Efficient Transistor-level Timing Yield Estimation via Line Sampling

Hiromitsu Awano and Takashi Sato (Kyoto Univ.)

京都大学
KYOTO UNIVERSITY

Introduction: increasing device density

Shrinkage of semiconductor manufacturing process still continues

- **Advantage:**
  Integrate billions of transistor into small silicon chip, enhancing computational power with device cost remained

- **Problem:**
  Increasing process variability further complicates circuit design
  Representative examples include SRAM cell design:

    Modern processor embeds large cache memory, requiring extremely high-level of reliability for single bit cell

    ➡️ **Estimation of RARE circuit failure probability becoming increasingly important**

Introduction: increasing demand for computational resource

- **Development of computationally heavy task: machine learning**
  - **Deep learning: stacked layers to achieve high performance but require high computational cost**

- **Massively parallel processor to cope with increasing demand for computational resource**
  - **Graphical processing unit (GPU) as general purpose accelerator**
  - **TrueNorth, neuromorphic chip from IBM**

- **Small arithmetic circuit is highly repeated to form an entire processor**

  Facing similar problem as SRAM cell design, i.e. extremely high-level of reliability is required for elemental circuit

  Accurate timing yield estimation is thus an important challenge

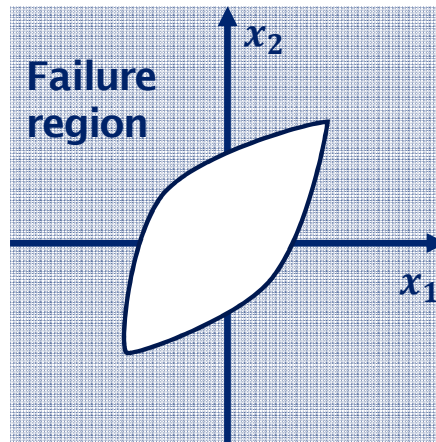# Difference between SRAM yield estimation and timing yield estimation

SRAM cell yield estimation

Timing yield estimation of combinational circuit

No. of random variables required

No more than 100

1000 or more

⎧ **Random variable represents Vth, gate length or width mismatches** ⎫

Shape of failure boundary

Failure region x_2 x_1

Failure region x_2 x_1

Complicated failure boundary
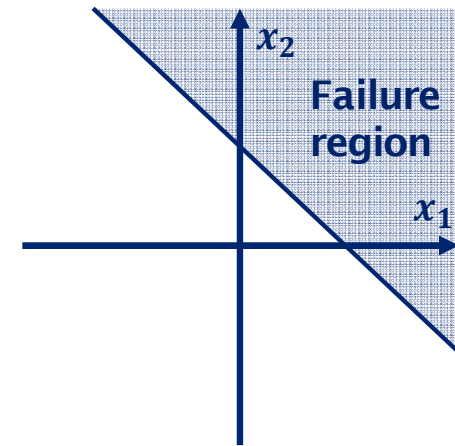
Hyper-plane-like failure boundary

Efficient algorithm for LARGE but SIMPLE system is required

Line sampling: suitable for simple failure boundary problem

1. **Initialize a sampling direction: α**

for i in 1 to N do

> 2. **Randomly generate line l_i such that $l_i \parallel \alpha$**
>
> 3. **Probe variability space along l_i**
>
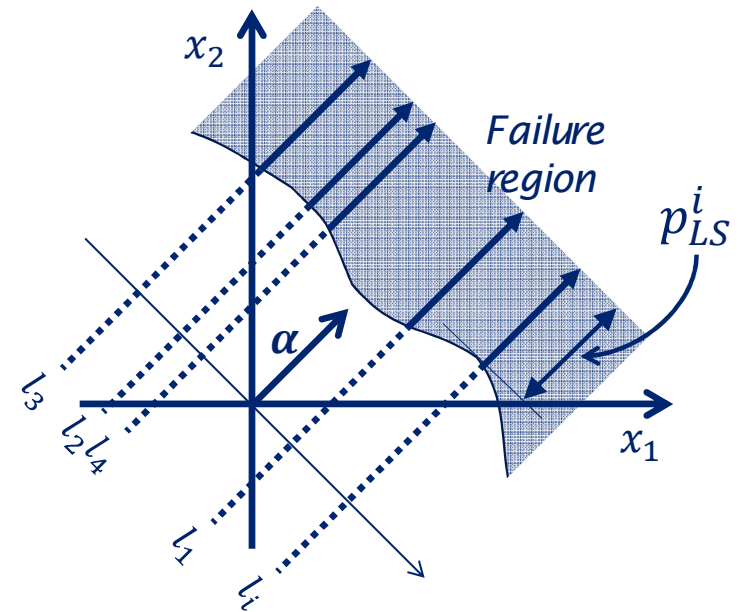> **Calculate failure probability when random variables are conditioned on line l_i:**
> $$p_{LS}^i = P(F|\boldsymbol{x} \text{ on } l_i)$$

end for

Contributions from all of lines are summed up to obtain failure probability: $\quad p_{\text{fail}} = \frac{1}{N}\sum_{i=1}^{N} p_{LS}^i$
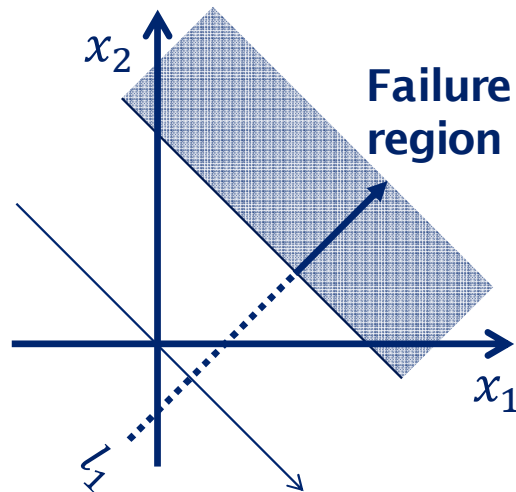
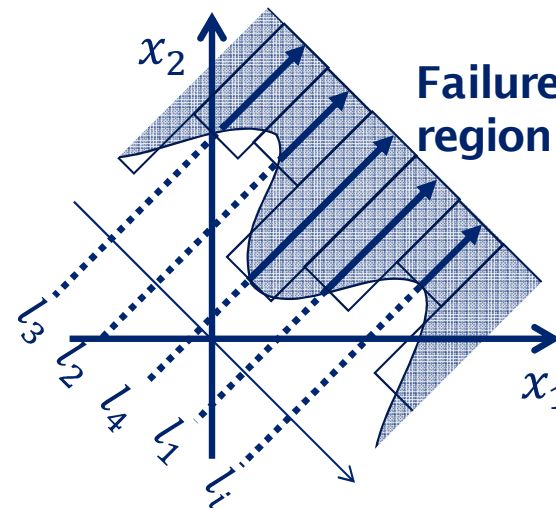**Probe variability space using LINES not POINTS**

Efficiency of line sampling

Shape of failure boundary have huge impact on sampling efficiency

➡️ **If failure boundary is more closer to hyperplane, line sampling can achieve better efficiency**

Single line is sufficient to probe failure region

Require large number of lines to probe failure region

Selecting sampling direction α

Direction α should be almost perpendicular to achieve good sampling efficiency

Almost linear relationship between signal propagation delay and variability can be assumed

➡ **α is approximated by gradient of signal propagation delay:**

$$s_d \approx \left.\frac{\partial y(x)}{\partial x_d}\right|_{x_d=0} \approx \frac{y(\Delta \cdot \mathbf{1}_d) - y(-\Delta \cdot \mathbf{1}_d)}{2\Delta}$$

$\boldsymbol{s} = (s_1, s_2, \cdots, s_D)$ *is normalized to obtain $\boldsymbol{\alpha}$:*

$\boldsymbol{\alpha} \approx \boldsymbol{s}/|\boldsymbol{s}|$

Failure region

Failure region

Bad direction deteriorates sampling efficiency
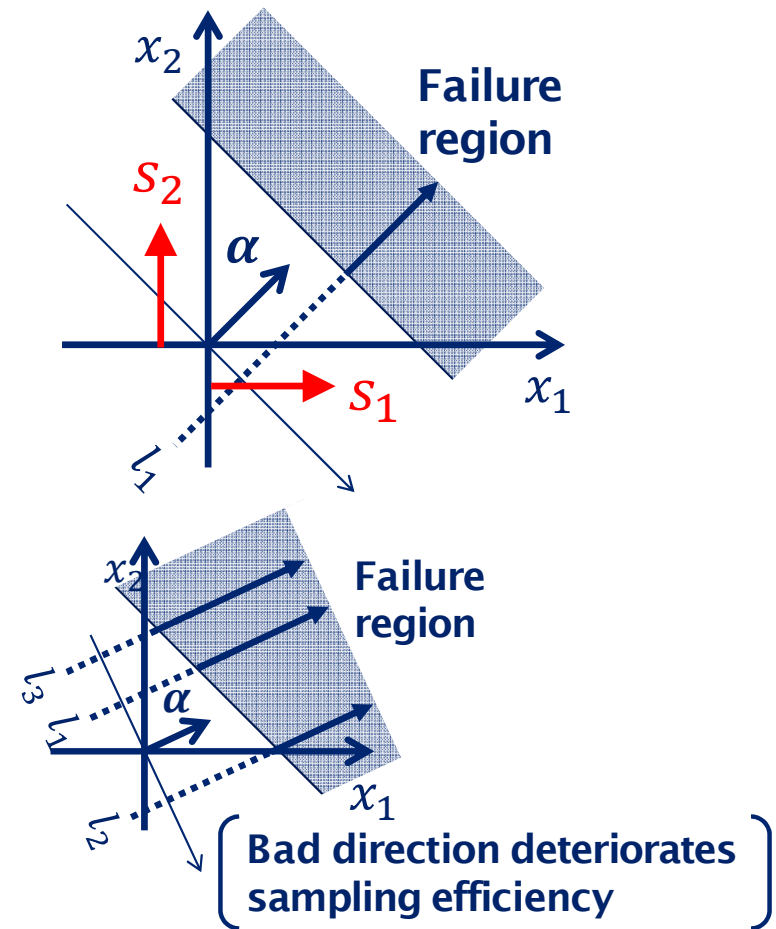
Experimental condition

- **Target circuit: ISCAS'85**

  c432 :   interrupt controller
  c499/c1355 : 32-bit SEC circuit
  c880 :   8-bit ALU
  c1908 :   16-bit SEC/DED circuit

  c2670 : 12-bit ALU and controller
  c3540 : 8-bit ALU
  c5315 : 9-bit ALU
  c6288 : 16x16 multiplier
  c7552 : 32-bit adder/comparator

- **Synthesized assuming a commercial 65-nm technology and critical path is extracted**

- **Random variations are introduced to threshold voltages and gate lengths so as to assume process variability**

Threshold voltages (V_{TH}): $\Delta V_{TH} \sim N(0, A_{V_{TH}}/$
$\sqrt{L \cdot W}$) Gate length (l_g): $\Delta l_g \sim N(0, 5 \times 10^{-9})$

$N(\mu, \sigma)$: Gaussian distribution
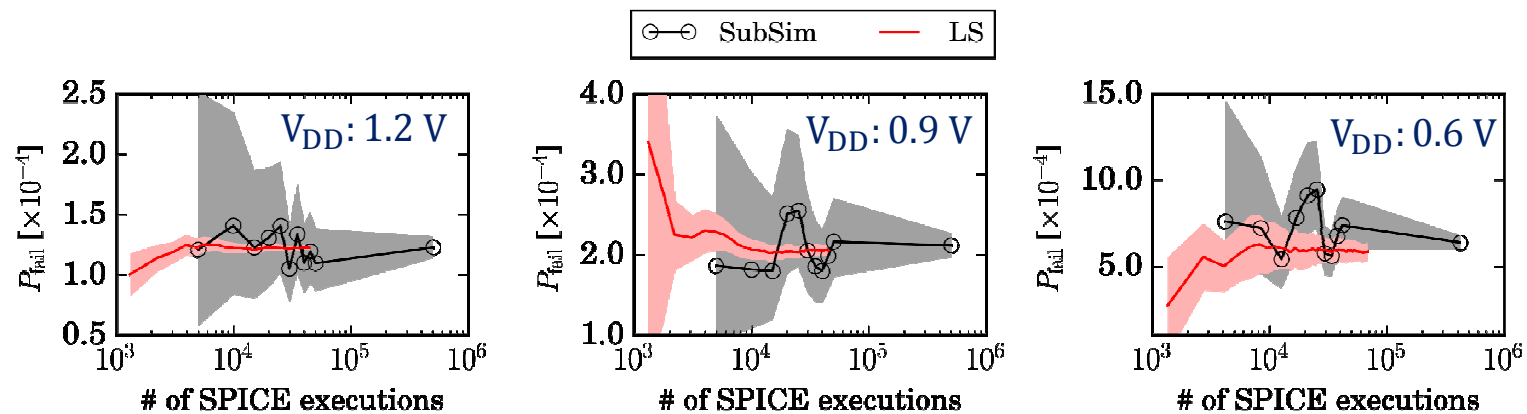*$A_{V_{TH}}$: Pelgrom coefficient*

Relationship between # of circuit sim. and estimated results are shown

Target circuit: c432

Red lines: Line sampling (LS)

Black lines: Subset simulation (SubSim)

Shaded regions show 95% of confidence interval


- **Both LS and SubSim converge to same result, indicating correctness of LS**
- **Even under low V_{DD} condition (less linearity, harsh condition for LS), LS converges faster than SubSim**

Relationship between # of sim. and estimation error is shown

Error is defined by: $\dfrac{\text{95\% of confidence interval}}{\text{Estimated failure probability}}$

1/14 (V_{DD} is set to 0.6V) to
1/300 (V_{DD} is set to 1.2V)
**reduction of sim. runs is achieved
without deteriorating accuracy**
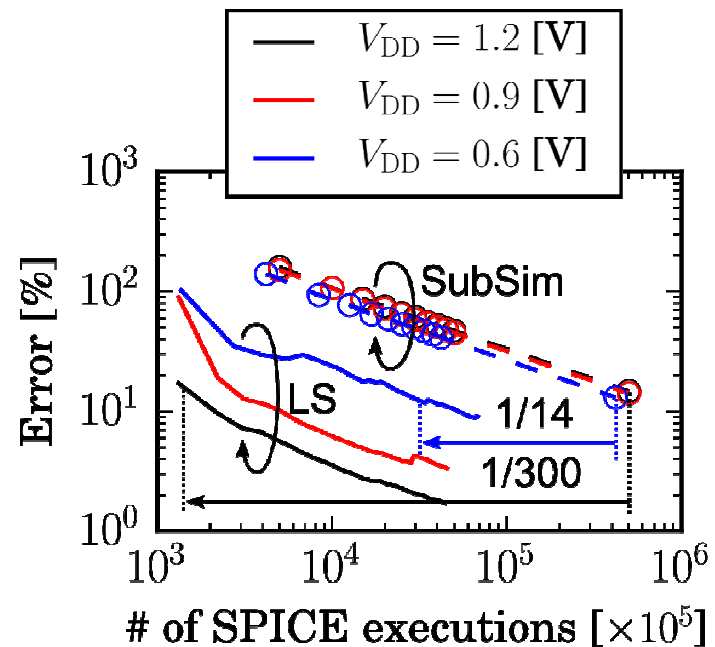
Experimental results for other circuits

# of sim. is set to 10k
i.e. achievable accuracy with same calculation time is compared

Circuit	Subet Simulation		Line sampling		A/B	Dim
	Pfail	Error (A)	Pfail	Error (B)		
C432	1.35e-4	106	1.28e-4	4.18	25.4	616
c499	7.72e-5	112	1.10e-4	2.62	42.7	468
C880	1.06e-4	110	1.19e-4	2.53	43.5	608
C1355	1.54e-4	108	1.16e-4	2.68	40.3	472
C1908	9.45e-5	110	9.08e-5	2.18	50.5	584
C2670	9.97e-5	111	1.04e-4	3.87	28.7	548
C3540	1.26e-4	109	9.92e-4	3.98	27.4	804
C5315	1.13e-4	108	1.09e-4	4.60	23.5	596
C6288	9.57e-5	110	9.60e-5	8.06	13.6	1984
C7552	9.99e-5	111	8.82e-5	4.48	24.8	2536

13.6 times to 50.5 times more accuracy can be achieved

Summary

Massive parallel architecture attracts increasing attention
- **Facing similar problem as SRAM design (high-level of reliability is required for each core)**
- **Accurate timing yield estimation is thus required**

Our proposal: Application of line sampling (LS)
- **LS perfectly fits the analysis of simple but large system**

Numerical experiment using ISCAS'85 c432 showed that…

LS achieved **14** **times (when V_{DD} is 0.6V)**

300 **times (when V_{DD} is 1.2V)**

faster convergence compared with Subset simulation