Importance Sampling for Determining SRAM Yield and Optimization with Statistical Constraint

E.J.W. ter Maten, O. Wittich, A. Di Bucchianico, T.S. Doorn, and T.G.J. Beelen

Abstract Importance Sampling allows for efficient Monte Carlo sampling that also properly covers tails of distributions. From Large Deviation Theory we derive an optimal upper bound for the number of samples to efficiently sample for an accurate fail probability $P_{\text{fail}} \leq 10^{-10}$. We apply this to accurately and efficiently minimize the access time of Static Random Access Memory (SRAM), while guaranteeing a statistical constraint on the yield target.

1 Introduction

As transistor dimensions of Static Random Access Memory (SRAM) become smaller with each new technology generation, they become increasingly susceptible to statistical variations in their parameters. These statistical variations may result

E.J.W. ter Maten (✉)
Eindhoven University of Technology, Department of Mathematics and Computer Science, CASA/LIME, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
NXP Semiconductors, High Tech Campus 32 and 46, 5656 AE Eindhoven, The Netherlands
Bergische Universität Wuppertal, Fachbereich C, Wicküler Park Rm 503, Bendahler Str. 29, D-42285, Wuppertal, Germany
e-mail: E.J.W.ter.Maten@tue.nl; Jan.ter.Maten@nxp.com; Jan.ter.Maten@math.uni-wuppertal.de

A. Di Bucchianico
Eindhoven University of Technology, Department of Mathematics and Computer Science, CASA/LIME, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: A.D.Bucchianico@tue.nl

O. Wittich
RWTH Aachen, Lehrstuhl A für Mathematik, Analysis und Zahlentheorie, Schinkelstr. 4, D-52056 Aachen, Germany
e-mail: Olaf.Wittich@mathA.rwth-aachen.de

T.S. Doorn · T.G.J. Beelen
NXP Semiconductors, High Tech Campus 32 and 46, 5656 AE Eindhoven, The Netherlands
e-mail: Toby.Doorn@nxp.com; Theo.G.J.Beelen@nxp.com

B. Michielsen and J.-R. Poirier (eds.), *Scientific Computing in Electrical Engineering SCEE 2010*, Mathematics in Industry 16, DOI 10.1007/978-3-642-22453-9_5,
© Springer-Verlag Berlin Heidelberg 2012

in failing memory. An SRAM is used as a building block for the construction of large Integrated Circuits (ICs). To ensure that a digital bit cell in SRAM does not degrade the yield (fraction of functional devices) of ICs with Megabits of memory, very small failure probabilities $P_{\text{fail}} \leq 10^{-10}$ are necessary. To simulate this, regular Monte-Carlo (MC) simulations require too much computing time. Importance Sampling (IS) [1] is a more advanced technique that provides sufficiently accurate results and is relatively easy to implement. A speed up of several orders can be achieved when compared to regular Monte Carlo methods.

2 Regular Monte Carlo

Let Y be a real-valued random variable with probability density function f. We assume that N independent random observations Y_i $(i = 1, \ldots, N)$ of Y are taken. We define $X_i = I_A(Y_i)$ for a given set $A = (-\infty, x)$ where $I_A(Y_i) = 1$ if $Y_i \in A$ and 0 otherwise. Then $p_f^{\text{MC}}(A) = \frac{1}{N}\sum_{i=1}^{N} X_i$ estimates $p = \int_{-\infty}^{x} f(z)dz = P(Y \in A)$. The X_i are Bernoulli distributed, hence $Np_f^{\text{MC}} \sim \text{Bin}(N, p)$, $\text{E}(p_f^{\text{MC}}) = \frac{1}{N}Np = p$, and $\sigma^2(p_f^{\text{MC}}) = \frac{p(1-p)}{N}$. Let $\Phi(x) = \int_{-\infty}^{x} \text{e}^{-z^2/2}dz$ and define z_α by $\Phi(-z_\alpha) = \alpha$. From the Central Limit Theorem (CLT) we derive

$$P(|p_f^{\text{MC}} - p| > \varepsilon) = P\left(\frac{|p_f^{\text{MC}} - p|}{\sigma(p_f^{\text{MC}})} > z\right) \overset{N_{\text{MC}}\to\infty}{\longrightarrow} 2\Phi(-z) \leq 2\Phi(-z_{\alpha/2}) = \alpha,$$

where $z = \varepsilon / \sqrt{p(1-p)/N_{\text{MC}}}$ and $N_{\text{MC}} = N$. Hence, if $z \geq z_{\alpha/2}$ we deduce

$$N_{\text{MC}} \geq p(1-p)\left(\frac{z_{\alpha/2}}{\varepsilon}\right)^2 = \frac{1-p}{p}\left(\frac{z_{\alpha/2}}{\nu}\right)^2, \tag{1}$$

for $\varepsilon = \nu p$. We take $\nu = 0.1$ and $p = 10^{-10}$. Now let $\alpha = 0.02$, then $z_{\alpha/2} \approx 2$. Then $N_{\text{MC}} \geq 4\ 10^{12}$. If we do not know p, we can use $p(1-p) \geq 1/4$ yielding $N_{\text{MC}} \geq \frac{1}{4}\left(\frac{z_{\alpha/2}}{\varepsilon}\right)^2 = 10^{22}$. And if N_{MC} is not large enough to apply the CLT, Chebyshev's inequality even results to $N_{\text{MC}} \geq 10^{24}$. These general bounds are much too pessimistic. Large Deviations Theory (LDT) [1,4] results in a sharp upper bound [6]

$$P(|p_f^{\text{MC}} - p| > \nu p) \leq \exp\left(-\frac{N_{\text{MC}}}{2}\frac{p}{1-p}\nu^2\right). \tag{2}$$

For $\nu = 0.1$, $p = 10^{-10}$ and $\alpha = 0.02$, as above, we find: $N_{\text{MC}} \geq 8\ 10^{12}$ (which is a sharp result – see at the end of the next proof). Note that an extra k-th decimal in ν increases N_{MC} with a factor k^2.

Proof of (2) [6]. The sequence of the Monte Carlo results $P_N(A) := p_f^{\mathrm{MC}}$ satisfies a *Large-Deviation Principle* [1, 4, 5], meaning that there is some *'rate function'* $I : \mathbb{R} \to \mathbb{R} \cup \{-\infty, +\infty\}$ such that

(i) $\limsup_{N\to\infty} \frac{1}{N} \ln P_N(C) \le -\inf_{x\in C} I(x)$ for all closed subsets $C \subset \mathbb{R}$,
(ii) $\liminf_{N\to\infty} \frac{1}{N} \ln P_N(G) \ge -\inf_{x\in G} I(x)$ for all open subsets $G \subset \mathbb{R}$.

Let X be a Bernoulli variable with success probability p. The *logarithmic moment generating function* for X is given by $\ln\left(\mathrm{E}\left[e^{\lambda X}\right]\right) = \ln\left(q + e^{\lambda} p\right)$, where as usual $q = 1 - p$. We define the following function [5]

$$J(x,\lambda) = \lambda x - \ln\left(\mathrm{E}\left[e^{\lambda X}\right]\right) = \lambda x - \ln(q + e^{\lambda} p), \tag{3}$$

where $x, \lambda \in \mathbb{R}$. We note that an optimum value λ^* must satisfy

$$\frac{\partial J}{\partial \lambda} = x - \frac{pe^{\lambda^*}}{q + pe^{\lambda^*}} \quad = \quad 0, \quad \text{hence}$$

$$\lambda^* = \ln\left(\frac{qx}{p(1-x)}\right), \quad \text{and} \quad pe^{\lambda^*} = \frac{qx}{1-x}, \quad \text{and} \quad q + pe^{\lambda^*} = \frac{q}{1-x}. \tag{4}$$

In our case, the rate function can be shown to be equal to

$$I(x) = \sup_{\lambda\in\mathbb{R}} J(x,\lambda) = J(x,\lambda^*) = x \ln\left(\frac{qx}{p(1-x)}\right) - \ln\left(\frac{q}{1-x}\right), \tag{5}$$

a function which is continuous on the interval $(0,1)$. With $C = [p - \nu p, p + \nu p] \subset (0,1)$ and $G = \mathbb{R} \setminus C$, the Large-Deviation Principle above implies

$$\lim_{N\to\infty} \frac{1}{N} \ln P\left(\left|\frac{1}{N}\sum_{k=1}^{N} X_k - p\right| \ge \nu p\right) = -\inf_{|x-p|\ge\nu p} I(x).$$

From (5) we can calculate $I'(x)$ and $I''(x)$ explicitly. For $x \in (0,1)$ we have $I''(x) > 0$, which implies that I' is increasing and that I is *convex*. Also $I(0^+) = -\ln(q) > 0$ and $I(1^-) = \ln(q/p) \in \mathbb{R}$. Clearly I can be extended continuously at both $x = 0$ and $x = 1$. Furthermore $I(p) = 0$ and $I'(p) = 0$. Hence $I(p) = 0$ is a global minimum. This implies that actually the infimum of I on $\{x : |x - p| > \nu p\}$ is assumed at $x = p \pm \nu p$. This can be analyzed further using Taylor expansion [6]. Thus from part (i) of the Large Deviation Principle, we obtain (2) for all N with a possible exception of finitely many. Part (ii) implies that the exponential bound in (2) is also valid from below and thus is sharp. □

3 Importance Sampling

With Importance Sampling we sample the Y_i according to a different distribution function g and observe that $p_f(A) = \int_{-\infty}^{x} f(z)dz = \int_{-\infty}^{x} \frac{f(z)}{g(z)} g(z)dz$. Define $V_i = I_A(Y_i) f(Y_i)/g(Y_i)$ and $V = V(A) = I_A(Y) f(Y)/g(Y)$. Let $p_f^{\mathrm{IS}}(A) = \frac{1}{N}\sum_{i=1}^{N} V_i$. Then $\mathrm{E}_g\left(p_f^{\mathrm{IS}}\right) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}_{\mathrm{g}}(V_i) = p_f(A)$. When $\frac{f(z)}{g(z)} \leq 1$ on A we have $\mathrm{Var}_g\left(p_f^{\mathrm{IS}}\right) \leq \mathrm{Var}_f\left(p_f^{\mathrm{MC}}\right)$ (variance reduction, using the same number of samples). This does not yet imply more efficiency. However, similar to (2), we derive (in which $N_{\mathrm{IS}} = N$) [6]

$$P\left(\left|p_f^{\mathrm{IS}} - p\right| > \nu p\right) \leq \exp\left(-\frac{N_{\mathrm{IS}}\ p^2}{2\mathrm{Var}_g(V)}\,\nu^2\right). \tag{6}$$

Assuming the same upper bounds, comparing (2) and (6) gives $\frac{N_{IS}}{N_{MC}} = \frac{\mathrm{Var_g}(V)}{p(1-p)} = \frac{\mathrm{E}_g(V^2)-p^2}{p(1-p)}$. Suppose $\frac{f(z)}{g(z)} \leq \kappa < 1$ on A and $p \leq \kappa$, then, with $q = 1-p$,

$$\frac{N_{IS}}{N_{MC}} = \frac{\mathrm{E}_g(V^2)}{pq} - \frac{p}{q} \leq \frac{\kappa}{q} - \frac{p}{q} \leq \kappa(1+\zeta) \tag{7}$$

for $|(1-\frac{1}{\kappa})p + \mathcal{O}(p^2)| \leq \zeta$, which for $\kappa = 0.1$ and $p = 10^{-10}$ means that $\zeta \leq 10^{-9}$. Hence for $\kappa = 0.1$ we can take an order less samples with Importance Sampling to get the same accuracy as with Monte Carlo. This even becomes better with smaller κ. Efficiency is the main message. Indeed the asymptotic accuracy also improves, but less: $\mathrm{Var}_g\left(p_f^{\mathrm{IS}}\right) \leq \kappa\ \mathrm{Var}_f\left(p_f^{\mathrm{MC}}\right) - \frac{1-\kappa}{N} p^2$ and thus $\sigma_g\left(p_f^{\mathrm{IS}}\right) \leq \sqrt{\kappa}\,\sigma_f\left(p_f^{\mathrm{MC}}\right)$, which for $\kappa = 0.1$ means that here not an order is gained, but a factor $\sqrt{\kappa} \approx 0.316$.

Proof of (6) [6]. Let Y be distributed according to g, $V = I_{(-\infty,x)}(Y) f(Y)/g(Y)$ and $v(y) = I_{(-\infty,x)}(y) f(y)/g(y)$. Then

$$\mathrm{E}_g\left[e^{\lambda V}\right] = \int_{-\infty}^{\infty} g(y)\, e^{\lambda I_{(-\infty,x)} f(y)/g(y)}\, dy = \int_{-\infty}^{x} g(y)\, e^{\lambda f(y)/g(y)}\, dy + 1 - G(x),$$

where $G(x) = \int_{-\infty}^{t} g(y)\, dy$. We will restrict ourselves to simple *sufficient* conditions and we will not strive for full generality. We assume:

1. There is no $y \in \mathbb{R}$ such that $P(Y = y) = 1$ (Y is not supported by a single point),
2. $0 < \mathrm{E}_g\left[e^{\lambda V}\right] < \infty$ for all $\lambda \in \mathbb{R}$,
3. Introduce the density function $\rho_\lambda(y)$

$$\rho_\lambda(y) = \frac{e^{\lambda v(y)} g(y)}{\mathrm{E}_g\left[e^{\lambda V}\right]} \quad \left(\text{thus} \int \rho_\lambda(y)\, dy = 1\right)$$

(which is well-defined for all $\lambda \in \mathbb{R}$) and let Y_λ be a random variable distributed according to ρ_λ. We assume that for all $\lambda \in \mathbb{R}$

$$\mathrm{E}_{\rho_\lambda}(Y_\lambda) = \int y_\lambda \rho_\lambda(y_\lambda)\, dy_\lambda = \int y \frac{e^{\lambda v(y)} g(y)}{\mathrm{E}_g\left[e^{\lambda V}\right]}\, dy < \infty$$

and

$$\mathrm{Var}_{\rho_\lambda}(Y_\lambda) = \mathrm{E}\left[Y_\lambda^2\right] - \mathrm{E}^2_{\rho_\lambda}(Y_\lambda) < \infty.$$

Now let $\varphi(\lambda) = \ln \mathrm{E}_g\left[e^{\lambda V}\right]$. Then, $\varphi(\lambda)$ is a well-defined, two times differentiable, real function with derivatives

$$\varphi'(\lambda) = \frac{\mathrm{E}_g\left[V\, e^{\lambda V}\right]}{\mathrm{E}_g\left[e^{\lambda V}\right]} = \mathrm{E}_{\rho_\lambda}(Y_\lambda), \quad \varphi''(\lambda) = \frac{\mathrm{E}_g\left[V^2\, e^{\lambda V}\right]}{\mathrm{E}_g\left[e^{\lambda V}\right]} - \frac{\mathrm{E}_g^2\left[V\, e^{\lambda V}\right]}{\mathrm{E}_g^2\left[e^{\lambda V}\right]} = \mathrm{Var}_{\rho_\lambda}(Y_\lambda).$$

Clearly, $\mathrm{Var}(Y_\lambda) > 0$ and φ is therefore *strictly convex*. Let $J(x,\lambda) = \lambda x - \varphi(\lambda)$. As in Sect. 2 we again consider the function $I(x) = \sup_{\lambda \in \mathbb{R}} J(x,\lambda)$ [5]. Clearly $I(x) \geq J(x,0) = -\varphi(0) = -\ln e^0 = 0$. To compute the supremum in $I(x)$, we consider

$$\frac{d}{d\lambda} J(x,\lambda) = x - \frac{d}{d\lambda}\varphi(\lambda) = x - \frac{\mathrm{E}_g\left[V e^{\lambda V}\right]}{\mathrm{E}_g\left[e^{\lambda V}\right]}\,. \tag{8}$$

We observe that

$$\frac{d}{d\lambda} J(x,\lambda) = 0 \Longrightarrow x = \Psi(\lambda), \quad \text{where} \quad \Psi(\lambda) = \frac{\int y\, e^{\lambda\, v(y)} g(y)\, dy}{\int e^{\lambda\, v(y)} g(y)\, dy}. \tag{9}$$

Here we note that

$$\Psi'(\lambda) = \frac{\int e^{\lambda\, v(y)} g(y)\, dy \int y^2 e^{\lambda\, v(y)} g(y)\, dy \quad - \quad [\int y e^{\lambda\, v(y)} g(y)\, dy]^2}{[\int e^{\lambda\, v(y)} g(y)\, dy]^2}. \tag{10}$$

At the right-handside we can recognize a weighted inner-product (using weight function $e^{\lambda\, v(y)}$): $< 1, y > \equiv \int 1 \cdot y e^{\lambda v(y)} g(y)\, dy$. By the Cauchy-Schwarz inequality, $< 1, y > \leq \sqrt{< 1, 1 >}\sqrt{< y, y >}$ we obtain $\Psi'(\lambda) > 0$ because $y \neq 1$. This implies that Ψ is invertible and hence (9) defines $\lambda = \lambda(x) = \Psi^{-1}(x)$. Hence

$$I(x) = J(x, \lambda(x)) \tag{11}$$

and we can write $x = \Psi(\lambda) = \mathrm{E}_{\rho_\lambda}[Y]$. Clearly $\rho_{\lambda=0}(y) = g(y)$. Further, to calculate the first (total) derivative of $I(x)$, we differentiate (11) with respect to x and substitute (9) to obtain $I'(x) = \lambda(x)$ and $I''(x) = \lambda'(x) = 1/\frac{\partial x}{\partial \lambda} = 1/\mathrm{Var}_{\rho_\lambda}(V)$ [6]. By [5, Lemma I.4, p. 8], $I(x)$ is strictly (proper) convex which means that the minimizer of I is unique. Now let p be as in Sect. 2. Then $I(p) = 0$, since the Strong Law of Large Numbers implies that the empirical measure of every neighbourhood of p tends to one. Hence, p is the unique minimizer of I and $I'(p) = 0$. Since p is also an internal point, we obtain that $0 = I'(p) = \lambda(p)$. Hence,

$$I''(p) = \frac{1}{\mathrm{Var}_{\rho_{\lambda(p)}}(V)} = \frac{1}{\mathrm{Var}_{\rho_{\lambda=0}}(V)} = \frac{1}{\mathrm{Var}_g\,(V)}. \tag{12}$$

Finally, by Taylor expansion, $I(p \pm \nu p) = \frac{1}{2}\nu^2 p^2 I''(p) + \mathscr{O}(\nu^3 p^3) = \frac{1}{2}\frac{\nu^2 p^2}{\mathrm{Var}_g\,(V)}$. Thus, after applying the *Large-Deviation Principle* [1,4,5], as in Sect. 2,

$$P\left(\left|\frac{1}{N}\sum_{k=1}^{N} V_k - p\right| > \nu p\right) \leq \exp\left(-N \inf_{|x-p|>\nu p} I(x)\right) \approx \exp\left(-\frac{Np^2}{2\mathrm{Var}_g(V)}\,\nu^2\right), \tag{13}$$

for all sufficiently large N. This implies (6), which completes the proof.
We finally note that, if $g(x) \equiv 1$, as in Sect. 2, we have $\mathrm{Var}_g\,(V) = \frac{1}{pq}$, see (2). □

4 Accurate Estimation of SRAM Yield

The threshold voltages V_t of the six transistors in an SRAM cell are the most important parameters causing variations of the characteristic quantities of an SRAM cell [2] like Static Noise Margin (SNM) and Read Current (I_{read}). In [2, 6] Importance Sampling (IS) was used to accurately and efficiently estimate low failure probabilities for SNM and I_{read}. SNM $= \min(\text{SNM}_h, \text{SNM}_l)$ is a measure for the read stability of the cell. SNM_h and SNM_l are identically Gaussian distributed. The min() function is a non-linear operation by which the distribution of SNM is no longer Gaussian. Figure 1-left, shows the cumulative distribution function (CDF) of the SNM, using 50k trials, both for regular MC (solid) and IS (dotted). Regular MC can only simulate down to $P_{\text{fail}} \leq 10^{-5}$. Statistical noise becomes apparent below $P_{\text{fail}} \leq 10^{-4}$. With IS (using a broad uniform distribution g), $P_{\text{fail}} \leq 10^{-10}$ is easily simulated (we checked this with more samples). The correspondence between regular MC and IS is very good down to $P_{\text{fail}} \leq 10^{-5}$. Figure 1-left clearly shows that using extrapolated MC leads to overestimating the SNM at $P_{\text{fail}} = 10^{-10}$. The Read Current I_{read} is a measure for the speed of the memory cell. It has a non-Gaussian distribution. Figure 1-right shows that extrapolated MC (dashed) can result in serious underestimation of I_{read}. This can lead to over-design of the memory cell. Also here IS is essentially needed for sampling I_{read} appropriately.

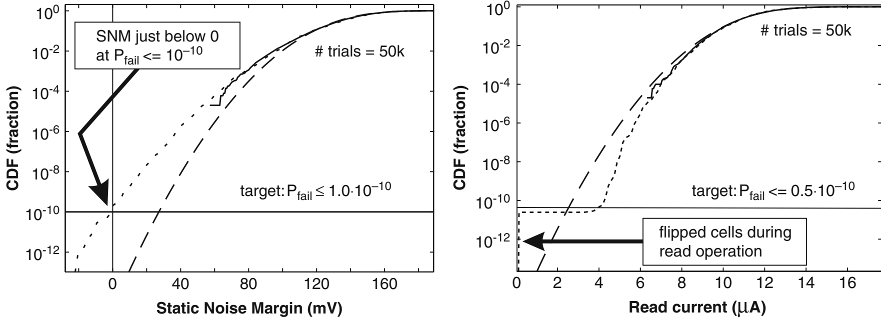

Fig. 1 SNM (left) and I_{read} (right) cumulative distribution function for extrapolated MC (dashed), regular MC (solid) and IS (dotted). Extrapolation assumes a normal distribution

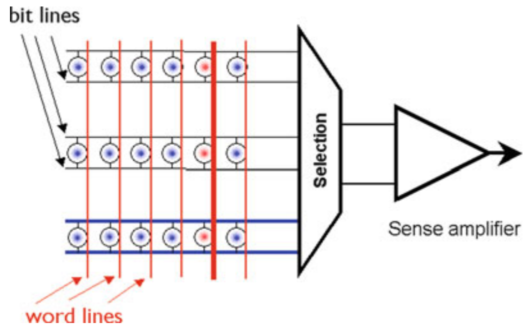

Fig. 2 Block of SRAMs (rotated 90°)

5 Optimization of SRAM Block

The block in Fig. 2 (rotated 90°) contains a Sense Amplifier (SA), a selector, and a number of SRAM cells. The selector chooses one "column" of cells. Then the voltage difference is $\Delta V_{\text{cell}} = \Delta V_{\text{k}}$. A block B works if $\min_{\text{k}}(\Delta V_{\text{k}}) \geq \Delta V_{\text{SA}}$. With m blocks B and n cells per block we define Yield Loss by $YL = P(\#B \geq 1) \leq m\,P(B)$, where the fail probability $P(B) = P_{\text{fail}}(B)$ of one block is (accurately) approximated by the lower bound $P(B) \approx \frac{YL}{m} = \frac{nYL}{N}$, where $N = n\,m$. For $YL = 10^{-3}$, $m = 10^4$ blocks, $n = 1000$ we find $P(B) \leq 10^{-7}$.
For $X = \min_k(\Delta V_k)$, and $Y = \Delta V_{\text{SA}}$ we have

$$P(B) = P(X < Y) = \int\int_{-\infty \leq x < y \leq \infty} f_{X,Y}(x, y)dx\,dy = \int_{-\infty}^{\infty} f_Y(y)\,F_X(y)dy.$$

Thus we need the pdf $f_Y(y)$ and the cdf $F_X(y)$ (probability and cumulative density functions of Y and X). Note that

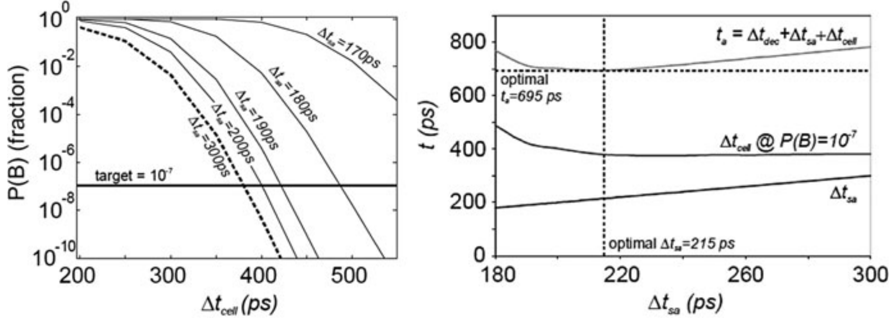

Fig. 3 Left: $P(B)$ as function of $\Delta t_{\rm cell}$ and $\Delta t_{\rm SA}$. Right: Delay time t as function of $\Delta t_{\rm SA}$

$$F_X(y) = P(X < y) = P(\min_k \Delta V_k < y) \\ = 1 - [1 - P(\Delta V_k < y)]^n \leq n\, P(\Delta V_k < y).$$

For each simulation of the block we can determine the access times $\Delta t_{\rm cell}$ and $\Delta t_{\rm SA}$. We come down to an optimization problem with a statistical constraint:

Minimize $\Delta t_{\rm cell} + \Delta t_{\rm SA}$ such that $P(B) \leq 10^{-7}$.

This has led to the following algorithm. We only give a sketch; for details see [3].

- By **Importance Sampling** sample ΔV_k. Collect ΔV_k at same $\Delta t_{\rm cell}$.
- By **Monte Carlo** sample $\Delta V_{\rm SA}$. Collect $\Delta V_{\rm SA}$ at same $\Delta t_{\rm SA}$.
- For given $\Delta t_{\rm cell}$:
  - Estimate pdf $f_{\Delta V_k}$ and cdf $P(\Delta V_k < y)$.
  - From this calculate $F_X(y) = F_X(y; \Delta t_{\rm cell})$. Note that $\frac{\partial F_X(y;\Delta t_{\rm cell})}{\partial \Delta t_{\rm cell}} \leq 0$.
- For given $\Delta t_{\rm SA}$:
  - Estimate pdf of $\Delta V_{\rm SA}$: $f_Y(y)$.
- Calculate (numerical integration)
  - $P(B) = \int_{-\infty}^{\infty} f_Y(y)\, F_X(y) dy$.

Hence $P(B) = G(\Delta t_{\rm cell}, \Delta t_{\rm SA})$ for some function G. For given $\Delta t_{\rm SA}$ $G_1(\Delta t_{\rm cell}; \Delta t_{\rm SA}) = G(\Delta t_{\rm cell}, \Delta t_{\rm SA})$ is monotonically decreasing in $\Delta t_{\rm cell}$, see Fig. 3. Hence we *Minimize* $G_1^{-1}(10^{-k}; \Delta t_{\rm SA}) + \Delta t_{\rm SA}$. The optimization with the statistical constraint on $P(B)$ led to a reduction of 6% of the access time of an already optimized SA while simultaneously reducing the silicon area [3].

6 Conclusions

Large Deviation Theory allows to derive sharp lower and upper bounds for estimating accuracy of tail probabilities of quantities that have a non-Gaussian distribution. For Monte Carlo this leads to a realistic number of samples that should

be taken. We extended this to Importance Sampling (IS). IS was applied to estimate fail probabilities $P_{\text{fail}} \leq 10^{-10}$ of SRAM characteristics like Static Noise Margin (SNM) and Read Current (I_{read}). We also applied IS to minimise the access time of an SRAM block while guaranteeing that the fail probability of one block is small enough.

In our experiments we used a fixed distribution g in the parameter space. In [6] ideas with an adaptively determined distribution g can be found.

References

1. Bucklew, J.A.: Introduction to Rare Event Simulation. Springer, Berlin (2004)
2. Doorn, T.S., ter Maten, E.J.W., Croon, J.A., Di Bucchianico, A., Wittich, O.: Importance Sampling Monte Carlo simulation for accurate estimation of SRAM yield. In: Proceedings of the IEEE ESSCIRC'08, 34th European Solid-State Circuits Conference, Edinburgh, Scotland, pp. 230–233 (2008)
3. Doorn, T.S., Croon, J.A., ter Maten, E.J.W., Di Bucchianico, A.: A yield statistical centric design method for optimization of the SRAM active column. In: Proceedings of the IEEE ESSCIRC'09, 35th European Solid-State Circuits Conference, Athens, Greece, pp. 352–355 (2009)
4. de Haan, L., Ferreira, A.: Extreme Value Theory. Springer, Berlin (2006)
5. den Hollander, F.: Large Deviations. Fields Institute Monographs 14, The Fields Institute for Research in Math. Sc. and AMS, Providence, RI (2000)
6. ter Maten, E.J.W., Doorn, T.S., Croon, J.A., Bargagli, A., Di Bucchianico, A., Wittich, O.: Importance Sampling for high speed statistical Monte-Carlo simulations – Designing very high yield SRAM for nanometer technologies with high variability. TUE-CASA 2009-37, TU Eindhoven (2009), http://www.win.tue.nl/analysis/reports/rana09-37.pdf