Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Design-Point Transfer With Conformal Calibration for SRAM Reliability Sign-Off Across PVT Conditions

Nabanita Sarker

2006162@eee.buet.ac.bd

Bangladesh University of Engineering and Technology https://orcid.org/0009-0000-0727-9137

Research Article

Keywords: Conformal prediction, design point, failure probability, first-order reliability method (FORM), importance sampling, minimum operating voltage, PVT corners, reliability sign-off, SRAM, static noise margin, surrogate models

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11455984/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Design-Point Transfer With Conformal Calibration for SRAM Reliability Sign-Off Across PVT Conditions

Nabanita Sarker

***Abstract*—SRAM yield requires per-cell failure probabilities of 10^{-6} to 10^{-12} at every process, voltage, and temperature (PVT) condition. We evaluate cross-PVT estimators against importance-sampling (IS) truth at all 242 unseen conditions of a SkyWater 130-nm foundry bitcell and an ASAP7 7-nm FinFET cell (read and write), cross-checked by plain Monte Carlo and a second IS proposal. A Gaussian tail with the exact margin moments errs by up to 11.9 decades, so moment surrogates fail through tail shape, not extrapolation. Extrapolating per-lobe reliability indices is accurate for the FinFET cell but unsafe by up to 1.9 decades for the planar cell, whose failure direction rotates across PVT. We transfer each lobe's design point, correct it with one SPICE run per lobe, which removes the first-order error, and calibrate the result: split-conformal margins, computed from IS on a random subset of the sign-off conditions, bound the probability of an unsafe acceptance at a randomly selected condition (a marginal, not grid-wide, guarantee), and undecided conditions escalate to IS. With 20 calibration conditions (nominal risk 0.15), no calibrated decision was unsafe, and the minimum operating voltage read from the conformal upper limit lay −1 to +34 mV from IS, whereas FORM, the unguarded transfer, and moment surrogates were unsafe by up to −7, −10, and −219 mV. Signed off this way, a 1-Mb macro targeting 99.9 % yield reached at least 99.88 %, versus 99.66 % with FORM-class estimators and near zero with moment surrogates, using 3.1–4.9 times fewer SPICE runs than IS everywhere.**

***Index Terms*—Conformal prediction, design point, failure probability, first-order reliability method (FORM), importance sampling, minimum operating voltage, PVT corners, reliability sign-off, SRAM, static noise margin, surrogate models**

NOMENCLATURE

AK-IS	Active-learning Kriging with importance sampling.
$\mathbf{c}$	Global PVT condition (T, V_{DD}, corner).
d	Number of standardized mismatch variables.
FORM	First-order reliability method.
GP	Gaussian process.
IS	Importance sampling.
MC	Monte Carlo.
m_0, m_1	Signed margins of the two mirror lobes (V).
MNIS	Mean-shifted IS based on norm minimization.
MPFP	Most probable failure point (design point).
P_f	Per-cell failure probability.
PVT	Process, voltage, and temperature.
r	Anchored residual, in units of σ.
s_j	Slope $\lvert\nabla m_j\rvert$ of lobe j at its design point.
SuS	Subset simulation.
$\mathbf{u}_j$	Unit failure direction of lobe j.
V_{min}	Minimum supply meeting a target P_f.
$\mathbf{x}$	Standardized mismatch vector, $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$.
α_1, α_2, α_c	Risk shares of the conformal limit, IS escalation, and IS calibration labels.
n_{cal}	Number of calibration conditions.
q	Conformal margin on $\log_{10}P_f$.
β_j	Reliability index of lobe j.
$\hat{\beta}$, $\hat{u}$, $\hat{s}$	Values transferred from the training conditions.
δ	Escalation threshold on r.
θ, θ_{95}	Direction error and its calibrated 95th percentile.
ρ	Correlation of the two lobes.
Φ, Φ_2	Standard normal and bivariate normal distribution functions.

Manuscript received [date]; revised [date]; accepted [date]. Date of publication [date]; date of current version [date]. This work did not receive external funding. Associate Editor: [assigned by the journal]. *(Corresponding author: Nabanita Sarker.)*

Nabanita Sarker is with the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1205, Bangladesh (e-mail: 2006162@eee.buet.ac.bd).

This article has supplementary downloadable material available at [DOI link assigned by IEEE], provided by the author: the simulation scripts, model conversions, and result files needed to reproduce every table and figure.

Digital Object Identifier [assigned by IEEE]

I. INTRODUCTION

Embedded static random-access memory (SRAM) occupies a large fraction of the area of modern systems on chip, and its yield is dominated by rare parametric failures of individual bitcells. The failures are caused by local mismatch between nominally identical transistors, which arises from random dopant fluctuation, line-edge roughness, and gate-stack variation [1], [2] and grows as devices shrink [3], [4]. Silicon measurements on millions of cells confirm that the cell stability metrics follow the statistics of this mismatch [5]. Because an array holds millions of cells, the per-cell failure probability P_f that matters for yield lies between 10^{-6} and 10^{-12}, five to seven standard deviations into the tail of the cell metric [6].

Plain Monte Carlo (MC) simulation needs on the order of $100/P_f$ samples for a 10 % relative error, which is out of reach at these probabilities. A large literature therefore develops rare-event estimators for SRAM: statistical blockade [7], mixture and mean-shifted importance sampling (IS) [8]–[11], gradient-based and flow-based proposals [12]–[14], non-Monte-Carlo boundary methods [15], scaled-sigma and full-

region sampling [16], [17], and, adopted from structural reliability, subset simulation (SuS) [18] and active-learning Kriging [19], [20].

P_f must hold at every process, voltage, and temperature (PVT) condition, and it changes by many orders of magnitude across them. Repeating a rare-event estimate over a dense grid of corners, supplies, and temperatures is expensive, and the direction in which the tail moves is not obvious: in nanometer CMOS the temperature dependence of drive current can reverse at low supply [21], [22]. A common shortcut is to characterize a subset of conditions and extrapolate the rest with a regression or Gaussian-process (GP) surrogate, the family of models used in Bayesian, surrogate-based, and multi-corner yield analysis [23]–[27]. Three questions are open: whether such extrapolation is safe in the far tail, where yield is decided; if not, what the cheapest estimate at a new condition is that remains safe; and how a designer can attach a stated risk to the decision that a target is met without running a rare-event simulation at every condition.

This paper addresses all three with SPICE-level IS ground truth at every one of the 242 unseen conditions of a SkyWater SKY130 foundry bitcell with its foundry mismatch model and of an ASAP7 7-nm FinFET cell, for read and write failures, cross-checked by plain MC and, at every condition below 10^{-7}, by a second, independent IS proposal. Earlier cross-PVT and multi-corner yield methods return an estimate whose uncertainty, if stated, rests on the surrogate model; the procedure proposed here returns a sign-off decision whose risk of unsafe acceptance is bounded, under exchangeability, by calibration against IS, at a known SPICE cost, and we evaluate it down to the yield of a memory macro. The contributions are as follows.

1) *A diagnosis of why moment surrogates fail.* A Gaussian tail fitted to the *exact* margin moments at the same condition already errs by up to 11.9 decades in the design-relevant range, so the error of moment surrogates is caused by the tail shape of the minimum of two failure lobes, not by extrapolation; an analytical example shows why.

2) *Evidence on extrapolation in reliability-index space.* Modeling the per-lobe reliability index and lobe correlation removes most of the error for the FinFET cell, but leaves errors of up to 1.9 decades, in the unsafe direction, for the planar foundry cell, which we trace to a rotation of the failure direction across PVT.

3) *Cross-PVT design-point transfer with a characterized correction error.* The design point of each lobe is predicted across PVT from the training conditions and corrected with one SPICE run per lobe at the new condition. FORM parametric sensitivity [28], [29] gives the derivative of β with respect to a parameter at a design point that has already been found, and single-loop reliability-based design [30] reuses the previous iterate's design point inside an optimization loop; neither predicts the design point at a new operating condition or quantifies the error of doing so. We show that the anchored correction removes the first-order error exactly for a locally linear lobe and leaves a product of prediction errors plus a one-sided, second-order direction term (Proposition 1), which we verify against FORM and IS.

4) *A statistically calibrated sign-off.* We show that a deterministic upper bound on P_f requires convexity that cannot be verified and that finite-sample IS bounds are uninformative at these probabilities, and we instead calibrate the transferred estimate with split-conformal margins [31], [32] computed from IS on a random subset of the sign-off conditions; conditions that the margin cannot decide are escalated to IS. Under exchangeability, the probability of an unsafe acceptance at a randomly selected condition, equivalently the expected fraction of unsafe acceptances, is bounded (Proposition 2); there is no simultaneous guarantee over the grid. Conformal calibration of failure-probability estimates has been proposed for Kriging-based structural reliability [33]; to our knowledge it has not been applied to cross-PVT circuit sign-off.

5) *A full-grid, independent evaluation* against IS at every unseen condition, including mean-shifted IS [9], SuS [18], AK-IS [20], low-budget IS at the two design points, the one-step and full FORM estimators, the unsafe-acceptance rate of sign-off decisions, the interpolated V_{min}, and the yield delivered by a 1-Mb macro signed off with each estimator.

Section II reviews related work. Section III formulates the problem. Section IV describes the estimators and the calibrated sign-off, Section V the setup, and Section VI the results. Section VII discusses the scope of the guarantee, a sign-off workflow, threats to validity, and limitations. Section VIII concludes.

II. Related Work

A. SRAM Stability Metrics

The static noise margin (SNM) obtained from the butterfly plot of the cross-coupled inverters [34] is the standard static measure of read and hold stability. Calhoun and Chandrakasan observed that the cell SNM is the smaller of two lobe margins and modeled it as the minimum of two independent variables [35]. The N-curve offers current-based alternatives for read stability and write-ability [36], several static and dynamic write margins have been compared [37], and dynamic stability measured with finite word-line pulses can be less pessimistic than static metrics [38]. Hold stability determines the data-retention voltage [39], and read and write assists trade the margins against each other [40]. We use signed butterfly margins for read and write because they are smooth functions of the mismatch, so they can serve as limit-state functions.

B. Rare-Event Estimation for SRAM

Mukhopadhyay *et al.* related read, write, access, and hold failures to device variation [6]. Statistical blockade filters samples with a classifier and fits a generalized Pareto tail [7]. Kanj *et al.* introduced mixture IS for SRAM [8], Dolecek *et al.* shifted the sampling density to a minimum-norm failure point [9], sequential IS adapts the proposal in high dimensions [10], and Wang *et al.* locate several failure regions by multistart optimization before sampling [11]. Gradient IS finds

the most probable failure point (MPFP) by local optimization [12], and normalizing-flow proposals learn the failure region [13], [14]. Gu and Roychowdhury estimated the failure volume without MC [15], scaled-sigma sampling extrapolates from inflated-variance runs [16], and REscope covers disjoint failure regions [17]. These estimators target one operating condition at a time and pay their full cost again at every condition.

C. Surrogate and Multi-Corner Yield Analysis

Surrogate models reduce the number of SPICE calls by learning the circuit response: adaptive online surrogates [23], low-rank tensor meta-models [24], and Bayesian yield analysis with active learning [25]. Across corners, correlated Bayesian model fusion encodes the similarity of the models of different corners in a shared prior [26], and learned priors share information across many corners [27]. OpenYield provides an open benchmark with realistic second-order effects [41]. These methods model a performance or its distribution and evaluate accuracy at the conditions where they are trained or refined; the tail-shape question raised here concerns any method that reads P_f from modeled moments.

D. Structural Reliability and Worst-Case Distances

The first-order reliability method (FORM) linearizes a limit-state function at its MPFP, located by the Hasofer–Lind/Rackwitz–Fiessler (HL-RF) iteration [42], [43], and second-order corrections account for curvature [44]. In circuit design the same quantity appears as the worst-case distance, used together with worst-case operating conditions for yield analysis and design centering [45], [46]. Systems with several failure modes are series or parallel systems of limit states [47], [48]. The sensitivity of the reliability index to a deterministic parameter follows from the limit-state value at the design point without re-solving the optimization [28], [29], and single-loop reliability-based design reuses the design point of the previous design iterate to avoid a nested reliability analysis [30]. SuS writes a small probability as a product of larger conditional probabilities [18], [49]; AK-MCS and AK-IS learn the limit state with Kriging [19], [20], [50], [51]; and defensive mixtures make IS robust to a misplaced proposal [52], [53]. The anchored correction (8) is a finite-difference form of the parametric sensitivity; what differs is that the design point at the new condition is not solved for but predicted across PVT, so the error of that prediction, which the sensitivity analysis does not address, must be characterized and controlled (Sections IV-D, IV-E, and IV-G).

E. Error Bounds and Certification

Rigorous bounds on a failure probability need structural assumptions. Error bounds for approximate rare-event models require a convex failure domain [54], and the classical statement that $\Phi(-\beta)$ bounds P_f from above likewise assumes convexity and the exact design point [47], [48]. Sampling estimators admit finite-sample confidence bounds when the likelihood ratio is bounded, e.g., by empirical Bernstein inequalities [55], but their range term scales with the bound on the weights rather than with P_f. Conformal prediction gives distribution-free, finite-sample coverage of a predictor's error from a calibration set that is exchangeable with the test points [31], [32], [56], and it has recently been combined with active-learning Kriging to certify failure-probability estimates in structural reliability [33]. We use split conformal calibration in a different role: to calibrate a physics-based cross-PVT estimator and the sign-off decisions and $V_{\min}$ built on it.

F. Positioning of This Work

Table I compares the method families from the viewpoint of a PVT sweep, where the relevant cost is the number of *new* SPICE runs at each additional condition. Sampling estimators are unbiased or nearly so but pay their cost again at every condition. Worst-case-distance and FORM analyses need tens to a hundred runs per condition. Surrogates need none, but read P_f from modeled moments. The proposed transfer spends two runs per new condition, and its calibration spends IS on a small random subset of the sign-off conditions instead of on every condition.

TABLE I
METHOD FAMILIES FOR SRAM FAILURE PROBABILITY ACROSS PVT

Family	Examples	Cost at a new condition	Training data	Tail representation	Runs here
Mixture / mean-shift IS	[8], [9], [11]	Samples to reach target error	No	Sampled	9.5k–23k
Gradient / flow IS	[12]–[14]	Optimization + samples	No	Sampled	—
Subset simulation	[18], [49]	Levels × samples	No	Sampled	2.8k–10.9k
Active-learning Kriging	[19], [20], [25]	Added points + fits	No	Learned limit state	275
Worst-case distance, FORM	[45], [46]	Gradient iterations	No	Linearized limit state	39–93
Moment surrogates, multi-corner	[23], [26], [27]	None if extrapolated	Yes	Modeled moments	0
Conformal Kriging (structural)	[33]	Added points + fits	No	Learned limit state	—
Transfer (proposed)	Sec. IV-C–E	Two anchored runs	Yes	Two-lobe series system	2
Calibrated sign-off (proposed)	Sec. IV-G	Two runs; IS if undecided	Yes + IS on a random subset	Conformal margin	$\approx 10^3$ incl. calibration

Runs here: SPICE simulations per condition measured in Section VI for d = 12, including amortized calibration for the calibrated sign-off; —: not implemented.

III. PROBLEM FORMULATION

A. Mismatch Model and Limit States

Let $\mathbf{x} \in \mathbb{R}^d$ collect the standardized local-mismatch variables of the six transistors of a 6T cell, $\mathbf{x} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$. A global PVT condition $\mathbf{c} = (T, V_{DD}, \text{corner})$ is fixed for each analysis. For a symmetric cell, every static failure mode has two mirror limit states, one for each stored value or write direction, with signed margins $m_0(\mathbf{x}; \mathbf{c})$ and $m_1(\mathbf{x}; \mathbf{c})$. Failure occurs when either margin is negative, so the cell is a series system with failure probability

$$P_f(\mathbf{c}) = \mathrm{P}[\ \min\{m_0(\mathbf{x}; \mathbf{c}), m_1(\mathbf{x}; \mathbf{c})\} < 0\ ]. \tag{1}$$

Fig. 1 shows the cell. The pull-up (PU), pull-down (PD), and pass-gate (PG) transistors of each half carry independent mismatch, with two variables per transistor in this work (d = 12). Lobe 0 concerns the stored state Q = 1 and lobe 1 the state Q = 0. For read, each lobe is dominated by the pass-gate and pull-down of the node that stores 0, which are disturbed by the precharged bit line, and by the trip point of the opposite inverter, so the two lobes weight the mirror devices oppositely and are correlated through the shared feedback.

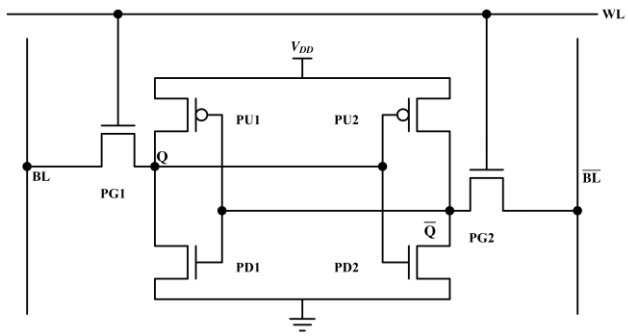

Fig. 1. 6T SRAM cell. Local mismatch is applied independently to the six transistors PU1, PU2 (pull-up), PD1, PD2 (pull-down), and PG1, PG2 (pass-gate). For the margins, the cell is opened into half-cells, each an inverter loaded by its pass-gate, so that every butterfly curve is obtained from a DC sweep of independent half-cells.

B. Signed Read and Write Margins

The read margin uses two half-cells, each an inverter loaded by its access transistor with word line and bit line at V_{DD}. Their voltage-transfer curves form the butterfly plot (Fig. 2(a)). After rotating the plane by 45°, $u = (x-y)/\sqrt{2}$ and $v = (x+y)/\sqrt{2}$, the vertical distance $\Delta v(u)$ between the curves is evaluated separately on each half of the rotated axis, and the lobe margin is

$$m_j = (1/\sqrt{2}) \cdot \max_{u \in half\,j} s_j \, \Delta v(u), \quad s_0 = +1, \; s_1 = -1, \qquad (2)$$

which equals the side of the largest square inscribed in lobe j when the lobe exists and becomes negative when the lobe vanishes. Measuring each lobe on its own half is essential: an extremum taken over the whole axis picks up millivolt slivers of opposite sign near the supply rails and hides failing cells. The write margin uses the same construction with the bit line of the node being written at 0 V: a write succeeds when the lobe of the old state has vanished, so the write margin is the negated old-state lobe of the write butterfly (Fig. 2(b)). Both metrics come from independent half-cells driven by one swept source, so one DC sweep evaluates hundreds of cells and no bistable circuit is simulated. We use static margins because they are the common first-pass stability metric and their statistics have been confirmed on large silicon arrays [5]; they are a deliberate choice, not a requirement of the method. The transfer, the anchored correction, and the calibration act on any scalar margin with two mirror lobes, so a dynamic metric [37], [38] changes the cost of each margin evaluation but not the procedure (Section VII-D).

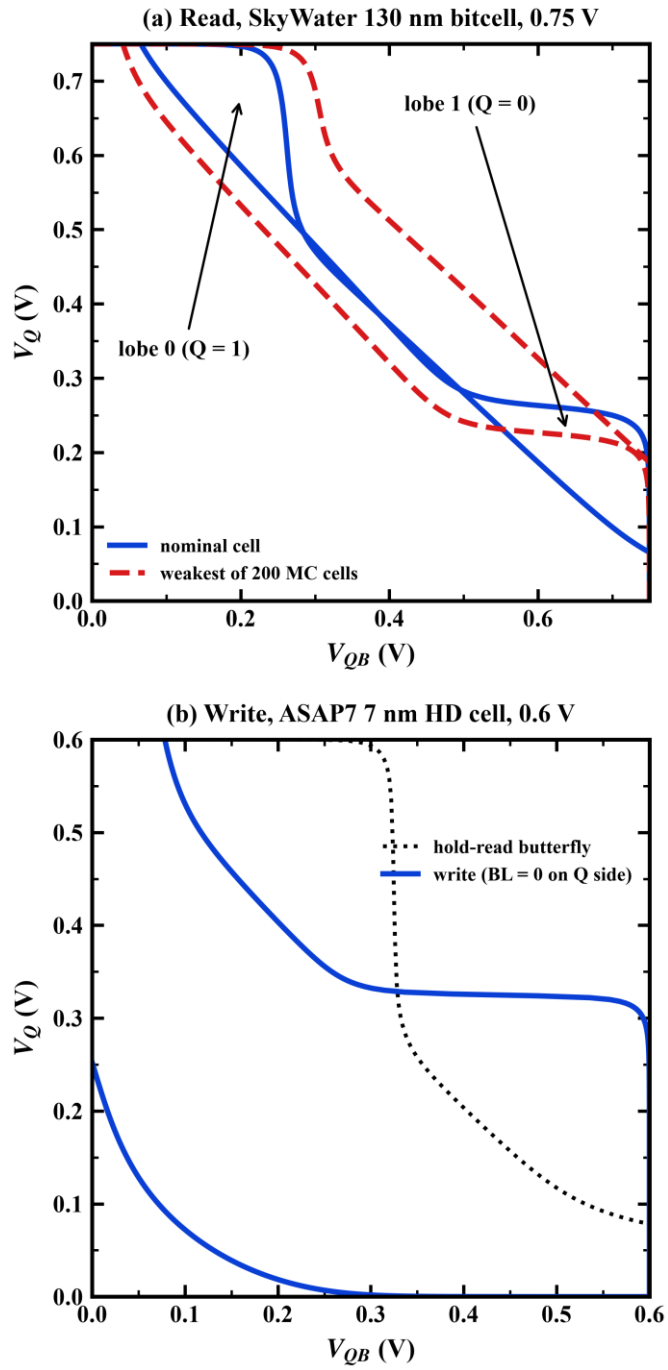

Fig. 2. Butterfly curves. (a) Read, SkyWater 130-nm foundry bitcell at 0.75 V, 27 °C: nominal cell and the weakest of 200 Monte Carlo cells, whose lobe 1 has nearly collapsed. (b) Write of a 0 into node Q for the ASAP7 7-nm high-density cell at 0.6 V: the write butterfly (solid) has a single intersection, so the old-state lobe has vanished and the write succeeds; the read butterfly (dotted) is shown for reference.

C. Extrapolation Task

Conditions are split into a *training box* of mild conditions (moderate temperature and supply) and the *unseen* conditions outside it. A method may use any amount of simulation at training conditions; at an unseen condition its cost is the number of new SPICE runs. Accuracy is measured against IS truth at unseen conditions only, and safety by the sign of the error: an underestimate of P_f is unsafe.

D. Why Moment-Based Models Fail in the Tail

Consider two lobes that are exactly linear in $\mathbf{x}$, each distributed as $N(a, s^2)$, with correlation ρ and reliability index $\beta = a/s$. The exact failure probability of their minimum is

$$P_f = 2\Phi(-\beta) - \Phi_2(-\beta, -\beta; \rho), \quad (3)$$

where Φ_2 is the bivariate standard normal distribution function. The first two moments of the minimum are known in closed form [57]:

$$E[\min] = a - s\sqrt{(1-\rho)/\pi}, \quad Var[\min] = s^2[1 - (1-\rho)/\pi]. \quad (4)$$

A Gaussian fitted to these *exact* moments gives $\Phi(-E[\min]/\sqrt{Var[\min]})$. Fig. 3 shows its error relative to (3). For the negative lobe correlation typical of read margins the Gaussian tail underestimates P_f by 3.1 decades at $\beta = 5$ and by 7.3 decades at $\beta = 7$ ($\rho = -0.6$); for independent lobes the errors are 1.2 and 2.9 decades, and only strongly positively correlated lobes give an accurate Gaussian tail. The minimum of two lobes has a heavier lower tail than any Gaussian with the same moments, and the discrepancy grows with depth. Any method that reads P_f from modeled moments inherits this error, however well it predicts the moments.

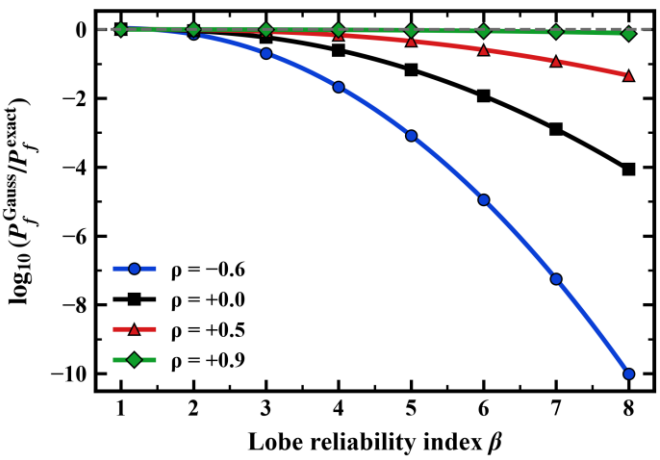

Fig. 3. Error of a Gaussian tail fitted to the exact mean and variance of the minimum of two equal, linear, correlated lobes, relative to the exact series-system probability (3), versus the lobe reliability index.

IV. Estimators

A. Moment Surrogates

Two surrogates are fitted to the mean μ and the log standard deviation $\log \sigma$ of the cell margin $\min(m_0, m_1)$ over the training box: a quadratic regression in (T, V_{DD}) with corner indicators and their interactions, and a GP [58] with an anisotropic squared-exponential plus linear kernel [59]. The failure probability follows from the Gaussian tail,

$$\hat{P}_f = \Phi(-\hat{\mu}(\mathbf{c})/\hat{\sigma}(\mathbf{c})). \quad (5)$$

Both use the same 1000 common-random-number cells at every training condition, which removes sampling noise from the condition-to-condition differences and favors the surrogates.

B. Reliability-Index-Space Surrogate

The tail-shape error of (5) is avoided by modeling the quantities that the series system needs. For jointly Gaussian (linearized) lobes with reliability indices β_0, β_1 and correlation ρ,

$$P_f = \Phi(-\beta_0) + \Phi(-\beta_1) - \Phi_2(-\beta_0, -\beta_1; \rho). \quad (6)$$

The *β-space surrogate* fits $\beta_0(\mathbf{c})$, $\beta_1(\mathbf{c})$, and $\operatorname{atanh} \rho(\mathbf{c})$ over the training box with the same GP family as above, using FORM at the training conditions as targets, and evaluates (6) at the unseen condition with no new runs. It is the strongest zero-run competitor we found, and it isolates the part of the error that is caused by extrapolation itself.

C. Design-Point Transfer

At a training condition, FORM gives for each lobe the design point $\mathbf{x}_j^* = \beta_j \mathbf{u}_j$, where $\mathbf{u}_j = -\nabla m_j/|\nabla m_j|$ is the unit failure direction, and the slope $s_j = |\nabla m_j|$ there. Near the design point a lobe is approximated by

$$m_j(\mathbf{x}) \approx s_j\,(\beta_j - \mathbf{u}_j^T\mathbf{x}). \quad (7)$$

The transfer fits GPs over the PVT features to β_j, $\log s_j$, and the components of $\mathbf{u}_j$ (renormalized after prediction), giving $\hat{\beta}_j(\mathbf{c})$, $\hat{s}_j(\mathbf{c})$, and $\hat{u}_j(\mathbf{c})$ at an unseen condition, and to $\operatorname{atanh} \rho$ for the correlation.

D. Anchored Correction and Its Error

At the unseen condition, each lobe is simulated once at its predicted design point $\hat{\mathbf{x}}_j = \hat{\beta}_j \hat{u}_j$; both lobes are evaluated in a single SPICE call. The anchored index is

$$\tilde{\beta}_j = \hat{\beta}_j + m_j(\hat{\mathbf{x}}_j) / \hat{s}_j. \quad (8)$$

Proposition 1: If the lobe is locally linear as in (7) and $\cos\theta = \mathbf{u}^T\hat{u}$, then

$$\tilde{\beta} - \beta = (\hat{\beta} - \beta)(1 - s/\hat{s}) + (s/\hat{s})\,\hat{\beta}\,(1 - \cos\theta). \quad (9)$$

Proof: By (7), $m(\hat{\mathbf{x}}) = s(\beta - \hat{\beta}\cos\theta)$. Substituting into (8) and rearranging gives (9). ■

Three consequences follow. First, the error of the transferred index, $\tilde{\beta} - \beta$, enters only multiplied by the relative slope error, so a single run removes it to first order; this is the finite-difference form of the parametric FORM sensitivity [28], [29]. Second, the direction error enters at second order, $\theta^2\hat{\beta}/2$, and is *one-sided*: when the slope is predicted well it always makes $\tilde{\beta}$ too large, i.e., the estimate unsafe. Third, the residual $r_j = |m_j(\hat{\mathbf{x}}_j)|/\hat{s}_j = (s/\hat{s})|\beta - \hat{\beta}\cos\theta|$ is an observable estimate of the first-order prediction error, available at no extra cost.

E. Empirical Guard and Escalation

The direction term cannot be observed from the anchored run, so it is bounded instead. Its size is estimated *inside* the training box by block cross-validation that imitates extrapolation: each edge block of the box (the hottest and the coldest temperature, the lowest and the highest supply) is held out in turn, the direction GP is fitted to the rest, and the angle between predicted and FORM directions is recorded. With θ_{95} the 95th percentile of these angles, the guarded index is

$$\beta_j^g = \tilde{\beta}_j - \hat{\beta}_j\,(1 - \cos\theta_{95}), \quad (10)$$

which, by (9), is conservative whenever $\theta \le \theta_{95}$ and the slope term is small. If the residual $r = \max_j r_j$ exceeds a threshold δ (0.25 by default), the transferred design point is too far from the actual one for (9) to hold and the condition is escalated to the one-step estimator of Section IV-F. Algorithm 1 and Fig. 4 summarize the procedure. The training-box

quantities are computed once; each new condition costs two runs, plus $3(d+1)$ when escalated.

Algorithm 1: Guarded design-point transfer at an unseen condition c

Once: FORM per lobe at the training conditions; GPs for β_j, $\log s_j$, $\mathbf{u}_j$, atanh ρ; θ_{95} by edge-block cross-validation.

1. Predict $\hat{\beta}_j$, $\hat{s}_j$, $\hat{u}_j$, $\hat{\rho}$ at $\mathbf{c}$; set $\hat{\mathbf{x}}_j = \hat{\beta}_j\hat{u}_j$.
2. Simulate $m_0(\hat{\mathbf{x}}_0)$ and $m_1(\hat{\mathbf{x}}_1)$ (one SPICE call, two cells).
3. $\tilde{\beta}_j \leftarrow \hat{\beta}_j + m_j(\hat{\mathbf{x}}_j)/\hat{s}_j$; $r \leftarrow \max_j |m_j(\hat{\mathbf{x}}_j)|/\hat{s}_j$.
4. If $r > \delta$: return the one-step estimate $(3(d+1)$ runs).
5. $\beta_j^g \leftarrow \tilde{\beta}_j - \hat{\beta}_j(1 - \cos\theta_{95})$; return (6) with $(\beta_0^g, \beta_1^g, \hat{\rho})$.

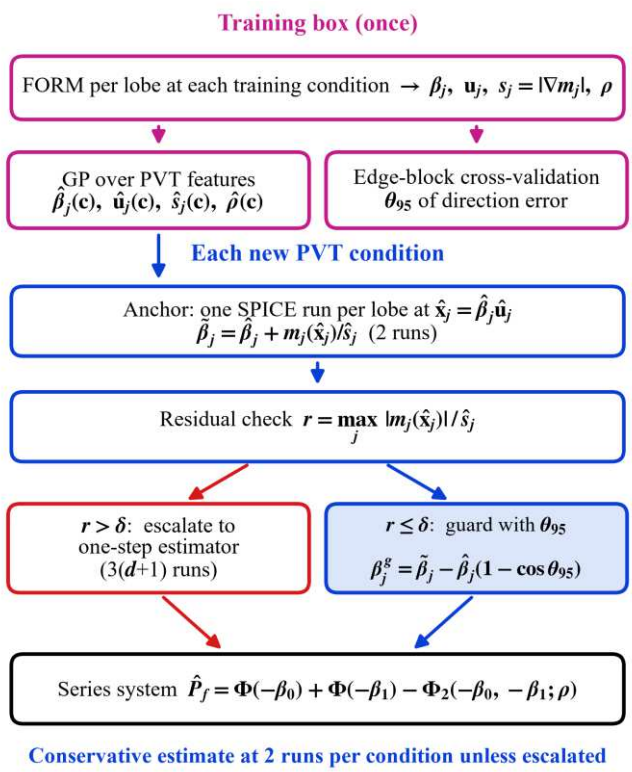

Fig. 4. Guarded design-point transfer (Algorithm 1). The training box is analyzed once; each new condition costs two SPICE runs (one per lobe), plus $3(d+1)$ runs if the anchored residual exceeds δ.

F. Reference and Baseline Estimators

One-step estimator. Each lobe is linearized at the origin by forward differences of step $h = 0.25\sigma$ ($d+1$ runs, one SPICE call), re-linearized once at the predicted design point $\mathbf{x}_j^{(1)} = -a_j\mathbf{g}_j/|\mathbf{g}_j|^2$, and β_j is taken from the new tangent plane; ρ follows from the corrected normals. This is one HL-RF step per lobe [42], [43] at a cost of $3(d+1) = 39$ runs. *FORM* iterates HL-RF to $|\Delta\beta| < 10^{-3}$ (at most ten iterations).

IS truth. Ground truth uses a defensive mixture proposal [52], [53] with weight 0.1 on $N(\mathbf{0}, \mathbf{I})$ and 0.45 on a Gaussian at each linearized design point. A first stage of 1000 SPICE samples re-centers each component on the weighted mean of the failing samples of its lobe, and an independent second stage of 4000 samples gives

$$\hat{P}_f = (1/N)\, \Sigma_i\, 1[\min_j m_j(\mathbf{x}_i) < 0]\, \varphi(\mathbf{x}_i)/q(\mathbf{x}_i). \qquad (11)$$

Because the proposal is a mixture over both lobes, includes the nominal density, and is re-centered on SPICE failures, it is unbiased for any lobe geometry; its accuracy is checked against plain MC in Section VI-A.

Sampling baselines. Mean-shifted IS with norm minimization (MNIS) [9] widens a Gaussian until ten failures are found, shifts to the minimum-norm failure, and samples to a 10 % relative error (cap 20 000); SuS [18] uses 1000 samples per level, $p_0 = 0.1$, and modified Metropolis [49]; AK-IS [20] uses a GP limit state, a 10^5-point population around the two linearized design points, U-learning, and a cap of 250 added points. The strongest sampling baseline is *MPFP-mixture IS*: after one linearization ($d+1$ runs), N samples from the same defensive mixture without re-centering, at a total of $d+1+N$ runs. Every simulated cell counts as one run.

G. Calibrated Sign-Off

Sign-off asks, for each condition $\mathbf{c}$ in a set $\mathcal{S}$ and a target P^* (e.g., 10^{-6} or 10^{-9}), whether $P_f(\mathbf{c}) \le P^*$; accepting a condition whose true P_f exceeds P^* is an *unsafe acceptance*. We first consider what can be guaranteed. *Deterministic bounds.* If the failure domain of lobe j is convex and β_j is its exact distance from the origin, the supporting hyperplane at the design point contains the domain on its far side, so $P_f \le \Phi(-\beta_0) + \Phi(-\beta_1) - \Phi_2(-\beta_0, -\beta_1; \rho)$ holds exactly. Convexity cannot be verified from a finite number of SPICE runs, the write lobes are visibly curved, and an estimated β that is too large invalidates the bound, so this route gives no guarantee in practice. *Finite-sample sampling bounds.* With the defensive mixture of Section IV-F the likelihood ratio is bounded by 10, and an empirical Bernstein bound [55] on the IS mean is rigorous, but its range term alone is 1.7×10^{-2} for 5000 samples and 1.7×10^{-3} for 50 000, far above the targets. *Statistical calibration.* What can be guaranteed without structural assumptions is a coverage statement for a randomly selected condition. Let a calibration set $\mathcal{C}$ of n_{cal} conditions be drawn uniformly at random from $\mathcal{S}$, and let IS be run at each of them by the same procedure, independently across conditions. The *label* of a condition is $Y = \log_{10}P_f^U$, its one-sided IS upper confidence limit at level $1 - \alpha_c$, and its score is $e = Y - \log_{10}\hat{P}_f$, with $\hat{P}_f$ the transferred estimate (8), (6). With q the $\lceil(n_{cal}+1)(1 - \alpha_1)\rceil$-th smallest calibration score, the conformal upper limit at a remaining condition is $\log_{10}\hat{P}_f + q$; the lower limit uses the $\lfloor(n_{cal}+1)\alpha_1\rfloor$-th smallest score computed with the IS lower confidence limits.

Proposition 2: Let $\mathbf{c}$ be drawn uniformly from $\mathcal{S} \setminus \mathcal{C}$, so that the scores of $\mathcal{C} \cup \{\mathbf{c}\}$ are exchangeable. (i) $P[Y(\mathbf{c}) > \log_{10}\hat{P}_f(\mathbf{c}) + q] \le \alpha_1$. (ii) If the IS upper limit at $\mathbf{c}$ covers the true P_f with probability at least $1 - \alpha_c$, then $P[\log_{10}P_f(\mathbf{c}) > \log_{10}\hat{P}_f(\mathbf{c}) + q] \le \alpha_1 + \alpha_c$. (iii) The rule that accepts $\mathbf{c}$ when the conformal upper limit is $\le P^*$, rejects it when the conformal lower limit is $> P^*$, and otherwise runs IS and accepts only if the IS upper limit at level $1 - \alpha_2$ is $\le P^*$, accepts $\mathbf{c}$ unsafely with probability at most $\alpha_1 + \alpha_c + \alpha_2$.

Proof: (i) is the split-conformal guarantee [31], [32]: by exchangeability the rank of the test score among the $n_{cal}+1$ scores is uniform, so it exceeds the $\lceil(n_{cal}+1)(1 - \alpha_1)\rceil$-th smallest calibration score with probability at most α_1. (ii) The

event $\log_{10}P_f > \log_{10}\hat{P}_f + q$ is contained in the union of $\{Y > \log_{10}\hat{P}_f + q\}$ and $\{\log_{10}P_f > Y\}$, whose probabilities are at most α_1 and α_c. (iii) An unsafe acceptance requires that the true P_f exceed either the conformal upper limit or the IS upper limit at level $1 - \alpha_2$; the union bound gives the sum. ■

Scope of the guarantee. The probabilities are taken over the random calibration set, the random choice of the test condition, and the IS randomness. By linearity of expectation, the bound therefore also limits the expected *fraction* of unsafe acceptances among the conditions of $\mathcal{S}$. It is *not* a simultaneous statement: the probability that at least one condition of the grid is accepted unsafely can be much larger, up to $|\mathcal{S}|$ times the bound in the worst case, and a simultaneous guarantee would require per-condition risks of the order of $\alpha/|\mathcal{S}|$ and hence more calibration runs than running IS everywhere. Part (i) is exact for finite samples; parts (ii) and (iii) rely on the IS confidence limits, which are asymptotic (central limit theorem) and require well-behaved weights, which we monitor through the failure-weight effective sample size. The guarantee does not extend to conditions outside $\mathcal{S}$. A finite quantile requires $n_{cal} \geq \lceil 1/\alpha_1 \rceil - 1$, i.e., 19 calibration conditions for $\alpha_1 = 0.05$. In the main configuration $\alpha_1 = \alpha_2 = \alpha_c = 0.05$, so the nominal bound is 0.15.

H. Cost Accounting

Plain MC needs about $100/P_f$ runs; IS and SuS need hundreds to tens of thousands, paid at every condition; FORM needs one gradient (d+1 runs) per HL-RF iteration and lobe. The transfer pays FORM at the training conditions once (7,150 runs for the 100 SkyWater training conditions) and then two runs per new condition, plus 39 at escalated conditions. The calibrated sign-off adds IS at the calibration conditions and at the undecided ones; we charge each IS estimate at the cost of the recommended FORM-centered procedure, i.e., converged FORM at that condition plus 8000 samples. Table II reports the total number of SPICE runs needed to sign off every unseen condition, including training, calibration, and escalation. On grids of the size studied here the guarded transfer costs about as much as converged FORM at every condition (7.8k versus 7.7k runs for SkyWater) and more than the one-step estimator; its benefit on these grids is safety on the planar cell, where FORM-class estimators are sometimes unsafe (Section VI-E), not cost. A cost advantage arises only for denser sweeps or several design variants sharing one training box, since each additional condition costs two runs instead of 39 or 65–93. The calibrated sign-off is 3.1 to 4.9 times cheaper than IS at every condition but more than an order of magnitude more expensive than FORM everywhere.

V. Experimental Setup

All simulations use ngspice 47 [60]. Table III summarizes the studies. For SkyWater 130 nm [61], the foundry SRAM devices special_nfet_latch (pull-down, 0.21/0.15 µm), special_nfet_pass (access, 0.14/0.15 µm), and special_pfet_pass (pull-up, 0.14/0.15 µm) are used with the foundry corner libraries; the PDK's mismatch model is converted to per-instance parameters (Appendix A). For ASAP7 [62], the BSIM-CMG [63] cards are run through ngspice's OSDI interface. The kit has no statistical model, so mismatch is an assumption: per-fin $\sigma_{Vth} = A_{VT}/\sqrt{(W_{eff}L)} = 31$ mV with $A_{VT} = 1.2$ mV·µm [3] and $W_{eff} = 2H_{fin} + T_{fin}$, and a 4 % mobility spread, both scaling as $1/\sqrt{n_{fin}}$; Section VI-I repeats the study at 20 and 45 mV. The high-density cell with one fin per device is used. At every condition the same 1000 mismatch vectors are simulated for the moment surrogates. All experiments ran on an eight-core laptop.

TABLE II
TOTAL SPICE RUNS TO SIGN OFF ALL UNSEEN CONDITIONS (RUNS PER CONDITION)

Study	Cond.	IS everywhere	FORM everywhere	One-step everywhere	Guarded transfer	Calibrated sign-off
SkyWater 130 nm, read	110	887.7k (8070)	7.7k (70)	4.3k (39)	7.8k (71)	180.9k (1645)
ASAP7 7 nm, read	66	532.3k (8065)	4.3k (65)	2.6k (39)	4.0k (61)	171.4k (2598)
ASAP7 7 nm, write	66	534.2k (8093)	6.2k (93)	2.6k (39)	5.7k (87)	171.9k (2605)

IS: converged FORM plus 8000 FORM-centered samples per condition. Guarded transfer: training FORM plus two runs per condition plus 39 per escalated condition. Calibrated sign-off: training FORM plus calibration IS, two runs per condition, and IS at undecided conditions (mean over the two targets and 500 splits).

TABLE III
TECHNOLOGY STUDIES

Study	Mismatch	d	V_{DD} (V)	Corners	Train / unseen
SKY130 bitcell, read	Foundry (per instance)	12	0.75–1.2	TT/FF/SS/FS/SF	100 / 110
ASAP7 FinFET, read	Assumed, derived	12	0.35–0.7	TT/FF/SS	60 / 66
ASAP7 FinFET, write	Assumed, derived	12	0.55–0.8	TT/FF/SS	60 / 66

Seven temperatures from −40 to 125 °C. Training box: $T \leq 85$ °C and $V_{DD} \geq$ 0.9 V (SKY130), 0.45 V (ASAP7 read), 0.65 V (ASAP7 write).

IS truth campaign. IS was run at every unseen condition (242 in total: 110 SkyWater, 66 ASAP7 read, 66 ASAP7 write) with the estimator of Section IV-F (proposal A), with seeds disjoint from all other runs and with the failure-weight effective sample size (ESS) recorded. Every condition with $P_f < 10^{-7}$ was then re-estimated independently with proposal B, a defensive mixture centered at the converged FORM design points (8000 samples). The two proposals agreed within two standard errors at 59 of 63 conditions and disagreed by more than three at 3 deep write conditions, in every case with proposal A too low (by 0.07, 0.17, 0.42 decades); at four of the deepest write conditions proposal A also kept an ESS below 50 even when re-run with 18 000 samples. Proposal A misplaces its mixture on the most curved write lobes. At all six of these conditions proposal B was run twice with independent seeds (8000 and 16 000 samples), and the two B estimates agreed within 1.8 standard errors with ESS of at least 750. The canonical truth therefore uses the inverse-variance combination of the B runs wherever B is available and proposal A elsewhere (Appendix B); the smallest B ESS is 266, and relative standard errors are 0.5–12 %.

VI. Results

A. Validation of the IS Truth

Table IV compares the IS truth with plain MC at the shallowest truth point of each study and at a point near 10^{-3}, where MC with up to 200 000 cells is affordable. Over the 5 comparisons the difference is at most 0.09 decade; two comparisons differ by about two standard errors (largest $|z|$ = 2.1), in opposite directions, so there is no sign of a systematic bias. At the SkyWater point with z = +2.1, three further IS runs with fresh seeds gave 5.18e-04, 5.39e-04, 5.31e-04, and a second 200 000-cell MC run brought the pooled MC estimate to 4.80e-04 ± 3.5e-05 (z = +1.6 for the mean of the four IS runs). Together with the two-proposal cross-check of Section V, we treat the IS truth as accurate to about ±0.1 decade, which is also the threshold used below to call an error unsafe.

Table IV
Plain MC Versus IS at Moderate Probabilities

Study	Condition	MC cells	MC	IS	z
SkyWater 130 nm, read	125 °C, 1 V, SF	20,000	3.7×10^{-2} ± 4 %	3.6×10^{-2} ± 3 %	−0.6
SkyWater 130 nm, read	125 °C, 1.2 V, SF	200,000	4.6×10^{-4} ± 10 %	5.6×10^{-4} ± 4 %	+2.1
ASAP7 7 nm, read	−40 °C, 0.35 V, FF	20,000	7.0×10^{-4} ± 27 %	6.2×10^{-4} ± 3 %	−0.4
ASAP7 7 nm, write	−40 °C, 0.55 V, SS	20,000	3.0×10^{-2} ± 4 %	2.8×10^{-2} ± 3 %	−1.5
ASAP7 7 nm, write	27 °C, 0.6 V, FF	200,000	5.9×10^{-4} ± 9 %	4.7×10^{-4} ± 4 %	−2.0

± values are one standard error; z = (IS − MC)/combined standard error.

B. The Worst-Case Direction Is Not Universal

Fig. 5 shows μ/σ of the cell margin for the TT corner. For the SkyWater bitcell −40 °C is worst, and at 0.75 V the cold SS and SF corners fail even on average, consistent with the reversal of the temperature dependence of transistor current at low supply [21], [22]. For the ASAP7 read margin the three temperatures nearly coincide, whereas the write margin of the same cell is worst at −40 °C. A model trained on mild conditions has no information about which way the tail moves outside its training box.

C. Moment Surrogates Fail by Tail Shape, Not by Extrapolation

Table V separates the two sources of error of the moment surrogates in the design-relevant range $10^{-12} \le P_f < 10^{-3}$. A Gaussian tail with the *exact* sample moments at the same condition, i.e., without any model or extrapolation, already underestimates P_f by up to 7.1, 5.2, and 11.9 decades in the three studies. The extrapolated surrogates are only moderately worse. Modeling the reliability indices instead (Section IV-B) removes most of the error with the same training data and the same GP family. Fig. 6 shows all truth points: the moment surrogate falls increasingly below the truth as P_f decreases, as predicted by Section III-D.

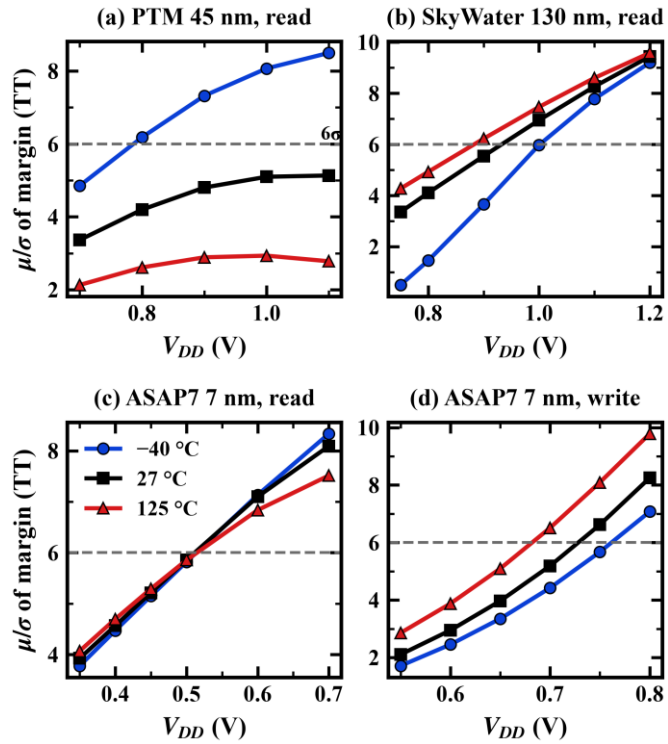

Fig. 5. Ratio μ/σ of the cell margin versus supply at −40, 27, and 125 °C (TT corner, 1000 cells per point). The dashed line marks 6σ. Panel (a) is a 45-nm planar predictive model [64] shown for comparison.

Table V
Worst Signed Error $\log_{10}(\hat{P}_F/P_F^{IS})$ of Tail Models, $10^{-12} \le P_F < 10^{-3}$

Study	Points	Gaussian, exact moments	Gauss. regression	GP	β-space GP
SkyWater 130 nm, read	58	−7.10	−9.53	−10.94	−1.93
ASAP7 7 nm, read	66	−5.19	−5.43	−5.20	−0.27
ASAP7 7 nm, write	39	−11.90	−12.75	−11.70	+0.08

Exact moments: sample mean and standard deviation of the 1000 cells at the same condition (no model). All other columns are fitted on the training box and evaluated at the unseen conditions.

D. Extrapolation in Reliability-Index Space

The β-space surrogate is accurate for both ASAP7 studies (largest error 0.27 and 0.08 decades) but not for the SkyWater cell, where it underestimates P_f by up to 1.9 decades, mostly at 125 °C and high supply. Fig. 7(a) shows the reason. The failure direction of the FinFET lobes is predicted within a few degrees (median 0.9° read, 0.4° write), but that of the planar foundry cell rotates across PVT (median 6.0°, 90th percentile 13°), i.e., the devices that dominate a failure change with temperature and corner. Fig. 7(b) confirms Proposition 1: after the anchored run, the remaining index error follows the direction term $\hat{\beta}(1 - \cos\theta)$ (Spearman correlation 0.88 for SkyWater) and is non-negative in all but a few cases. The calibrated θ_{95} is 15.3° for SkyWater, 4.5° for ASAP7 read, and 0.8° for ASAP7 write, so the guard (10) is large only where the direction actually rotates.

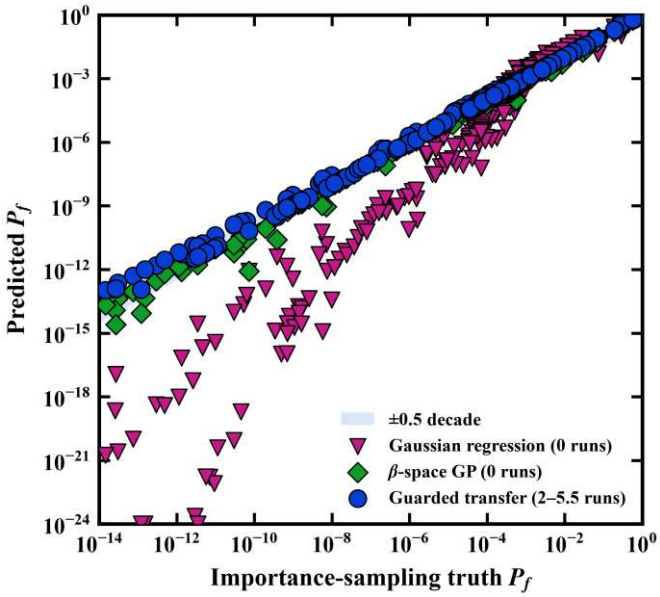

Fig. 6. Predicted versus IS truth at all unseen truth points of the three studies. The band is ±0.5 decade. Predictions below 10^{-24} are drawn at 10^{-24}.

TABLE VI
ERROR VERSUS IS AT UNSEEN CONDITIONS, $10^{-12} \le P_F < 10^{-3}$

Study	Estimator	Runs	Median / max |err|	Signed range	Unsafe %
SkyWater 130 nm, read	Gaussian regression	0	2.04 / 9.53	−9.53 … +0.74	93
	Gaussian process	0	2.29 / 10.94	−10.94 … −0.46	100
	β-space GP	0	0.10 / 1.93	−1.93 … +0.20	34
	Anchored transfer	2	0.06 / 0.34	−0.34 … +0.05	33
	Guarded transfer	5.5	0.25 / 0.75	−0.09 … +0.75	0
	One-step	39	0.03 / 0.41	−0.41 … +0.03	9
	FORM	71	0.03 / 0.18	−0.18 … +0.03	5
ASAP7 7 nm, read	Gaussian regression	0	1.34 / 5.43	−5.43 … −0.26	100
	Gaussian process	0	1.43 / 5.20	−5.20 … −0.86	100
	β-space GP	0	0.07 / 0.27	−0.27 … −0.00	33
	Anchored transfer	2	0.02 / 0.08	−0.08 … +0.05	0
	Guarded transfer	2.0	0.01 / 0.07	−0.06 … +0.07	0
	One-step	39	0.02 / 0.08	−0.08 … +0.05	0
	FORM	65	0.02 / 0.08	−0.08 … +0.05	0
ASAP7 7 nm, write	Gaussian regression	0	1.02 / 12.75	−12.75 … +0.41	62
	Gaussian process	0	1.17 / 11.70	−11.70 … +0.45	67
	β-space GP	0	0.04 / 0.08	−0.03 … +0.08	0
	Anchored transfer	2	0.04 / 0.07	−0.02 … +0.07	0
	Guarded transfer	2.0	0.04 / 0.07	−0.02 … +0.07	0
	One-step	39	0.04 / 0.46	−0.03 … +0.46	0
	FORM	93	0.04 / 0.07	−0.02 … +0.07	0

Errors in decades, $\log_{10}(\hat{P}_f/P_f^{IS})$. Runs: new SPICE runs per unseen condition (mean, including escalations; FORM: mean over conditions). Unsafe: fraction of conditions with error below −0.1 decade. Guarded transfer: Algorithm 1.

E. Accuracy Against Independent IS Truth

Table VI compares all estimators with the IS truth in the design-relevant range (58, 66, and 39 conditions). Two anchored runs remove most of the extrapolation error of the β-space surrogate (SkyWater: 1.93 → 0.34 decades), but the unguarded transfer is still unsafe by more than 0.1 decade at 33 % of the SkyWater conditions, as Proposition 1 predicts for a rotating failure direction. With the guard and escalation of Section IV-E it is never more than 0.09 decade unsafe on SkyWater, at the price of a pessimism of up to 0.75 decade and 5.5 runs per condition on average; for ASAP7 it is within 0.07 (read) and 0.07 (write) decades at two runs per condition. The full-grid truth also shows that FORM-class estimators are not automatically safe: on SkyWater the one-step estimator is unsafe by up to −0.41 decade (9 % of conditions) and converged FORM by up to −0.18 decade (5 %). On the curved ASAP7 write lobes the one-step estimator is conservative by up to 0.46 decade, whereas FORM and the transfer are within a few hundredths. Over the full range, which on SkyWater extends to 10^{-27}, the guarded transfer is never more than 0.09 decade unsafe.

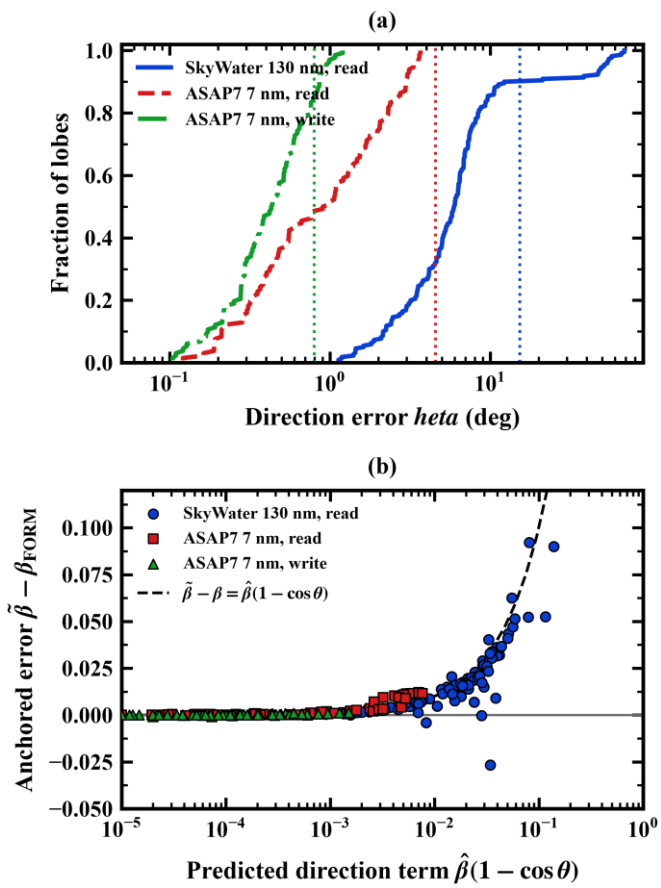

Fig. 7. (a) Distribution of the angle between the transferred and the FORM failure direction over all unseen lobes; dotted lines mark θ_{95} from edge-block cross-validation inside the training box. (b) Error of the anchored index (8) relative to FORM versus the direction term of (9).

F. Cost Against Sampling Baselines

Table VII and Fig. 8 compare the cost and accuracy of all estimators at the conditions where the sampling baselines were run. MNIS is biased low by up to several decades, because a single mean shift covers only one of the two symmetric lobes and random exploration in 12 dimensions finds a failure point far from the design point. SuS reaches 0.1–0.4 decades with thousands of runs, and AK-IS is accurate with 275 runs except

at the deepest write point. MPFP-mixture IS (Table VIII) is the strongest sampling competitor on the near-linear read lobes, reaching a median error of a few hundredths of a decade with a few hundred runs, but on the curved write lobes its 90th-percentile error stays above 0.4 decades even with 513 runs because the linearized design points misplace the proposal. The guarded transfer uses two runs; its error is conservative and is the price of not sampling.

TABLE VII
RUNS : MEDIAN / MAX |ERROR| AT THE BASELINE CONDITIONS

Study	MNIS	SuS	AK-IS	IS-MPFP 213	IS-MPFP 513	One-step	Guarded
SkyWater 130 nm, read	22,500: 0.42 / 2.13	5,500: 0.09 / 0.34	275: 0.03 / 0.04	213: 0.04 / 0.25	513: 0.04 / 0.12	39: 0.02 / 0.06	2: 0.31 / 0.60
ASAP7 7 nm, write	22,500: 0.60 / 3.47	5,500: 0.21 / 0.41	275: 0.01 / 0.39	213: 0.12 / 2.32	513: 0.13 / 1.41	39: 0.02 / 0.29	2: 0.03 / 0.07

Three unseen conditions per study and three seeds per stochastic method (five for IS-MPFP); errors in decades against the IS truth.

TABLE VIII
MPFP-MIXTURE IS: MEDIAN / 90TH PERCENTILE / MAX |ERROR| VERSUS BUDGET

Study	63 runs	113 runs	213 runs	513 runs
SkyWater 130 nm, read	0.08 / 0.36 / 0.91	0.09 / 0.21 / 0.35	0.05 / 0.14 / 0.25	0.03 / 0.09 / 0.20
ASAP7 7 nm, read	0.08 / 0.22 / 0.46	0.08 / 0.17 / 0.28	0.05 / 0.15 / 0.19	0.03 / 0.09 / 0.14
ASAP7 7 nm, write	0.24 / 1.44 / 2.33	0.19 / 0.78 / 1.91	0.12 / 0.91 / 2.32	0.08 / 0.62 / 1.41

All original truth points of each study, five seeds each; d+1 = 13 linearization runs included.

G. Calibrated Sign-Off and Minimum Operating Voltage

The calibrated sign-off of Section IV-G was evaluated with n_{cal} = 20 calibration conditions and $\alpha_1 = \alpha_2 = \alpha_c = 0.05$ (nominal bound 0.15) over 500 random calibration/test splits per study. Table IX lists the decisions at the two targets. The targets are rarely close calls on these grids: no calibrated decision was unsafe, and neither were the decisions of the guarded transfer, the one-step estimator, or FORM, so the decision study shows that the procedure is safe here but does not discriminate between the methods. Including the calibration IS, it cost 1573–2608 runs per condition (Table II adds the training cost). Part (i) of Proposition 2 can be checked directly: the fraction of test conditions whose IS upper confidence limit exceeded the conformal upper limit was 4.4 %, 4.8 %, and 4.5 % on average against a bound of 5 %, but in individual SkyWater splits it reached 14 % (95th percentile): the bound holds on average over calibration draws, not for every draw.

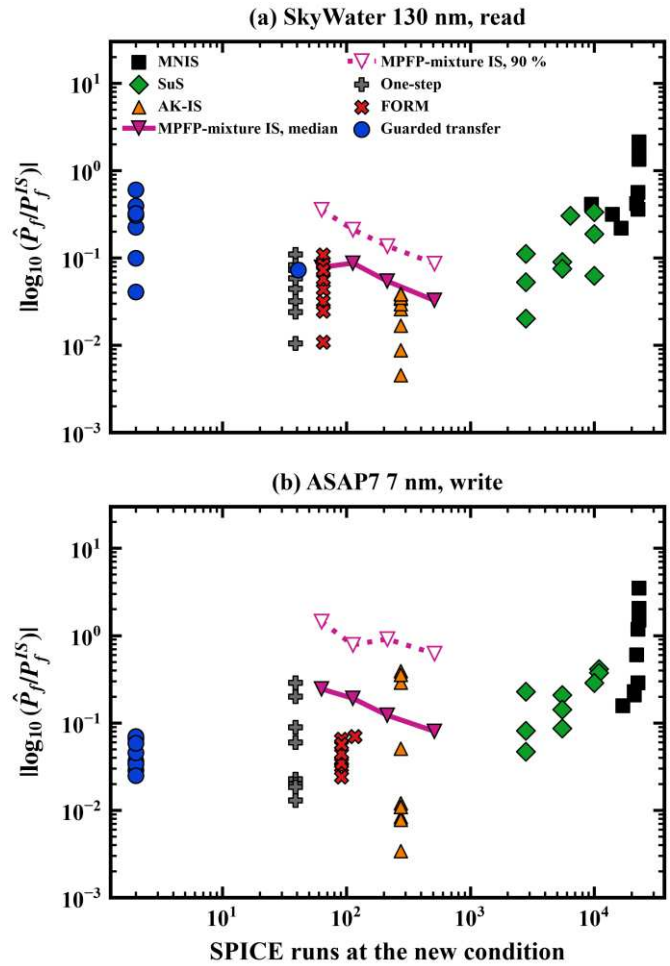

Fig. 8. Accuracy versus SPICE runs at a new condition. Sampling baselines at their three conditions per study; MPFP-mixture IS, one-step, FORM, and the guarded transfer at all original truth points. Errors below 10^{-3} are drawn at 10^{-3}.

TABLE IX
SIGN-OFF DECISIONS: UNSAFE ACCEPTANCE % / FALSE REJECTION % (MEAN OVER 500 SPLITS)

Study	Target	Failing	Calibrated	Guarded	One-step	FORM	IS all	Runs (cal.)
SkyWater 130 nm, read	10^{-6}	54/90	0.0 / 0.9	0.0 / 0.9	0.0 / 0.0	0.0 / 0.0	0.0 / 0.9	1587
	10^{-9}	62/90	0.0 / 0.9	0.0 / 2.8	0.0 / 0.0	0.0 / 0.0	0.0 / 0.9	1573
ASAP7 7 nm, read	10^{-6}	34/46	0.0 / 1.5	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 1.5	2608
	10^{-9}	43/46	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	2469
ASAP7 7 nm, write	10^{-6}	33/46	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	2455
	10^{-9}	39/46	0.0 / 0.0	0.0 / 0.0	0.0 / 2.9	0.0 / 0.0	0.0 / 0.0	2584

Failing: test conditions whose IS P_f exceeds the target / test conditions. Calibrated: transfer with split-conformal limits and IS escalation ($\alpha_1 = \alpha_2 = \alpha_c = 0.05$). Runs: SPICE runs per condition of the calibrated sign-off including the calibration IS, excluding training.

The decisive quantity for a designer is V_{min}. Table X compares the interpolated V_{min} of each estimator with that of the IS curves on every temperature–corner curve that lies entirely outside the training box. The moment surrogates set V_{min} too low by up to 219 mV and the β-space surrogate by up to 62 mV. More importantly for sign-off, the unguarded

transfer, the one-step estimator, and FORM are also unsafe by a few millivolts on many curves (worst −10, −7 mV), and the block-cross-validated guard of Section IV-E is unsafe on some ASAP7 curves (−3 mV) while pessimistic by up to 63 mV on SkyWater. $V_{\min}$ read from the conformal upper limit was never more than 1 mV too low and at most 34 mV pessimistic, the only estimator that was both safe on every curve and less pessimistic than the empirical guard. This is an empirical result: Proposition 2 concerns single conditions, whereas $V_{\min}$ combines the upper limits at several supplies of a curve, so it carries no formal guarantee. Fig. 9 shows two curves with the IS points and the conformal upper limit.

TABLE X
INTERPOLATED V_{MIN} ERROR VERSUS IS (MV): MOST UNSAFE … MOST PESSIMISTIC

Study	Cases	Gauss.	GP	β-space GP	Transfer	Guarded	FORM	Conformal
SkyWater 130 nm, read	12	−161 … −65	−219 … −53	−62 … +9	−10 … −1	+16 … +63	−7 … +0	−0 … +34
ASAP7 7 nm, read	9	−188 … −82	−190 … −86	−8 … −1	−5 … −1	−3 … +1	−5 … −1	−1 … +5
ASAP7 7 nm, write	12	−62 … −23	−63 … −23	+1 … +2	+1 … +2	+1 … +2	+1 … +2	+0 … +4

Cases: (curve, target) pairs, targets 10^{-6} and 10^{-9}, on temperature–corner curves entirely outside the training box; reference: IS point estimates. Negative: claimed $V_{\min}$ too low (unsafe). Conformal: $V_{\min}$ of the conformal upper limit over 200 random calibration draws (empirical, no formal guarantee).

Table XI shows how the calibrated sign-off depends on its settings. With $n_{\mathrm{cal}} = 10$ and $\alpha_1 = 0.05$ the conformal quantile is infinite, so every condition is escalated to IS and nothing is gained; with 20 to 40 calibration conditions and α_1 between 0.025 and 0.05 the conformal $V_{\min}$ is unchanged and the cost grows linearly with n_{cal}.

TABLE XI
CALIBRATED SIGN-OFF VERSUS CALIBRATION SIZE AND RISK (TARGET 10^{-9})

n_{cal}	$\alpha_1+\alpha_2$	SKY runs	SKY $V_{\min}$ (mV)	R runs	R $V_{\min}$ (mV)	W runs	W $V_{\min}$ (mV)
10	0.1	8072	none (all to IS)	8067	none (all to IS)	8095	none (all to IS)
20	0.1	1573	−0 … +34	2469	+1 … +5	2584	+0 … +4
30	0.1	2301	+1 … +34	3696	+2 … +5	3805	+0 … +4
40	0.05	3026	+1 … +34	4919	+2 … +5	5013	+0 … +4

SKY: SkyWater read; R: ASAP7 read; W: ASAP7 write; $\alpha_1 = \alpha_2$, $\alpha_c = 0.05$. Runs per condition include the calibration IS, not training. $V_{\min}$: most unsafe … most pessimistic error of the conformal upper limit over 200 calibration draws.

H. Case Study: Sign-Off of a 1-Mb Macro

A designer signs off a memory macro, not a cell. We therefore considered a 1-Mb macro of 1024 rows and 1024 columns of each cell with a yield target of 99.9 %. Without repair the macro yield is $(1 - P_f)^N$ with $N = 2^{20}$, which requires $P_f \leq 9.5 \times 10^{-10}$; with four spare columns and column repair it is the probability that at most four of the 1028 columns contain a failing cell, which requires $P_f \leq 7.0 \times 10^{-7}$, close to the two per-cell targets used above. On the fully unseen curves of Table X, each estimator claims the supply at which its P_f curve crosses this requirement, and the yield actually delivered at that supply follows from the IS truth interpolated there; no new simulation is involved. Table XII lists the worst delivered yield. Macros signed off with the moment surrogates would yield as little as 3.7 % without repair and 0.0 % with repair. FORM and the unguarded transfer fell short of the target on 20 and 21 of the 33 (curve, configuration) cases, down to 99.77 % and 99.66 %, a yield loss up to 3.4 times the budget of 0.1 %. The guarded transfer met the target on SkyWater at the price of its pessimism but fell short on 5 ASAP7 read cases. The conformal upper limit delivered at least 99.90 % and 99.88 %; it fell short of 99.9 % only in 0.2 to 0.5 % of the calibration draws of two studies with repair, where its $V_{\min}$ was at most 1 mV too low. At macro level, the millivolt errors of FORM-class estimators therefore become a yield loss of several times the budget, and the surrogate errors a failed product.

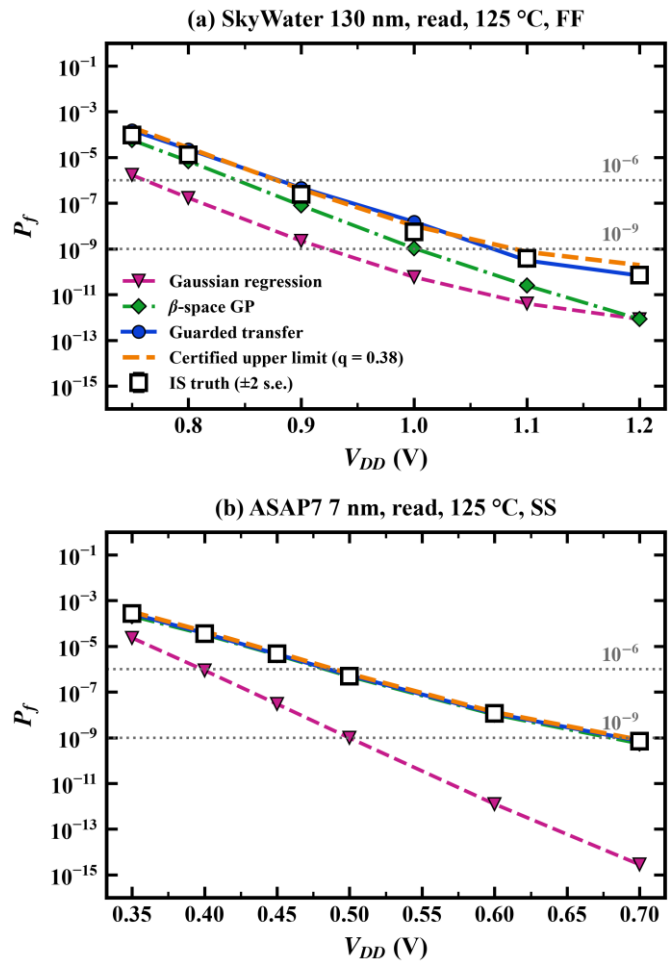

Fig. 9. Failure probability versus supply with IS truth (squares, ±2 standard errors). (a) SkyWater read at 125 °C, FF; (b) ASAP7 read at 125 °C, SS. The conformal upper limit is the transfer plus the median conformal margin.

TABLE XII
WORST YIELD (%) OF A 1-MB MACRO SIGNED OFF AT EACH ESTIMATOR'S V_{MIN} (TARGET 99.9 %)

Study	Repair	Moment	β-space GP	Transfer	Guarded	FORM	Conformal
SkyWater	none	3.7	99.46	99.86	99.96	99.87	99.90
SkyWater	4 col.	0.0	96.5	99.66	99.99	99.79	99.88
ASAP7 R	none	57.5	99.87	99.88	99.89	99.88	99.90
ASAP7 R	4 col.	0.0	99.74	99.77	99.83	99.77	99.89
ASAP7 W	none	74.4	99.90	99.90	99.91	99.90	99.90
ASAP7 W	4 col.	22.1	99.91	99.92	99.92	99.92	99.91

1024 × 1024 cells; repair: spare columns. Moment: worse of the Gaussian and GP surrogates. Delivered yield from the IS truth at the claimed V_{min} on the fully unseen curves of Table X; Conformal: worst over 200 calibration draws. R: read; W: write.

I. Robustness

Because the ASAP7 mismatch is assumed, the read study was repeated with σ_{Vth} = 20 and 45 mV per fin, each with its own dataset, training box, FORM, and IS truth. Table XIII shows that the conclusions do not depend on this choice.

TABLE XIII
ASAP7 READ: MAX |ERROR| VERSUS IS FOR DIFFERENT ASSUMED Σ_{VTH}

σ_{Vth} (mV)	Points	θ_{95}	Gauss.	β-space GP	Anchored	Proposed	One-step
20	8	5.1°	4.76	0.43	0.12	0.12	0.05
31 (nominal)	66	4.5°	5.43	0.27	0.08	0.07	0.08
45	8	4.8°	2.07	0.11	0.09	0.06	0.09

Points: unseen IS truth points of each variant; the nominal variant also has the additional stratified and V_{min}-curve points of Section V.

The guarded transfer has two settings, the quantile of the direction error and δ. Over quantiles of 90, 95, and 99 % and δ = 0.1, 0.25, and 0.5, the most unsafe SkyWater error stays between −0.23 and −0.05 decades, the most pessimistic between +0.86 and +2.14, and the mean cost between 4.1 and 13.0 runs per condition; for the ASAP7 studies the results are unchanged within 0.01 decade. The settings trade pessimism against cost, not safety.

VII. DISCUSSION

A. When Is Transfer Safe?

The results separate three effects that a single accuracy number would hide. The Gaussian tail of the two-lobe minimum is wrong by orders of magnitude at every condition, so moment-based models should not be used to read P_f at all in the far tail. Extrapolation in reliability-index space is accurate when the failure direction is stable across PVT, as for the FinFET cell, and fails when it rotates, as for the planar foundry cell; this rotation can be measured inside the training box by the block cross-validation of Section IV-E, before any unseen condition is simulated. The anchored run then removes the first-order part of the extrapolation error, and the remainder has a known, one-sided sign that the guard or the calibrated sign-off accounts for.

B. What the Calibration Does and Does Not Guarantee

The conformal calibration converts the empirically conservative behavior of the transfer into a stated risk, but a limited one. Under exchangeability of the calibration and sign-off conditions it bounds the probability of an unsafe acceptance at a *randomly selected* condition, equivalently the expected fraction of unsafe acceptances. It does not bound the probability that *any* condition of the grid is accepted unsafely, it does not bound P_f at a particular condition, it holds on average over calibration draws rather than for every draw, it does not transfer to V_{min}, which combines several conditions, and it does not cover conditions outside the sign-off set, such as a new supply range. Its IS part is asymptotic and requires valid IS confidence limits; our cross-check of proposal A against proposal B shows that this is not automatic on strongly curved limit states, so the effective sample size must be monitored. The price is IS at about 20 random conditions per sign-off set, which is worthwhile when the set is large, when a stated marginal risk is required, or when the empirical guard is too pessimistic.

C. A Sign-Off Workflow

1) Run FORM at the training conditions and fit the transfer. 2) Inspect θ_{95}: a few degrees indicates that the transfer is accurate; tens of degrees indicate that its errors will be larger. 3) Apply Algorithm 1 at every sign-off condition. 4) Draw at least $\lceil 1/\alpha_1 \rceil - 1$ calibration conditions at random from the sign-off set, run IS there with FORM-centered proposals and an ESS check, and compute the conformal margin. 5) Accept or reject with the conformal limits and run IS at the undecided conditions. 6) Read V_{min} from the conformal upper limit, keeping in mind that this V_{min} is empirically, not formally, safe.

D. Extensions

Cells with more than two failure mechanisms form a series system of k limit states; the transfer and the anchored correction apply per limit state, at k runs per condition, and P_f follows from the k-variate normal distribution. The read-decoupled 8T cell keeps the 6T latch for write and hold, so its write and hold lobes have the same mirror structure. Dynamic metrics [37], [38] replace the DC sweep by a transient simulation per cell without changing the number of runs. Assist techniques [40] change the limit states but not their number.

E. Threats to Validity

Ground truth. IS covers every unseen condition, agrees with plain MC within sampling error (Table IV), and was cross-checked by a second proposal at every condition below 10^{-7}; where the two disagreed, the proposal with the healthy effective sample size was used. *Favoring the baselines.* Common random numbers remove sampling noise from the surrogate training data, the training box includes all corners, and the β-space surrogate is given converged FORM targets. *No tuning on the test set.* The guard is calibrated only on training data, the conformal limits only on their random calibration subset, and the decision rules and settings were fixed before the full-grid truth was available. *Baseline implementations.* MNIS, SuS, and AK-IS were implemented by the author from their original descriptions; tuned implementations may do better, which would narrow the cost gap but not affect the conclusions on tail shape, transfer, and calibrated sign-off.

F. Limitations

1) The metrics are static noise margins; dynamic failures depend on timing [37], [38]. 2) Hold failures were not studied: near the data-retention voltage of the SkyWater cell at −40 °C the subthreshold currents fall below the simulator's minimum conductance (10^{-12} S). 3) Only the SkyWater study uses a

foundry statistical model; the ASAP7 mismatch is an assumption, mitigated by Section VI-I. 4) The ngspice BSIM-CMG implementation ignores five card parameters of the ASAP7 models (version, coremod, capmod, etaqm, nseg). 5) All estimators except IS are first order, and on the planar cell FORM itself is unsafe by up to 0.18 decade; a second-order correction [44] was not studied. 6) The empirical guard assumes that the direction error at unseen conditions is no larger than at the edges of the training box, and it was unsafe by a few millivolts of $V_{\min}$ on some ASAP7 curves; only the conformal calibration carries a stated guarantee, and that guarantee is marginal over conditions, partly asymptotic, requires exchangeability, and costs IS at about 20 calibration conditions. 7) The transfer gives no cost advantage over FORM at every condition on grids of the size studied here once training is counted (Table II). 8) Global variation is represented by discrete corners, and the macro study of Section VI-H assumes independent cell failures and column repair only. 9) No silicon data are available to the author; all conclusions are relative to SPICE.

VIII. Conclusion

Extrapolating SRAM failure probability across PVT is unsafe when the tail is read from modeled moments, and the main reason is the tail shape of the two-lobe cell rather than extrapolation: even exact moments err by up to 11.9 decades. Extrapolating the per-lobe reliability index is accurate when the failure direction is stable across PVT and fails when it rotates. Transferring the design point and correcting it with one SPICE run per lobe removes the first-order error and leaves a one-sided remainder. A deterministic bound on the failure probability is not available without unverifiable convexity, but a split-conformal calibration computed from IS on a random subset of the sign-off conditions bounds the probability of an unsafe acceptance at a randomly selected condition under exchangeability, a marginal rather than a grid-wide guarantee. Against IS at every unseen condition, cross-checked by plain MC and a second IS proposal, the calibrated sign-off made no unsafe decision, and the minimum operating voltage read from its upper limit lay −1 to +34 mV from the reference, an empirical result, whereas FORM and the unguarded transfer were unsafe by several millivolts and moment surrogates by up to hundreds. For a 1-Mb macro with a 99.9 % yield target, these errors meant a delivered yield down to 99.66 % for FORM-class estimators and a failed product for moment surrogates, whereas the calibrated limit delivered at least 99.88 %. Including training and calibration, the calibrated sign-off needed 3.1 to 4.9 times fewer SPICE runs than IS at every condition. The calibration costs IS at about 20 conditions per sign-off set, and on the grids studied here the transfer itself costs about as much as FORM at every condition; its benefit is the safety obtained on the planar cell, with a cost advantage only for denser sweeps. The scripts and data needed to reproduce all results are provided as supplementary material.

Appendix A

Per-Instance Foundry Mismatch for SKY130

In the SKY130 model files, local mismatch enters each binned BSIM4 model through expressions of the form $p = p_0 + \xi_p \cdot k_p/\sqrt{(L\ W \cdot \text{mult})}$, where k_p is the foundry mismatch slope and ξ_p a *global* parameter with unit standard deviation that a statistical simulator varies. Because a single global ξ_p would apply the same draw to every transistor, we turned each ξ_p into a subcircuit parameter with default 0 and passed an independent standard-normal value to every instance, leaving k_p and the $\sqrt{(LW)}$ scaling unchanged. Only the size bins used by the cell were kept. For the bitcell devices the oxide-thickness and threshold-voltage slopes are non-zero, giving two mismatch variables per device. A check on a logic nFET of 0.42/0.15 µm gave a threshold shift of 13.7 mV for a +1σ draw, against 13.4 mV from the foundry formula.

Appendix B

IS Truth Campaign

Table XIV summarizes the canonical truth. The value, standard error, proposal, and effective sample size at every unseen condition, together with the predictions of all estimators, are provided in the supplementary material.

Table XIV
Canonical IS Truth: Sources and Cross-Check

Study	Conditions	Proposal B used	A/B pairs	$\lvert z\rvert > 2$	$\lvert z\rvert > 3$	Min ESS (B)
SkyWater 130 nm, read	110	38	37	1	0	266
ASAP7 7 nm, read	66	12	12	0	0	858
ASAP7 7 nm, write	66	16	14	3	3	753

Proposal A: defensive mixture at the linearized design points with re-centering; proposal B: defensive mixture at the converged FORM design points. z: difference of the two estimates over their combined standard error.

Acknowledgment

During the preparation of this manuscript, the author used Claude (Anthropic) for assistance with code development, technical troubleshooting, data analysis, and manuscript preparation. The author reviewed and validated the results and takes full responsibility for the content of this publication.

Conflict of Interest

The author declares no conflict of interest.

Data Availability

The code and data supporting the findings of this study are provided as supplementary material with this article, available at [DOI link assigned by IEEE].

References

[1] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 µm MOSFET's: A 3-D 'atomistic' simulation study," *IEEE Trans. Electron Devices*, vol. 45, no. 12, pp. 2505–2513, Dec. 1998, doi: 10.1109/16.735728.

[2] K. J. Kuhn *et al.*, "Process technology variation," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2197–2208, Aug. 2011, doi: 10.1109/TED.2011.2121913.

[3] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989, doi: 10.1109/JSSC.1989.572629.

[4] A. J. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 658–665, Apr. 2001, doi: 10.1109/4.913744.

[5] Z. Guo, A. Carlson, L.-T. Pang, K. T. Duong, T.-J. K. Liu, and B. Nikolić, "Large-scale SRAM variability characterization in 45 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 11, pp. 3174–3192, Nov. 2009, doi: 10.1109/JSSC.2009.2032698.

[6] S. Mukhopadhyay, H. Mahmoodi, and K. Roy, "Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscaled CMOS," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 12, pp. 1859–1880, Dec. 2005, doi: 10.1109/TCAD.2005.852295.

[7] A. Singhee and R. A. Rutenbar, "Statistical blockade: Very fast statistical simulation and modeling of rare circuit events and its application to memory design," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 8, pp. 1176–1189, Aug. 2009, doi: 10.1109/TCAD.2009.2020721.

[8] R. Kanj, R. Joshi, and S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," in *Proc. 43rd Design Autom. Conf. (DAC)*, 2006, doi: 10.1145/1146909.1146930.

[9] L. Dolecek, M. Qazi, D. Shah, and A. Chandrakasan, "Breaking the simulation barrier: SRAM evaluation through norm minimization," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, 2008, pp. 322–329, doi: 10.1109/ICCAD.2008.4681593.

[10] K. Katayama, S. Hagiwara, H. Tsutsui, H. Ochi, and T. Sato, "Sequential importance sampling for low-probability and high-dimensional SRAM yield analysis," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, 2010, pp. 703–708, doi: 10.1109/ICCAD.2010.5654259.

[11] M. Wang, C. Yan, X. Li, D. Zhou, and X. Zeng, "High-dimensional and multiple-failure-region importance sampling for SRAM yield analysis," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 3, pp. 806–819, Mar. 2017, doi: 10.1109/TVLSI.2016.2601606.

[12] T. Haine, J. Segers, D. Flandre, and D. Bol, "Gradient importance sampling: An efficient statistical extraction methodology of high-sigma SRAM dynamic characteristics," in *Proc. Design, Autom. Test Eur. Conf. (DATE)*, 2018, pp. 195–200, doi: 10.23919/DATE.2018.8342002.

[13] Y. Liu, G. Dai, and W. W. Xing, "Seeking the yield barrier: High-dimensional SRAM evaluation through optimal manifold," in *Proc. 60th ACM/IEEE Design Autom. Conf. (DAC)*, 2023, pp. 1–6, doi: 10.1109/DAC56929.2023.10247952.

[14] Z. Gao, D. Zhang, L. Daniel, and D. S. Boning, "NOFIS: Normalizing flow for rare circuit failure analysis," in *Proc. 61st ACM/IEEE Design Autom. Conf. (DAC)*, 2024, pp. 1–6, doi: 10.1145/3649329.3658459.

[15] C. Gu and J. Roychowdhury, "An efficient, fully nonlinear, variability-aware non-Monte-Carlo yield estimation procedure with applications to SRAM cells and ring oscillators," in *Proc. Asia South Pacific Design Autom. Conf. (ASP-DAC)*, 2008, pp. 754–761, doi: 10.1109/ASPDAC.2008.4484052.

[16] S. Sun and X. Li, "Fast statistical analysis of rare circuit failure events via scaled-sigma sampling for high-dimensional variation space," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, 2015, pp. 1–4, doi: 10.1109/CICC.2015.7338409.

[17] W. Wu, W. Xu, R. Krishnan, Y.-L. Chen, and L. He, "REscope: High-dimensional statistical circuit simulation towards full failure region coverage," in *Proc. 51st Design Autom. Conf. (DAC)*, 2014, pp. 1–6, doi: 10.1109/DAC.2014.6881409.

[18] S.-K. Au and J. L. Beck, "Estimation of small failure probabilities in high dimensions by subset simulation," *Probabilistic Eng. Mech.*, vol. 16, no. 4, pp. 263–277, 2001, doi: 10.1016/S0266-8920(01)00019-4.

[19] B. Echard, N. Gayton, and M. Lemaire, "AK-MCS: An active learning reliability method combining Kriging and Monte Carlo simulation," *Struct. Saf.*, vol. 33, no. 2, pp. 145–154, 2011, doi: 10.1016/j.strusafe.2011.01.002.

[20] B. Echard, N. Gayton, M. Lemaire, and N. Relun, "A combined importance sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models," *Reliab. Eng. Syst. Saf.*, vol. 111, pp. 232–240, 2013, doi: 10.1016/j.ress.2012.10.008.

[21] R. Kumar and V. Kursun, "Reversed temperature-dependent propagation delay characteristics in nanometer CMOS circuits," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 10, pp. 1078–1082, Oct. 2006, doi: 10.1109/TCSII.2006.882218.

[22] D. Wolpert and P. Ampadu, "Temperature effects in semiconductors," in *Managing Temperature Effects in Nanoscale Adaptive Systems*. New York, NY, USA: Springer, 2012, pp. 15–33, doi: 10.1007/978-1-4614-0748-5_2.

[23] J. Yao, Z. Ye, and Y. Wang, "An efficient SRAM yield analysis and optimization method with adaptive online surrogate modeling," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 7, pp. 1245–1253, Jul. 2015, doi: 10.1109/TVLSI.2014.2336851.

[24] X. Shi, H. Yan, Q. Huang, J. Zhang, L. Shi, and L. He, "Meta-model based high-dimensional yield analysis using low-rank tensor approximation," in *Proc. 56th Design Autom. Conf. (DAC)*, 2019, pp. 1–6, doi: 10.1145/3316781.3317863.

[25] S. Yin, X. Jin, L. Shi, K. Wang, and W. W. Xing, "Efficient Bayesian yield analysis and optimization with active learning," in *Proc. 59th ACM/IEEE Design Autom. Conf. (DAC)*, 2022, pp. 1195–1200, doi: 10.1145/3489517.3530607.

[26] Z. Gao, F. Wang, J. Tao, Y. Su, X. Zeng, and X. Li, "Correlated Bayesian model fusion: Efficient high-dimensional performance modeling of analog/RF integrated circuits over multiple corners," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 42, no. 2, pp. 360–370, Feb. 2023, doi: 10.1109/TCAD.2022.3174170.

[27] W. W. Xing, K. Huang, J. Liu, H. Qiu, and S. Shen, "Breaking the tuning barrier: Zero-hyperparameters yield multi-corner analysis via learned priors," 2026, *arXiv:2603.13092*.

[28] M. Hohenbichler and R. Rackwitz, "Sensitivity and importance measures in structural reliability," *Civil Eng. Syst.*, vol. 3, no. 4, pp. 203–209, 1986, doi: 10.1080/02630258608970445.

[29] P. Bjerager and S. Krenk, "Parametric sensitivity in first order reliability theory," *J. Eng. Mech.*, vol. 115, no. 7, pp. 1577–1582, 1989, doi: 10.1061/(ASCE)0733-9399(1989)115:7(1577).

[30] J. Liang, Z. P. Mourelatos, and J. Tu, "A single-loop method for reliability-based design optimisation," *Int. J. Product Develop.*, vol. 5, no. 1/2, p. 76, 2008, doi: 10.1504/IJPD.2008.016371.

[31] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*, 2nd ed. Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-031-06649-8.

[32] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, "Distribution-free predictive inference for regression," *J. Amer. Statist. Assoc.*, vol. 113, no. 523, pp. 1094–1111, 2018, doi: 10.1080/01621459.2017.1307116.

[33] E. Jaber, V. Chabridon, and M. Mougeot, "AK-MCS-C2: Active Kriging Monte Carlo simulation method with conformal certification for failure probability estimation," 2026, *arXiv:2606.20191*.

[34] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. 22, no. 5, pp. 748–754, Oct. 1987, doi: 10.1109/JSSC.1987.1052809.

[35] B. H. Calhoun and A. Chandrakasan, "Analyzing static noise margin for sub-threshold SRAM in 65nm CMOS," in *Proc. 31st Eur. Solid-State Circuits Conf. (ESSCIRC)*, 2005, pp. 363–366, doi: 10.1109/ESSCIR.2005.1541635.

[36] E. Grossar, M. Stucchi, K. Maex, and W. Dehaene, "Read stability and write-ability analysis of SRAM cells for nanometer technologies," *IEEE J. Solid-State Circuits*, vol. 41, no. 11, pp. 2577–2588, Nov. 2006, doi: 10.1109/JSSC.2006.883344.

[37] J. Wang, S. Nalam, and B. H. Calhoun, "Analyzing static and dynamic write margin for nanometer SRAMs," in *Proc. Int. Symp. Low Power Electron. Design (ISLPED)*, 2008, doi: 10.1145/1393921.1393954.

[38] S. O. Toh, Z. Guo, T.-J. K. Liu, and B. Nikolić, "Characterization of dynamic SRAM stability in 45 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 46, no. 11, pp. 2702–2712, Nov. 2011, doi: 10.1109/JSSC.2011.2164300.

[39] H. Qin, Y. Cao, D. Markovic, A. Vladimirescu, and J. Rabaey, "SRAM leakage suppression by minimizing standby supply voltage," in *Proc. Int. Symp. Quality Electron. Design (ISQED)*, 2004, pp. 55–60, doi: 10.1109/ISQED.2004.1283650.

[40] R. W. Mann *et al.*, "Impact of circuit assist methods on margin and performance in 6T SRAM," *Solid-State Electron.*, vol. 54, no. 11, pp. 1398–1407, 2010, doi: 10.1016/j.sse.2010.06.009.

[41] S. Shen, X. Li, Z. Liu, J. Ma, Y. Wang, Y. Wu, Y. Sun, and W. W. Xing, "OpenYield: An open-source SRAM yield analysis and optimization benchmark suite," 2025, *arXiv:2508.04106*.

[42] A. M. Hasofer and N. C. Lind, "Exact and invariant second-moment code format," *J. Eng. Mech. Div.*, vol. 100, no. 1, pp. 111–121, 1974, doi: 10.1061/JMCEA3.0001848.

[43] R. Rackwitz and B. Fiessler, "Structural reliability under combined random load sequences," *Comput. Struct.*, vol. 9, no. 5, pp. 489–494, 1978, doi: 10.1016/0045-7949(78)90046-9.

[44] K. Breitung, "Asymptotic approximations for multinormal integrals," *J. Eng. Mech.*, vol. 110, no. 3, pp. 357–366, 1984, doi: 10.1061/(ASCE)0733-9399(1984)110:3(357).

[45] K. J. Antreich, H. E. Graeb, and C. U. Wieser, "Circuit analysis and optimization driven by worst-case distances," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 13, no. 1, pp. 57–71, Jan. 1994, doi: 10.1109/43.273749.

[46] H. E. Graeb, *Analog Design Centering and Sizing*. Dordrecht, The Netherlands: Springer, 2007, doi: 10.1007/978-1-4020-6004-5.

[47] O. Ditlevsen and P. Bjerager, "Methods of structural systems reliability," *Struct. Saf.*, vol. 3, no. 3–4, pp. 195–229, 1986, doi: 10.1016/0167-4730(86)90004-4.

[48] R. E. Melchers and A. T. Beck, *Structural Reliability Analysis and Prediction*, 3rd ed. Hoboken, NJ, USA: Wiley, 2018, doi: 10.1002/9781119266105.

[49] I. Papaioannou, W. Betz, K. Zwirglmaier, and D. Straub, "MCMC algorithms for subset simulation," *Probabilistic Eng. Mech.*, vol. 41, pp. 89–103, 2015, doi: 10.1016/j.probengmech.2015.06.006.

[50] R. Teixeira, M. Nogal, and A. O'Connor, "Adaptive approaches in metamodel-based reliability analysis: A review," *Struct. Saf.*, vol. 89, Art. no. 102019, 2021, doi: 10.1016/j.strusafe.2020.102019.

[51] M. Moustapha, S. Marelli, and B. Sudret, "Active learning for structural reliability: Survey, general framework and benchmark," *Struct. Saf.*, vol. 96, Art. no. 102174, 2022, doi: 10.1016/j.strusafe.2021.102174.

[52] T. Hesterberg, "Weighted average importance sampling and defensive mixture distributions," *Technometrics*, vol. 37, no. 2, pp. 185–194, 1995, doi: 10.1080/00401706.1995.10484303.

[53] A. Owen and Y. Zhou, "Safe and effective importance sampling," *J. Amer. Statist. Assoc.*, vol. 95, no. 449, pp. 135–143, 2000, doi: 10.1080/01621459.2000.10473909.

[54] F. Wagner, J. Latz, I. Papaioannou, and E. Ullmann, "Error analysis for probabilities of rare events with approximate models," *SIAM J. Numer. Anal.*, vol. 59, no. 4, pp. 1948–1975, 2021, doi: 10.1137/20M1359808.

[55] A. Maurer and M. Pontil, "Empirical Bernstein bounds and sample variance penalization," in *Proc. 22nd Conf. Learn. Theory (COLT)*, 2009. [Online]. Available: https://arxiv.org/abs/0907.3740

[56] A. N. Angelopoulos and S. Bates, *Conformal Prediction: A Gentle Introduction*. Hanover, MA, USA: Now Publishers, 2023, doi: 10.1561/9781638281597.

[57] C. E. Clark, "The greatest of a finite set of random variables," *Oper. Res.*, vol. 9, no. 2, pp. 145–162, 1961, doi: 10.1287/opre.9.2.145.

[58] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006, doi: 10.7551/mitpress/3206.001.0001.

[59] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[60] H. Vogt *et al.*, "Ngspice users manual, version 47," 2025. [Online]. Available: https://ngspice.sourceforge.io

[61] SkyWater Technology Foundry and Google, "SkyWater SKY130 open source process design kit," 2020. [Online]. Available: https://github.com/google/skywater-pdk

[62] L. T. Clark *et al.*, "ASAP7: A 7-nm finFET predictive process design kit," *Microelectron. J.*, vol. 53, pp. 105–115, Jul. 2016, doi: 10.1016/j.mejo.2016.04.006.

[63] J. P. Duarte *et al.*, "BSIM-CMG: Standard FinFET compact model for advanced circuit design," in *Proc. 41st Eur. Solid-State Circuits Conf. (ESSCIRC)*, 2015, pp. 196–201, doi: 10.1109/ESSCIRC.2015.7313862.

[64] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45 nm early design exploration," *IEEE Trans. Electron Devices*, vol. 53, no. 11, pp. 2816–2823, Nov. 2006, doi: 10.1109/TED.2006.884077.

Nabanita Sarker received the B.Sc. degree in electrical and electronic engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2026. Current research interests include reliability and variability of nanoscale circuits, rare-event simulation, and first-principles modeling of two-dimensional semiconductor heterostructures.