY, Sani Nassif, IBM Austin Research Labs, Austin, TX, DAC '06 Proceedings of the 43rd annual Design Automation Conference

[11] Banit Agrawal Timothy Sherwood Department of Computer Science, University ofCalifornia, Santa IEEE,. International Conference on computer Computer Design, 2006. ICCD 2006, at San Jose, CA

[12] Yun Yang, IEEE,Three-dimensional Image Processing VLSI System with Network-on-chip System and Reconfigurable Memory Architecture, page no 1345

[13] Paul Rosenfeld, Elliott Cooper-Balis, Bruce Jacob Department of Electrical and Computer Engineering University of Maryland College Park, IEEE computer architecture letters, vol. 10, no. 1, january-june 2011

[14] Rajiv V. Joshi, Fellow, Keunwoo Kim, Senior Member, Rouwaida Kanj, Senior Member, ,Ajay N. Bhoj, Matthew M. Ziegler, Phil Oldiges, IEEE, Pranita Kerber, Robert Wong, Terence Hook, Sudesh Saroop, Carl Radens, and Chun-Chen (Frank) Yeh IEEE transactions on very large scale integration systems 2014