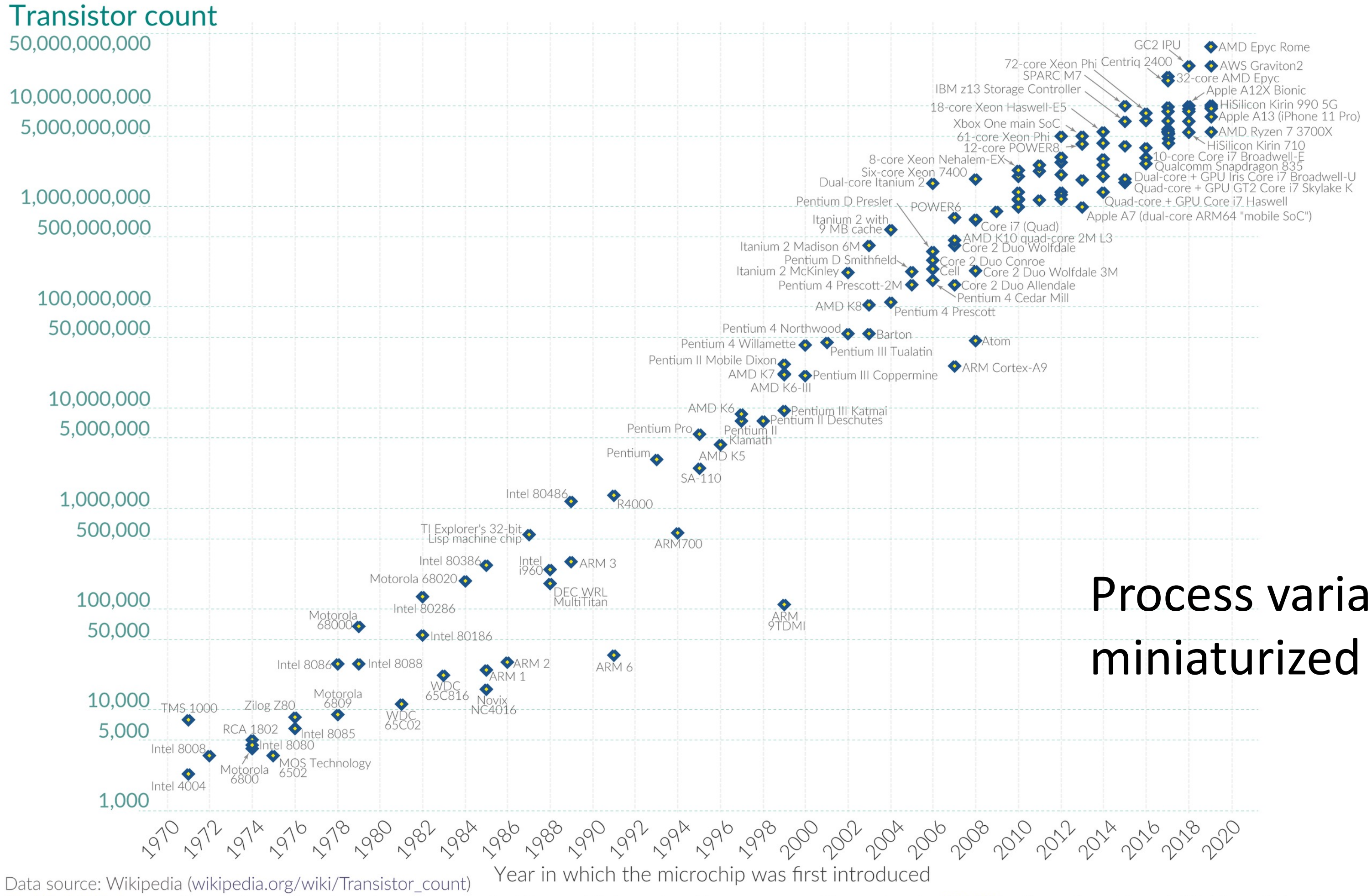
Distributionally Robust Circuit Design Optimization under Variation Shifts

Yifan Pan[*†], Zichang He[*†], Nanlin Guo[#], Zheng Zhang[*]

[*]University of California, Santa Barbara
[#]Fudan University
[†]Equal contributions

ICCAD, Oct. 2023

Process variation



Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years.
This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

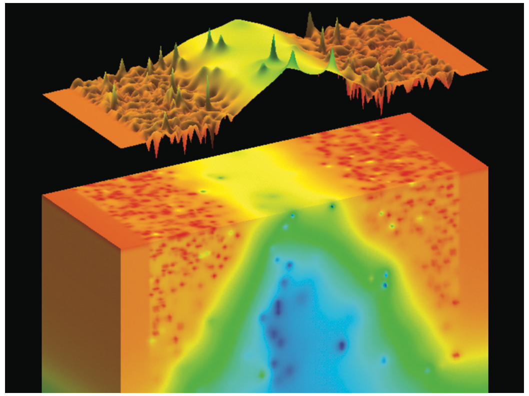
Our World in Data

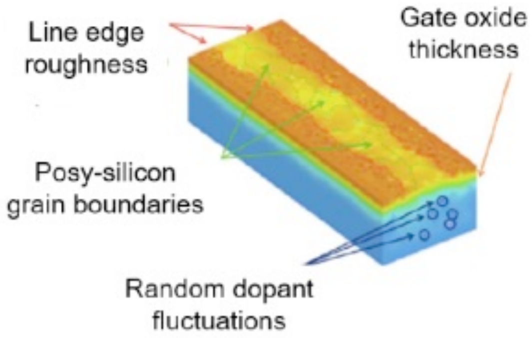
Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)

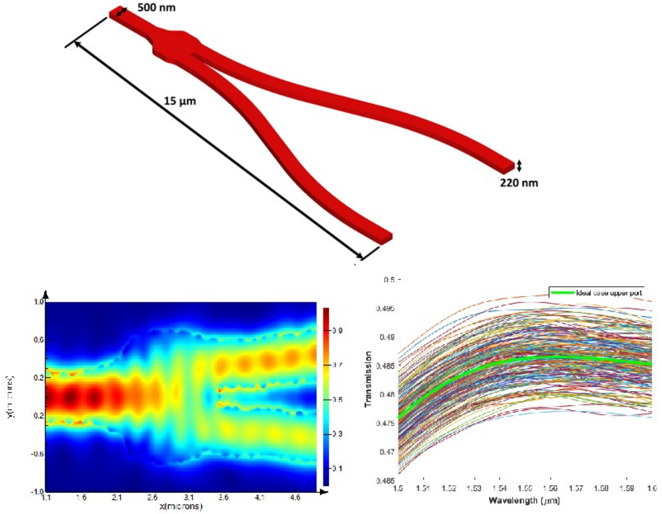
OurWorldinData.org – Research and data to make progress against the world's largest problems.

Process variation in the miniaturized system

Process variation







Sources of process variations: manufacturing & environment

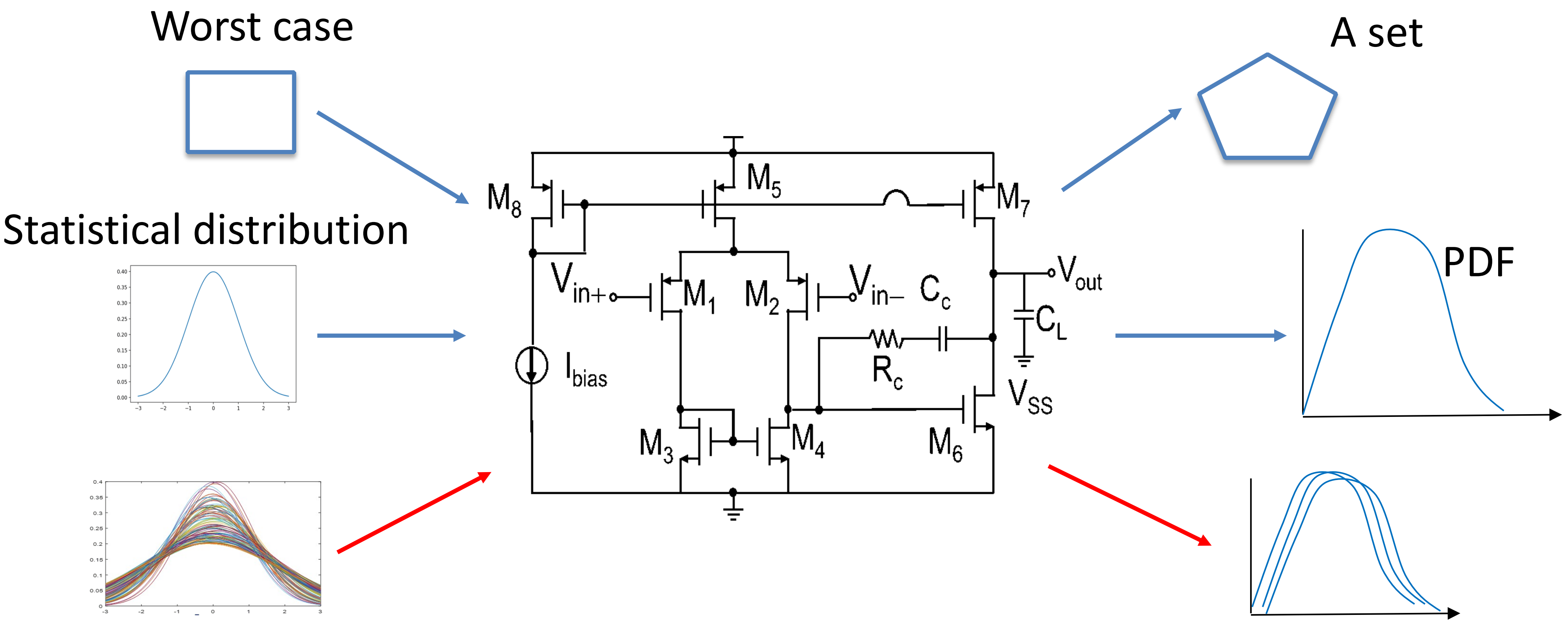
Importance

- ☐ Device performance, quality, and yield
- ☐ More significant as the scales of devices decrease

Rethink process variation modeling

Description of variation

Performance output



Worst case

Statistical distribution

A set

PDF

Reality: statistical but unknown distribution

Possible PDFs

Why? Case 1: poor data quality

A typical process of extracting process variation models

Testing chip design & fab

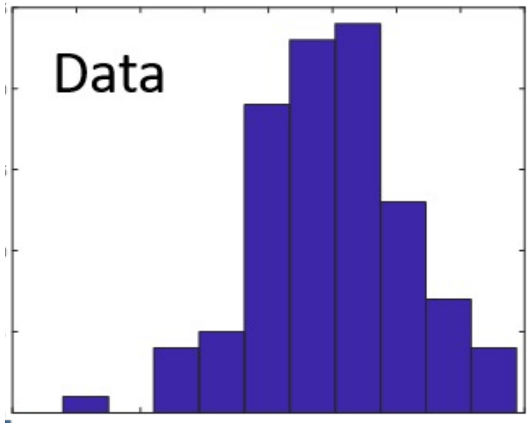
Data measurement

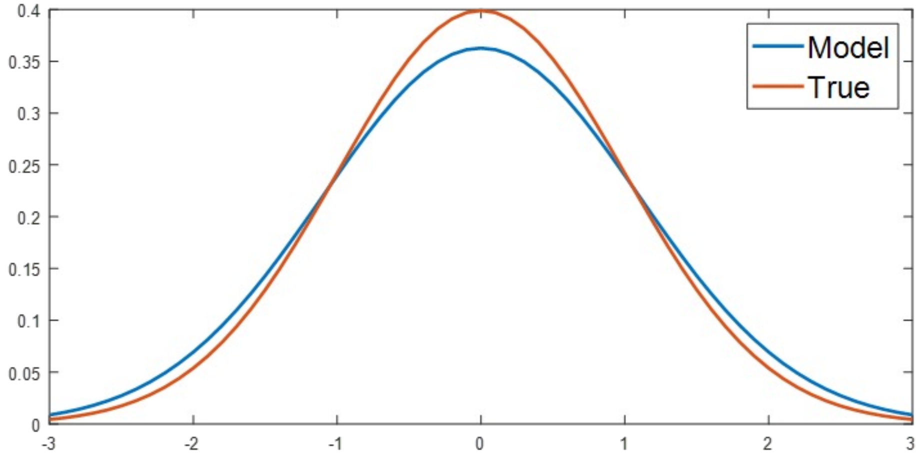
Statistical model (e.g., PDF)

Why? Case 1: poor data quality

A typical process of extracting process variation models

Testing chip design & fab → Data measurement → Statistical model (e.g., PDF)



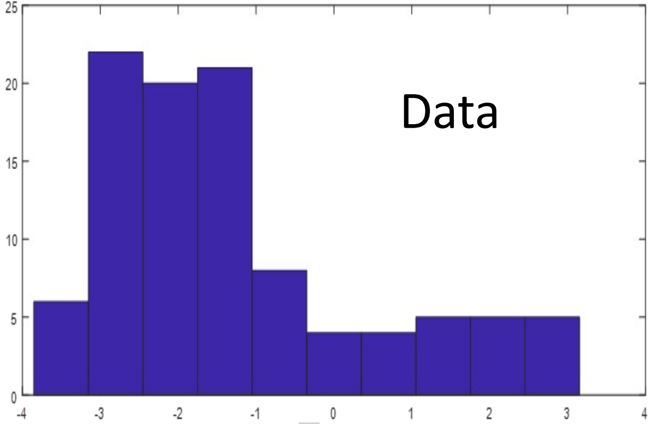


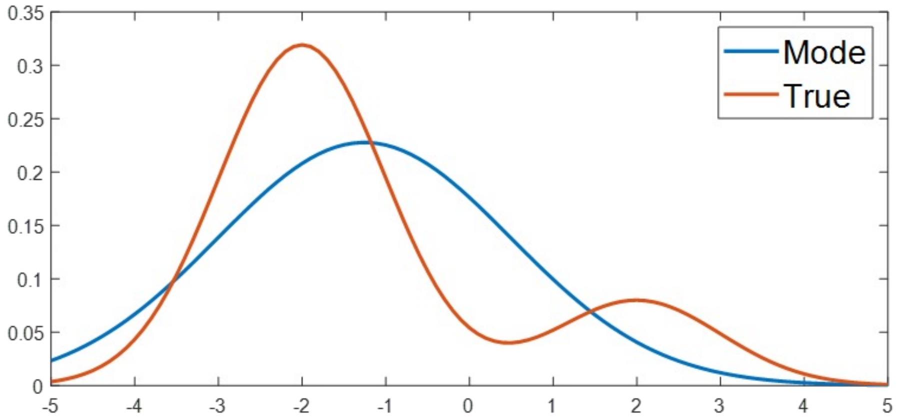
- Data is limited and noisy

Why? Case 2: model misfit

A typical process of extracting process variation models

Testing chip design & fab → Data measurement → Statistical model (e.g., PDF)

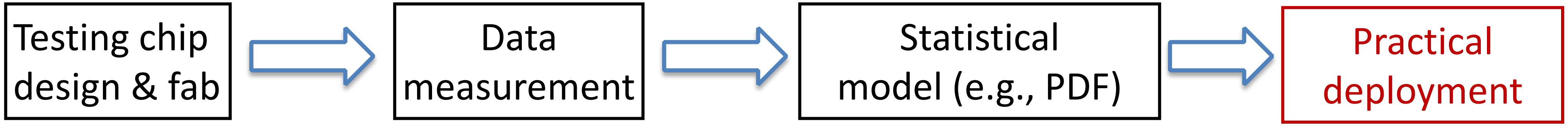




- The chosen statistical model is over-simplified

Why? Case 3: time shift

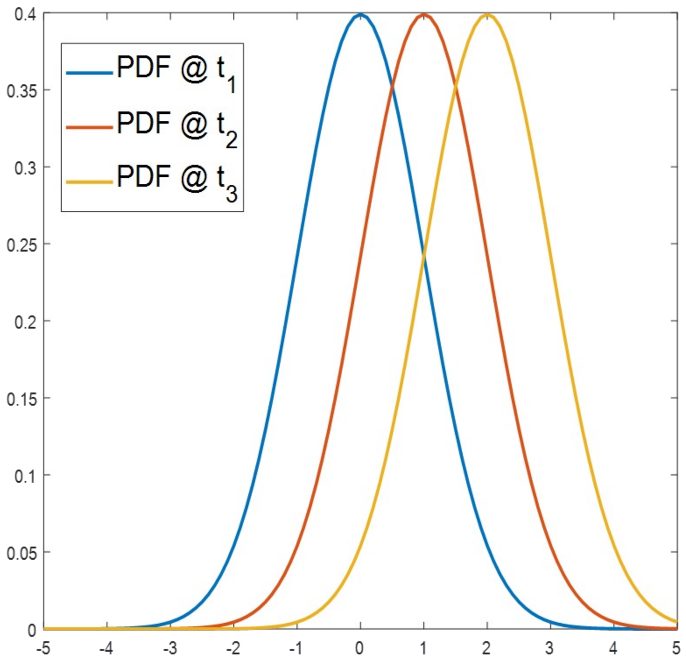
A typical process of extracting process variation models


Testing chip design & fab

Data measurement

Statistical model (e.g., PDF)

Practical deployment



- Data is limited and noisy
- The chosen statistical model is over-simplified
- PDF can change over time

Three cases can happen together, we name them *Variation shifts*

Shift-aware optimization

Traditional stochastic circuit optimization

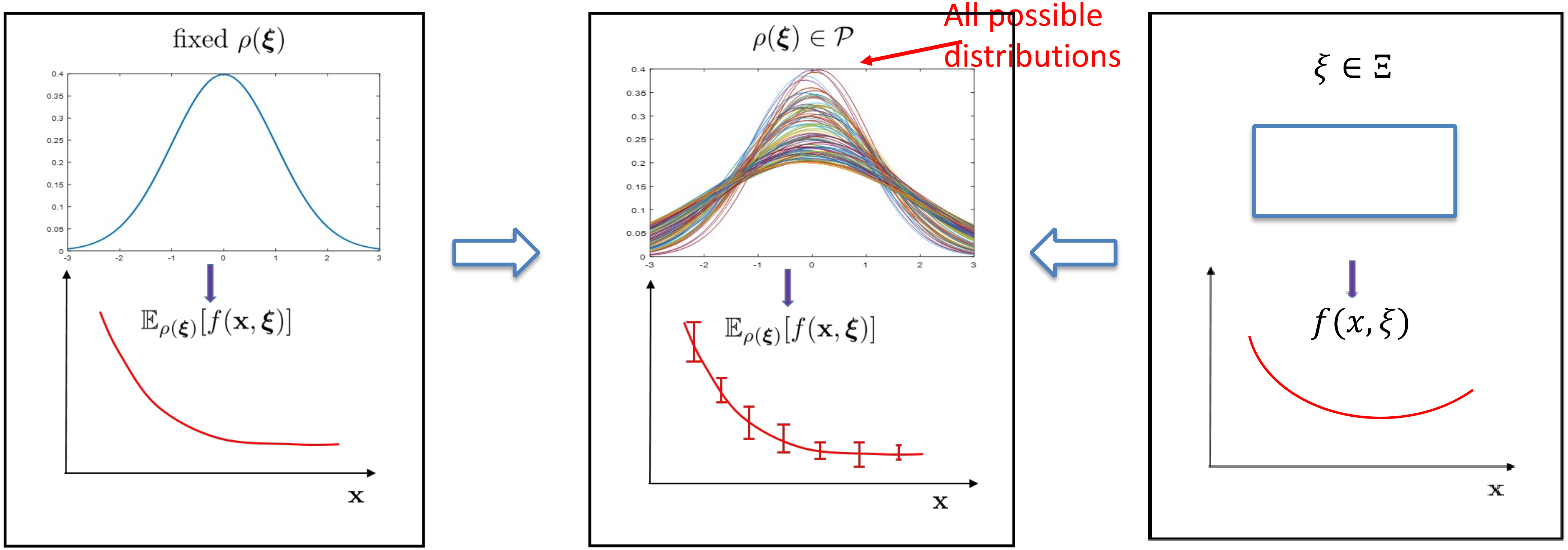
$$\min_{\mathbf{x}\in\mathcal{X}} \mathbb{E}_{\rho_0(\boldsymbol{\xi})}\left[f(\mathbf{x},\boldsymbol{\xi})\right]$$

Distributionally robust circuit optimization (DRO)

$$\min_{\mathbf{x}\in\mathcal{X}} \sup_{\rho(\boldsymbol{\xi})\in\mathcal{P}} \mathbb{E}_{\rho(\boldsymbol{\xi})}\left[f(\mathbf{x},\boldsymbol{\xi})\right]$$

Traditional robust circuit optimization

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\boldsymbol{\xi}\in\Xi} \left[f(\mathbf{x},\boldsymbol{\xi})\right]$$



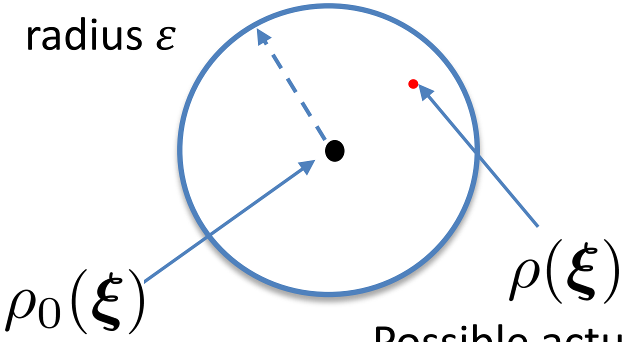
How to solve DRO?

Key: a careful modeling of the distributional uncertainty set

$$\mathcal{P} := \mathcal{B}(\rho_0) = \{\rho : \mathcal{D}(\rho_0, \rho) \leq \varepsilon\}$$

$$\min_{\mathbf{x}\in\mathbf{X}} \sup_{\rho(\boldsymbol{\xi})\in\mathcal{B}_\varphi(\rho_0)} \mathbb{E}_{\rho(\boldsymbol{\xi})}[f(\mathbf{x}, \boldsymbol{\xi})]$$

Distribution ball



radius ε

$\rho_0(\boldsymbol{\xi})$

Nominal PDF
(e.g., Gaussian)

$\rho(\boldsymbol{\xi})$

Possible actual PDF
(e.g., Gaussian mixture)

Degeneration

- Robust optimization: $\rho(\xi)$ is a delta function
- Stochastic programming: $\epsilon \to 0$

Challenges

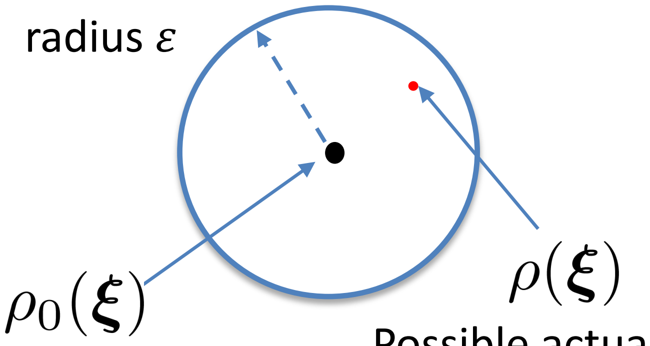
- The inner problem is non-trivial to solve
- Two-level problem is computationally heavy
- No analytical function for circuit performance

How to solve DRO?

Key: a careful modeling of the distributional uncertainty set

$$\mathcal{P} := \mathcal{B}(\rho_0) = \{\rho : \mathcal{D}(\rho_0, \rho) \leq \varepsilon\}$$

Distribution ball



Nominal PDF
(e.g., Gaussian)

Possible actual PDF
(e.g., Gaussian mixture)

$$\min_{\mathbf{x}\in\mathbf{X}} \sup_{\rho(\boldsymbol{\xi})\in\mathcal{B}_\varphi(\rho_0)} \mathbb{E}_{\rho(\boldsymbol{\xi})}[f(\mathbf{x}, \boldsymbol{\xi})]$$

$\Updownarrow$ $\mathcal{D}(\rho_0, \rho)$ as χ^2 divergence

$$\min_{\mathbf{x}\in\mathbf{X}} \mathbb{E}_{\rho_0(\boldsymbol{\xi})}[f(\mathbf{x}, \boldsymbol{\xi})] + \sqrt{\varepsilon \cdot \mathrm{Var}_{\rho_0(\boldsymbol{\xi})}[f(\mathbf{x}, \boldsymbol{\xi})]}$$

How to solve DRO?

- Two-level min-max → Single minimization (exact reformulation)
- The regularization parameter ε is physically meaningful
- Works for other φ-divergence measure

$$\min_{\mathbf{x}\in\mathbf{X}} \sup_{\rho(\boldsymbol{\xi})\in\mathcal{B}_{\varphi}(\rho_0)} \mathbb{E}_{\rho(\boldsymbol{\xi})}[f(\mathbf{x},\boldsymbol{\xi})]$$

$$\Big\Updownarrow \quad \mathcal{D}(\rho_0,\rho) \text{ as } \chi^2 \text{ divergence}$$

$$\min_{\mathbf{x}\in\mathbf{X}} \mathbb{E}_{\rho_0(\boldsymbol{\xi})}[f(\mathbf{x},\boldsymbol{\xi})] + \sqrt{\varepsilon\cdot\mathrm{Var}_{\rho_0(\boldsymbol{\xi})}[f(\mathbf{x},\boldsymbol{\xi})]}$$

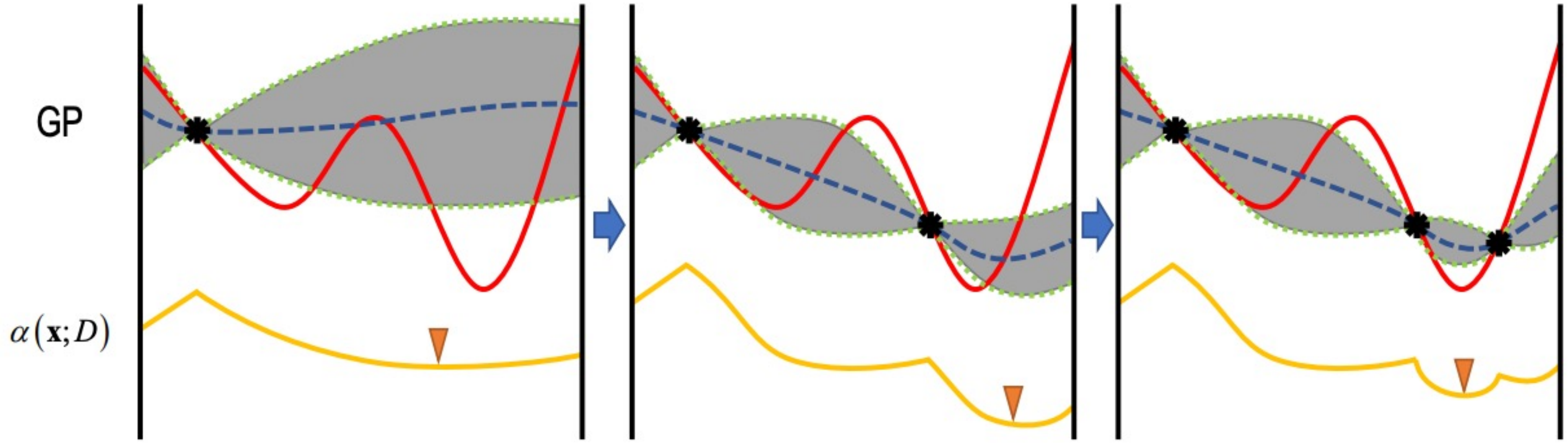
Penalizing the design that has a large variance under the nominal PDF ρ_0

Can be solved by many optimizers

- Bayesian optimization
- Gradient-based optimizers

Bayesian optimization (BO) solver

$$\min_{\mathbf{x}\in\mathbf{X}} \; \mathbb{E}_{\rho_0(\boldsymbol{\xi})}[f(\mathbf{x},\boldsymbol{\xi})]$$



Initial probabilistic surrogate model → Query based on uncertainty estimation → Update the surrogate model

Workflow: DRBO solver

$$\min_{\mathbf{x}\in\mathbf{X}} \ \mathbb{E}_{\rho_0(\boldsymbol{\xi})}[f(\mathbf{x},\boldsymbol{\xi})] + \sqrt{\varepsilon \cdot \mathrm{Var}_{\rho_0(\boldsymbol{\xi})}[f(\mathbf{x},\boldsymbol{\xi})]}$$

Step 1: Build a Gaussian process regression model

Step 2: Design exploration through minimizing the lower confidence bound (LCB) acquisition function

$$A(\mathbf{x}) := \frac{1}{L}\sum_{l=1}^{L}[\mu(\mathbf{x},\boldsymbol{\xi}^l) - \sqrt{\beta}\sigma(\mathbf{x},\boldsymbol{\xi}^l)] \boxed{+ \sqrt{\frac{\varepsilon}{L}\sum_{l=1}^{L}\left(\mu(\mathbf{x},\boldsymbol{\xi}^l) - \bar{\mu}\right)^2},}$$

Penalizing the design that has a large variance under the nominal PDF ρ_0

Step 3: If not converge, add samples of ξ, return to step 1

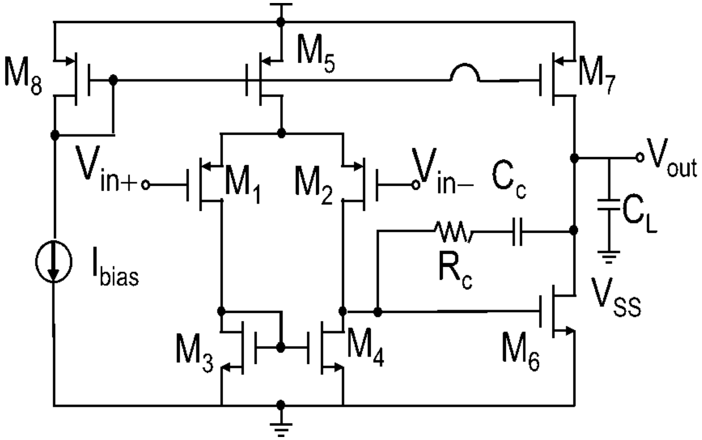
Demonstration in yield-aware optimization

Goal: under shifted variations, to ensure both high performance and high yield

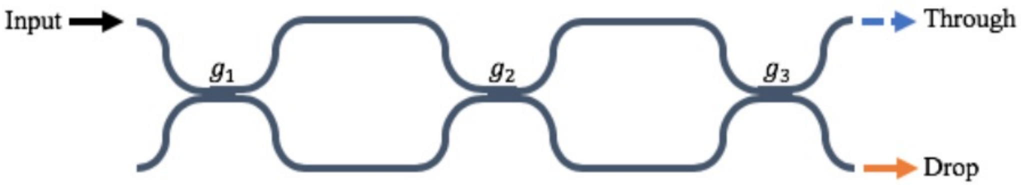
$$f(\mathbf{x}, \boldsymbol{\xi}) = f_{\text{obj}}(\mathbf{x}, \boldsymbol{\xi}) + \lambda I(\mathbf{x}, \boldsymbol{\xi})$$

Loss describing performance metric

Indicator function describing circuit failure (thus yield)



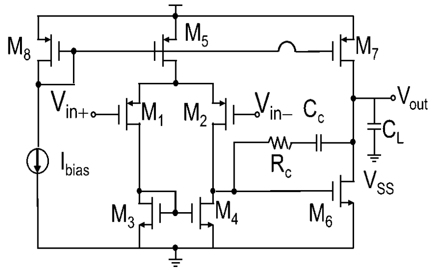
Two-stage amplifier



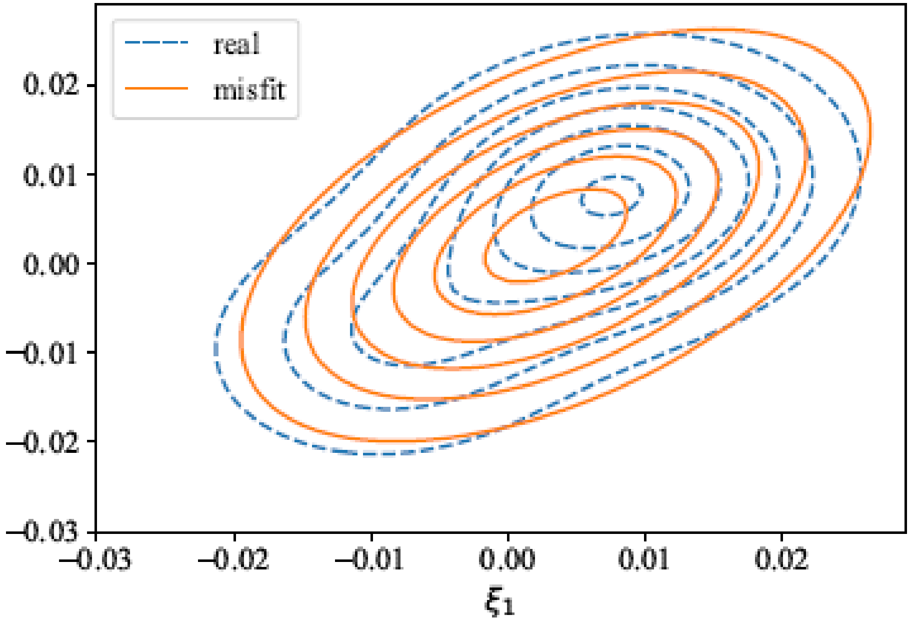
Mach-Zehnder interferometer

Although different circuits, share the same formulation

Demonstration in yield-aware optimization



Robustness against misfit of variation models & noisy data



Misfit model: Gaussian
Actual: Gaussian mixture

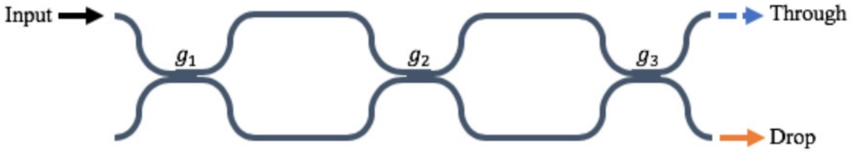
Misfit increases

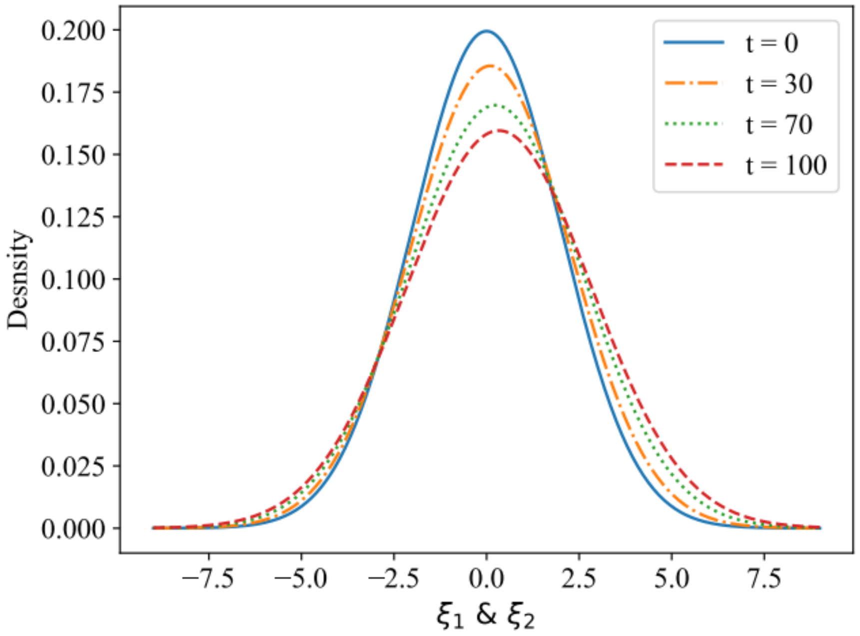
mixture rate	Method	Perf loss	yield
3:7	proposed	0.2195	91.63%
	LCB	0.2373	85.74%
5:5	proposed	0.2285	88.92%
	LCB	0.2343	86.85%
7:3	proposed	0.2249	87.73%
	LCB	0.2287	87.11%

LCB: optimize with fixed misfit model

The proposed method provides better yield and performance when the actual PDF of process variations differ from the given one.

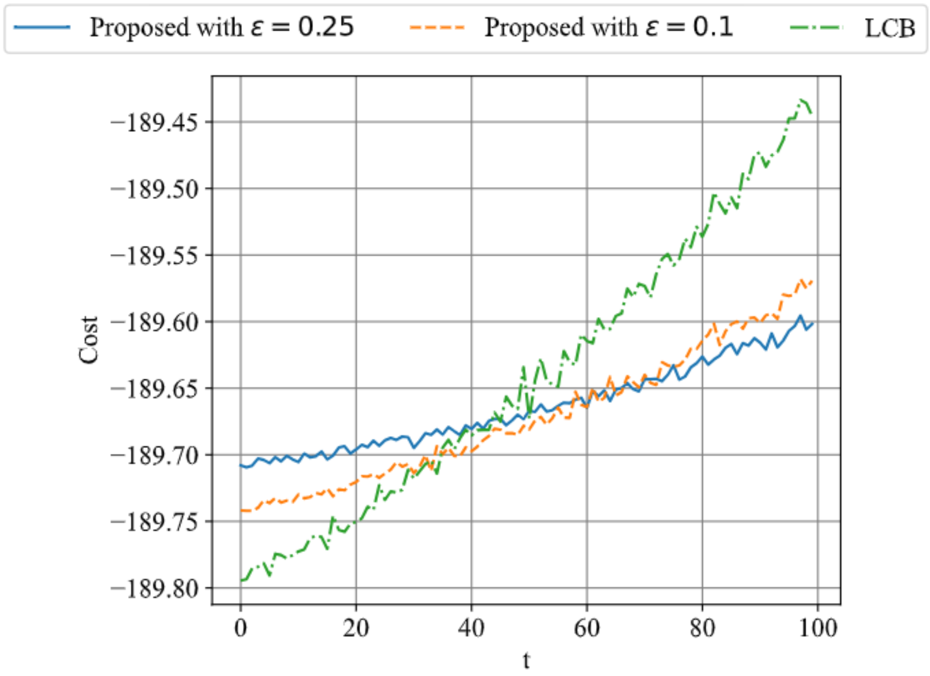
More numerical results (Time shift)





Practical variations gradually diverge from nominal distribution with time

LCB: optimize with fixed misfit model



- The LCB performs the best when the shift is not significant ($t \leq 40$)
- When $t > 40$, larger error bounds (radius ε) provide more robust solutions under distributional shift

Take-home message

- The practical issues of variation shifts: data quality, model misfit, time shift
- Distributionally robust formulation for optimization under shifts
- Efficient Bayesian optimization solver for the DRO formulation, showing robustness against shifted variations

Open questions:

- Higher-dimensional cases
- More domain customized modeling of the variations