A Novel and Efficient Bayesian Optimization Approach for Analog Designs with Multi-Testbench

Jingyao Zhao[1], Changhao Yan[1], Zhaori Bi[1],
Fan Yang[1], Xuan Zeng[1] and Dian Zhou[2]

[1] Fudan University, Shanghai, P. R. China

[2] University of Texas at Dallas, Dallas, USA

Resume

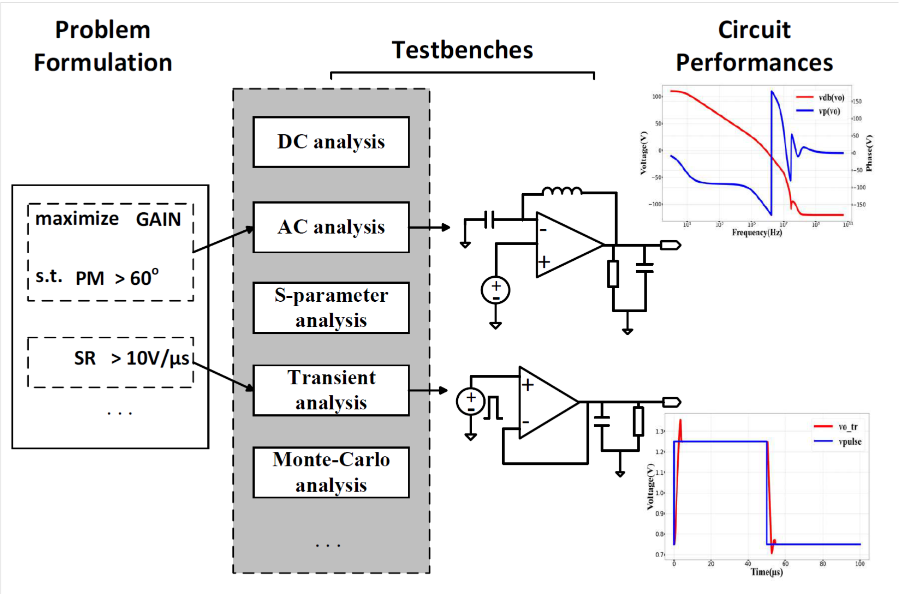
- Name
  - Jingyao Zhao
- Affiliation
  - Microelectronics Department, Fudan University, Shanghai, China
- Resume
  - I received the B.S. degree in microelectronics from Fudan University, Shanghai, China, in 2019. She is currently pursuing master's degree with State Key Laboratory of Application Specific Integrated Circuits and System, Microelectronics Department, Fudan University, Shanghai, China. Her current research interests include analog circuit design automation and optimization.

- Problem Definition
    - Multi-Testbench Optimization Problem
    - Circuit Sizing Formulation
- Review of Analog Circuit Optimization Algorithms
    - Model-based optimization algorithms
    - Simulation-based optimization algorithms
    - Bayesian Optimization
- Proposed Method
    - Predictive Entropy Search with Constraints
    - Feasibility Regions Exploration of Constraints
    - Multi-modal Optimization of FEI
- Experimental Results
    - Two-Stage Operational Amplifier
    - Low-Power Amplifier
- Conclusion

Multi-Testbench Optimization Problem

- Simulation testbench
  - A tuple of peripheral test circuit, excitation source and simulation type is termed a *testbench*
- Multi-testbench optimization problem
  - Target: find the design parameters that make the analog circuit work fine under several or dozens of different testbenches
  - Difficulties
    - Equations and derivatives are inaccessible
    - Circuit simulations are time-consuming

Operational amplifier
design with Multi-testbench



Circuit Sizing Formulation

- Traditional analog circuit sizing formulation

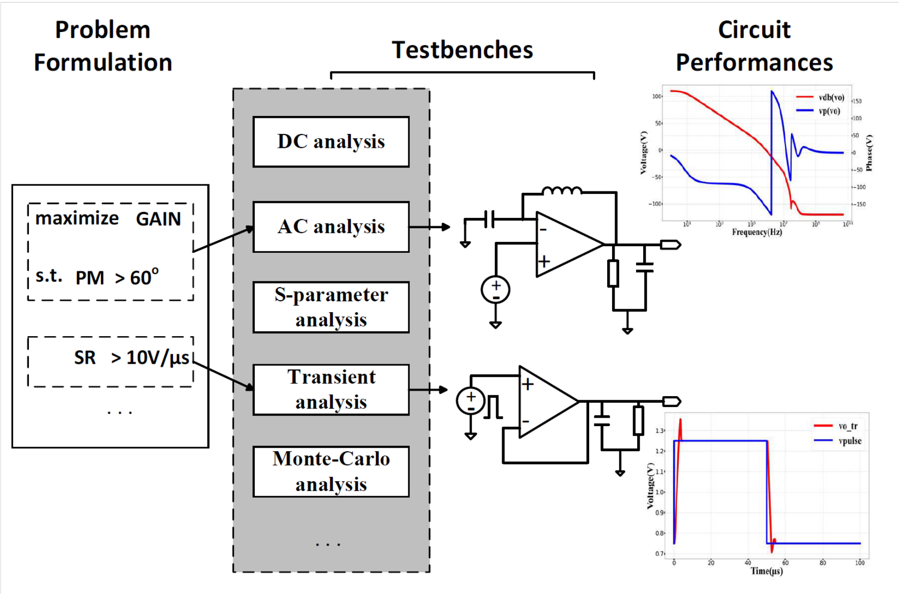
$$\begin{aligned} \textbf{minimize} \quad & f(x) = c_0(x) \\ \textbf{s.t.} \quad & c_i(x) > 0, \qquad i \in [1, N_c] \end{aligned}$$

- Multi-testbench optimization formulation

$$\begin{aligned} \textbf{minimize} \quad & f(x) \\ \textbf{s.t.} \quad & c_i^j(x) > 0\ , j \in [1, |F_t|], i \in [1, |\mathcal{F}|] \end{aligned}$$

- Partition performance set into $\mathcal{F} = \{F_t, t \in [1, |\mathcal{F}|]\}$
- F_t contains the circuit performances from the t-th testbench
- $c_t^j(x)$ is the j-th performance in the t-th testbench

Operational amplifier design with Multi-testbench



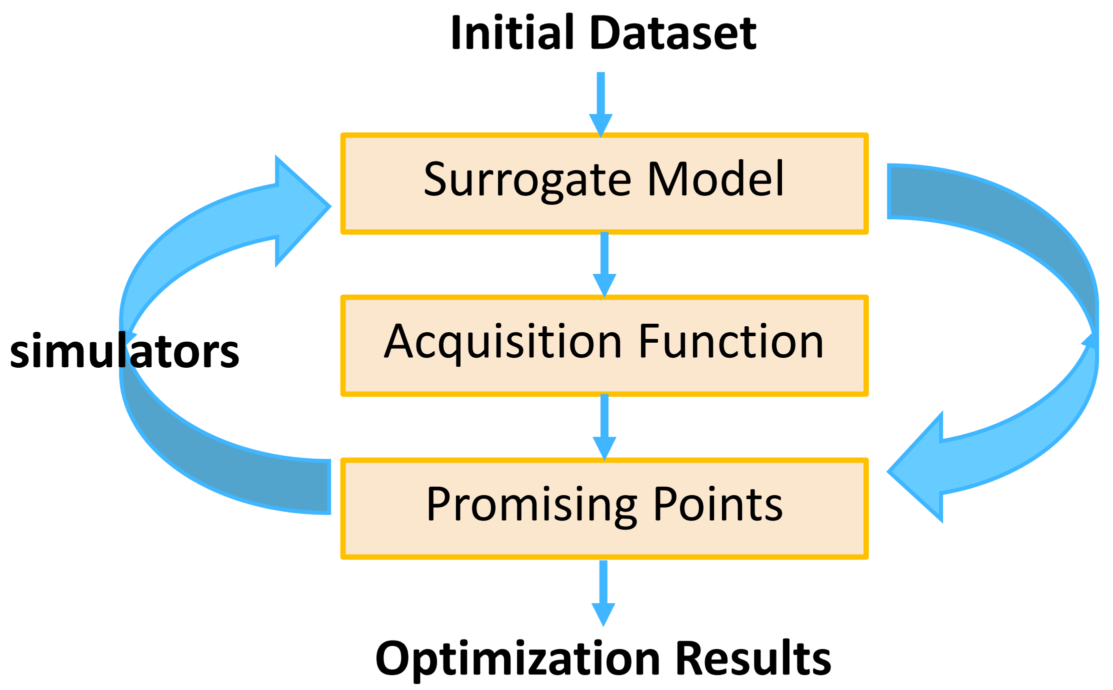
Review of Analog Circuit Optimization Algorithms

- Model-based optimization algorithms
  - Geometric programming [1]
    - Posynomial approximation for circuit performances
    - Guarantee transforming to a convex problem which has global optimum
  - Cannot guarantee the accuracy of the models especially for the large-scale circuits
- Simulation-based optimization algorithms
  - Differential evolutionary [2]
    - Add penalty functions of constraints into the objective
    - Transform the constrained optimization into an unconstrained one
  - Particle swarm optimization [3]
  - Simulation-based optimization is generally accurate, but convergence rate of heuristic algorithms is low

[1] M. del Mar Hershenson, "Design of pipeline analog-to-digital converters via geometric programming," in Proceedings of the 2002 IEEE/ACM international conference on Computer-aided design, 2002, pp. 317–324.
[2] B. Liu, Y. Wang, Z. Yu, L. Liu, M. Li, Z. Wang et al., "Analog circuit optimization system based on hybrid evolutionary algorithms," Integration, vol. 42, no. 2, pp. 137–148, 2009.
[3] R. Vural and T. Yildirim, "Analog circuit sizing via swarm intelligence," AEU-International journal of electronics and communications, vol. 66, no. 9, pp. 732–740, 2012.

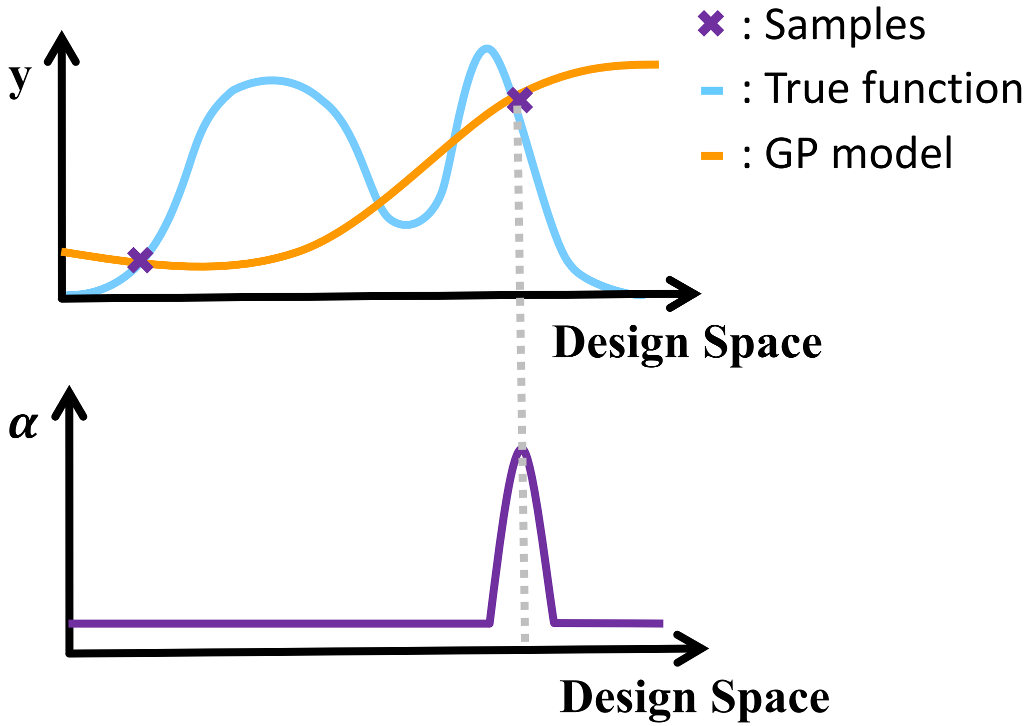
Bayesian Optimization

- Two components
  - Surrogate model
    - provide the posterior prediction and corresponding uncertainty
  - Acquisition function
    - guide proposing the next data points of simulation during the optimization



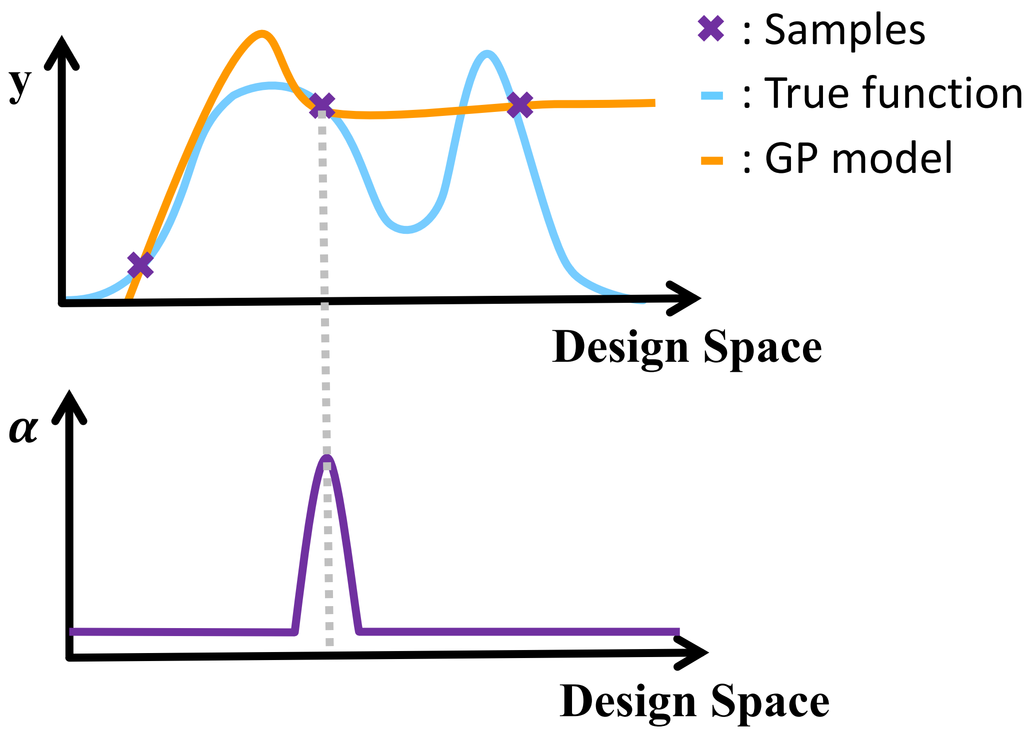
Bayesian Optimization

- Gaussian Process (GP) surrogate model
  - The prediction of GP is a normal distribution $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\sigma}^2(\boldsymbol{x}))$
  - $\boldsymbol{\mu}(\boldsymbol{x}_0)$ can be viewed as the predictive mean for a new sample $\boldsymbol{x}_0$
  - Predictive variance $\boldsymbol{\sigma}^2(\boldsymbol{x}_0)$ represents the uncertainty of prediction
- Traditional acquisition function
  - Evaluate both objective and constraints at the same selected point by maximizing acquisition function $\boldsymbol{\alpha}$



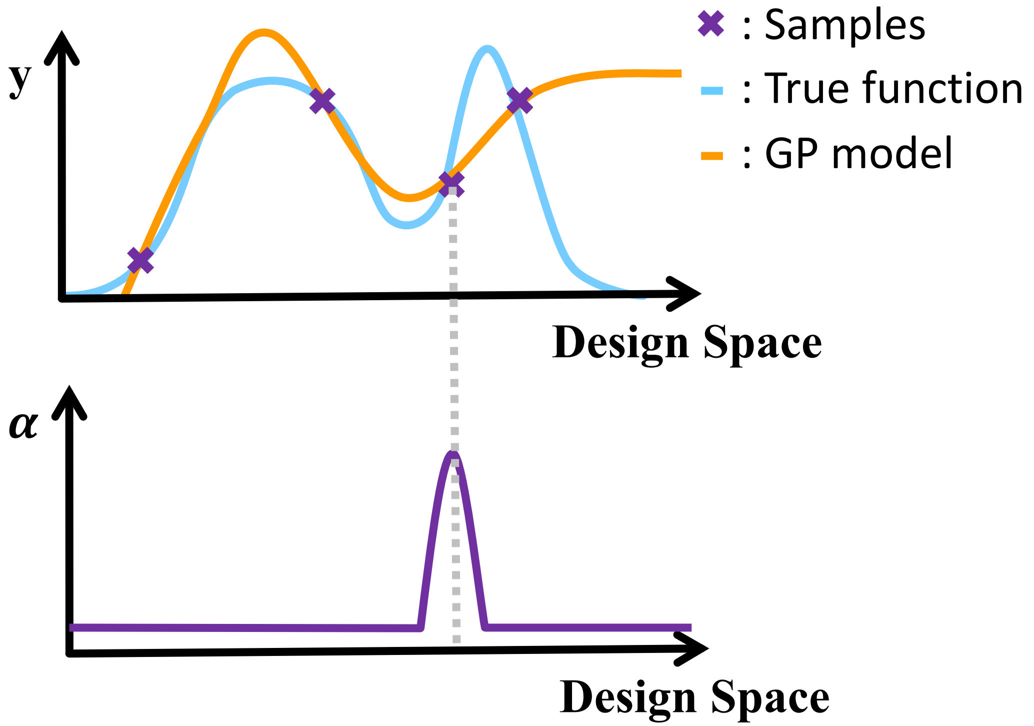
Bayesian Optimization

- Gaussian Process (GP) surrogate model
  - The prediction of GP is a normal distribution $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\sigma}^2(\boldsymbol{x}))$
  - $\boldsymbol{\mu}(\boldsymbol{x_0})$ can be viewed as the predictive mean for a new sample $\boldsymbol{x_0}$
  - Predictive variance $\boldsymbol{\sigma}^2(\boldsymbol{x_0})$ represents the uncertainty of prediction
- Traditional acquisition function
  - Evaluate both objective and constraints at the same selected point by maximizing acquisition function $\boldsymbol{\alpha}$



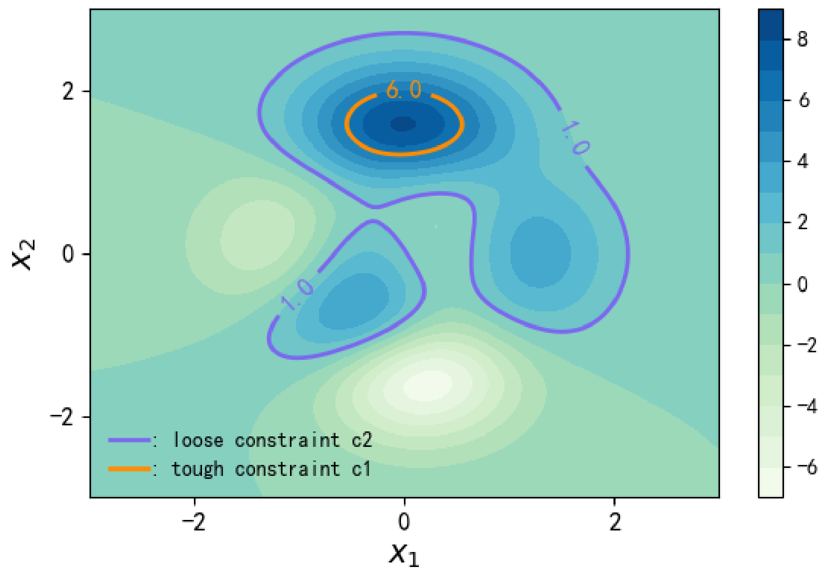
Bayesian Optimization

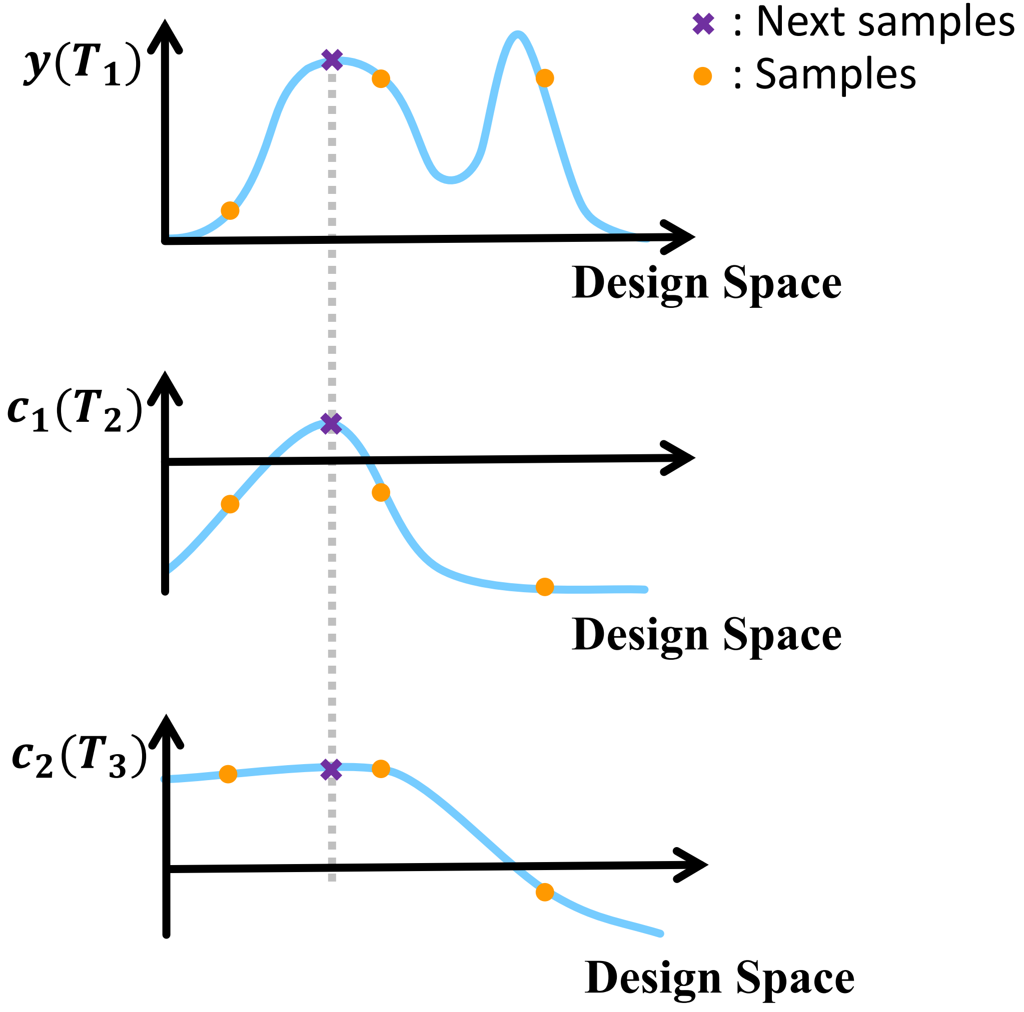
- Gaussian Process (GP) surrogate model
  - The prediction of GP is a normal distribution $y \sim N(\mu(x), \sigma^2(x))$
  - $\mu(x_0)$ can be viewed as the predictive mean for a new sample x_0
  - Predictive variance $\sigma^2(x_0)$ represents the uncertainty of prediction
- Traditional acquisition function
  - Evaluate both objective and constraints at the same selected point by maximizing acquisition function α



Multi-testbench Bayesian Optimization

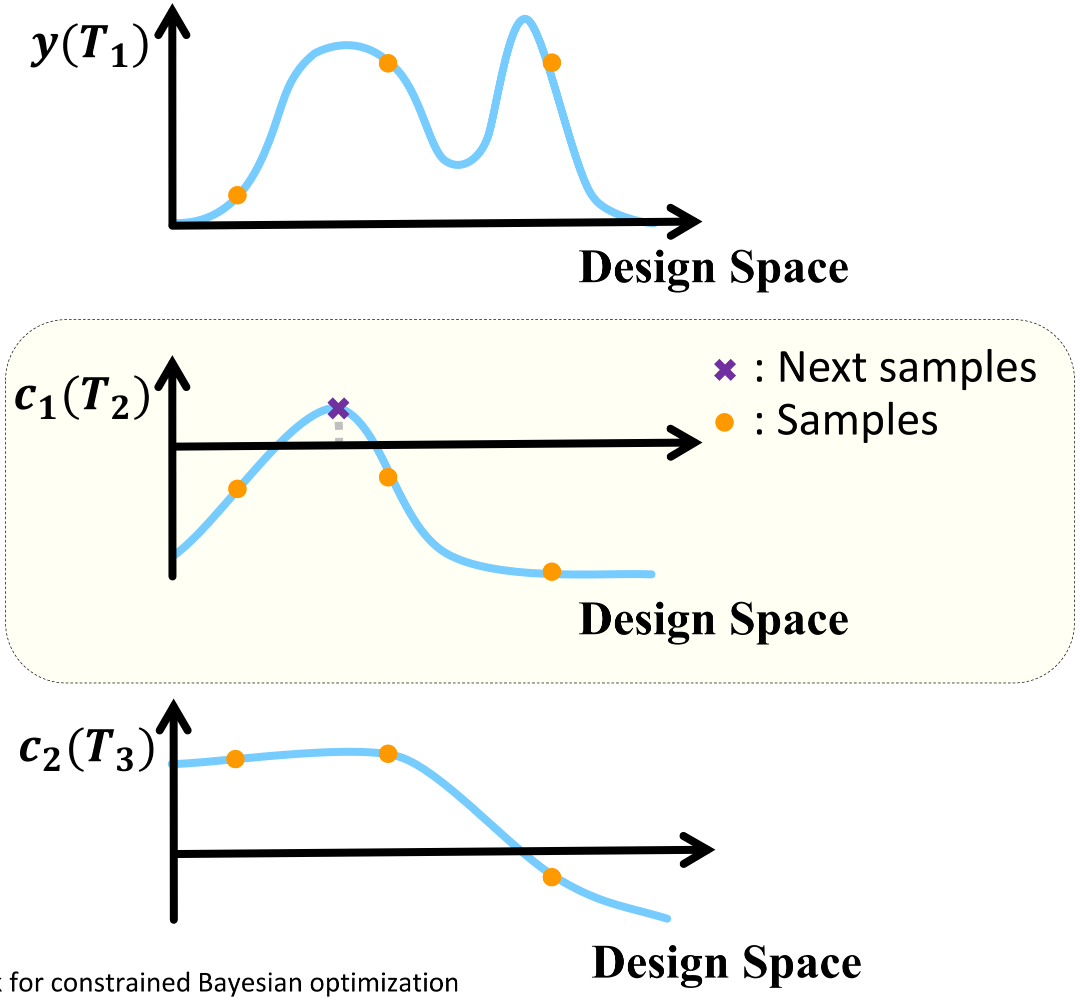
- Motivation
  - In optimization, the constraints and the objective are not equally important
  - Traditional acquisition function indiscriminately run all testbenches for a new design, which is time-consuming





Multi-testbench Bayesian Optimization

- How to select testbench to be simulated
  - Only simulate the most *important* testbench at each iteration for efficiency consideration
- Predictive Entropy Search with Constraints (PESC) [4]
  - PESC, i.e. *information gain*, is applied as the criterion of importance for each testbench
  - We only choose the testbench with maximal information gain for finding the global optimum $\boldsymbol{x}_*$



[4] J. M. Hern´andez-Lobato, M. A. Gelbart, R. P. Adams, M. W. Hoffman, and Z. Ghahramani, “A general framework for constrained Bayesian optimization using information-based search,” The Journal of Machine Learning Research, vol. 17, no. 1, pp. 5549–5601, 2016.

Acquisition Function of PESC

- PESC concept
  - Information gain measures the uncertainty reduction of the global optimum $\boldsymbol{x}_*$, if a new data pair $(\mathbf{x}, \mathbf{y})$ is added

$$\boldsymbol{\alpha}(\mathbf{x}) \;=\; \mathbf{H}[\boldsymbol{x}_*|\boldsymbol{\mathcal{D}}] - \mathbb{E}_{\mathbf{y}}\{\mathbf{H}[\boldsymbol{x}_*|\boldsymbol{\mathcal{D}} \cup (\mathbf{x}, \mathbf{y})]\}$$

Exact evaluation is infeasible!

Differential entropy:

$$\mathrm{H}[x] = \int_{\chi} \boldsymbol{p}(\boldsymbol{x})\boldsymbol{logp}(\boldsymbol{x})\boldsymbol{dx}$$

Training dataset:

$$\boldsymbol{\mathcal{D}} = \{(\boldsymbol{x}_n, \boldsymbol{y}_n)\}_{n \le N}$$

Acquisition Function of PESC

- PESC concept
  - Information gain measures the uncertainty reduction of the global optimum $\boldsymbol{x}_*$, if a new data pair $(\mathbf{x}, \mathbf{y})$ is added

$$\boldsymbol{\alpha}(\mathbf{x}) \;=\; \mathbf{H}[\boldsymbol{x}_*|\boldsymbol{\mathcal{D}}] - \mathbb{E}_{\mathbf{y}}\{\mathbf{H}[\boldsymbol{x}_*|\boldsymbol{\mathcal{D}} \cup (\mathbf{x}, \mathbf{y})]\}$$

the symmetry of mutual information

$$\boldsymbol{\alpha}(\mathbf{x}) \;=\; \mathbf{H}[\mathbf{y}|\boldsymbol{\mathcal{D}}, \mathbf{x}] - \mathbb{E}_{\mathbf{x}_*}\{\mathbf{H}[\mathbf{y}|\boldsymbol{\mathcal{D}}, \mathbf{x}, \mathbf{x}_*]\}$$

  - Both entropy terms are now with respect to the predictive distribution of $\mathbf{y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{K})$

Acquisition Function of PESC

- Information gain of a testbench
  - Under the assumption that the objective and constraints are independent, rewrite the PESC acquisition function in the form of summation

$$\boldsymbol{\alpha}(\mathbf{x}) = \sum_{i=1}^{N_c+1} \frac{1}{2}\log\sigma_i^2(x) - \sum_{i=1}^{N_c+1} \mathbb{E}_{x_*}[\frac{1}{2}\log\sigma_i^2(x|x_*)]$$

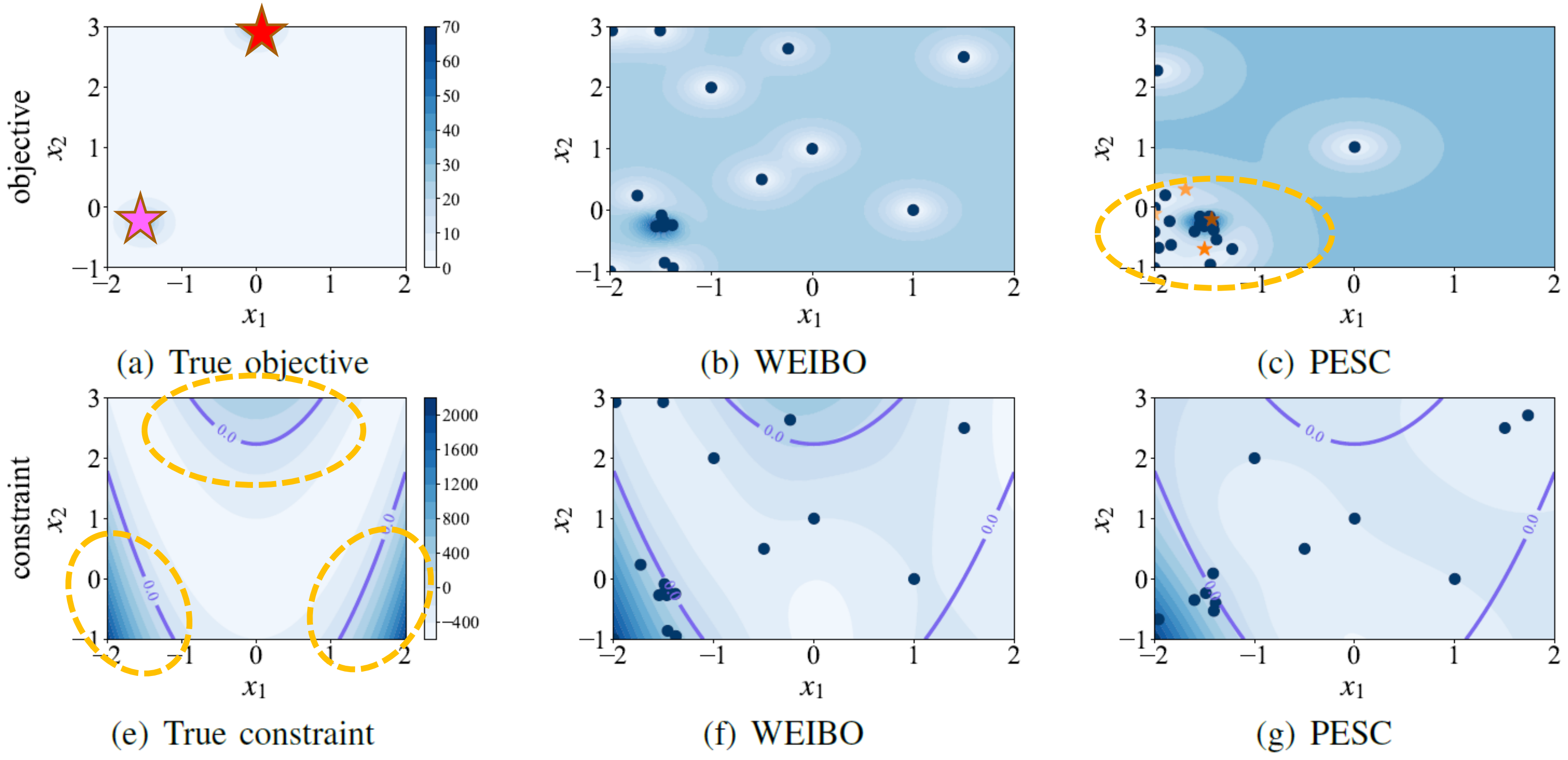
Predictive variance from GP

  - Simulators are invoked based on testbenches, thus information gain of a testbench is the sum of the information gain of the performances inside this testbench

$$\boldsymbol{\alpha}_t(\mathbf{x}) = \sum_{j=1}^{|F_t|} \alpha_t^j(x) = \sum_{j=1}^{|F_t|} \{\frac{1}{2}\log\sigma_{t,j}^2(x) - \mathbb{E}_{x_*}[\frac{1}{2}\log\sigma_{t,j}^2(x|x_*)]\}$$

Dilemma of PESC

- Inadequate exploration ability
  - The power of PESC is on selecting the suitable testbench
  - Inside a testbench, the exploration ability of PESC is inadequate



Feasibility Regions Exploration of Constraints

- Expected Improvement

Improvement of y **Current minimum objective**

$$\boldsymbol{I(y,\tau) = \max(0, \tau - y)}$$

$$\boldsymbol{EI(x) = \mathbb{E}[I(y,\tau)] = \big(\tau - \mu(x)\big)\Phi\left(\frac{\tau - \mu(x)}{\sigma(x)}\right) + \sigma(x)\phi\left(\frac{\tau - \mu(x)}{\sigma(x)}\right)}$$

Where $\boldsymbol{\Phi}$ and $\boldsymbol{\phi}$ are the Cumulative Distribution Function (CDF) and Probability Density Function (PDF) of standard normal distribution

Feasibility Regions Exploration of Constraints

- Feasibility Expected Improvement
  - Fix $\boldsymbol{\tau} = \mathbf{0}$ for exploring feasibility boundary

$$\boldsymbol{FEI(x)} = \mathbb{E}[\boldsymbol{I(y,0)}] = \boxed{\cancel{-\boldsymbol{\mu(x)}\boldsymbol{\Phi}\left(\frac{-\boldsymbol{\mu(x)}}{\boldsymbol{\sigma(x)}}\right)}} + \boldsymbol{\sigma(x)}\boldsymbol{\phi}\left(\frac{-\boldsymbol{\mu(x)}}{\boldsymbol{\sigma(x)}}\right)$$

search into the current feasible region

$$\boldsymbol{FEI(x)} = \boldsymbol{\sigma(x)}\boldsymbol{\phi}\left(\frac{-\boldsymbol{\mu(x)}}{\boldsymbol{\sigma(x)}}\right)$$

  - Large $\boldsymbol{\sigma(x)}$ → explore unknown regions
  - $\boldsymbol{\mu(x)}$ close to $\mathbf{0}$ → exploit boundaries of predicted feasible regions

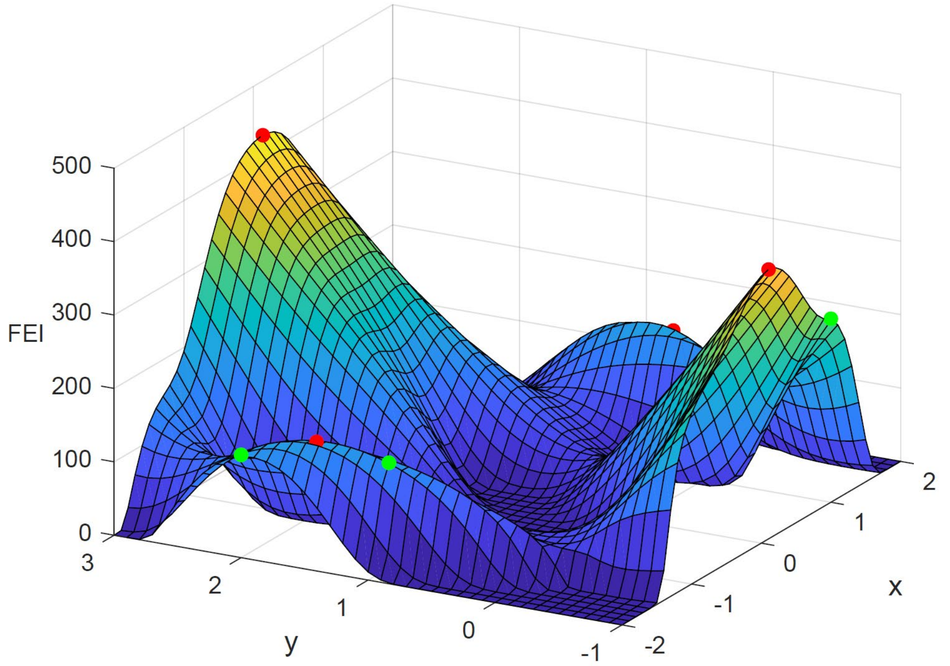
Multi-modal Optimization of FEI

- Multi-modal problem

$$\max_{\mathbf{XS} \subseteq \mathbf{D}} \quad |\mathbf{XS}|$$

Set size of local optimums

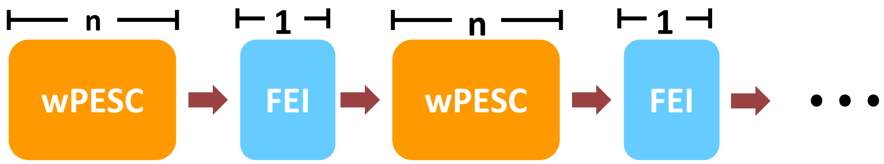
  - Locate all local optimums of FEI function by NMMSO method [5]
- DBSCAN clustering algorithm [6]
  - Avoid repeated exploration in adjacent regions
  - Clustering centers are still local peaks

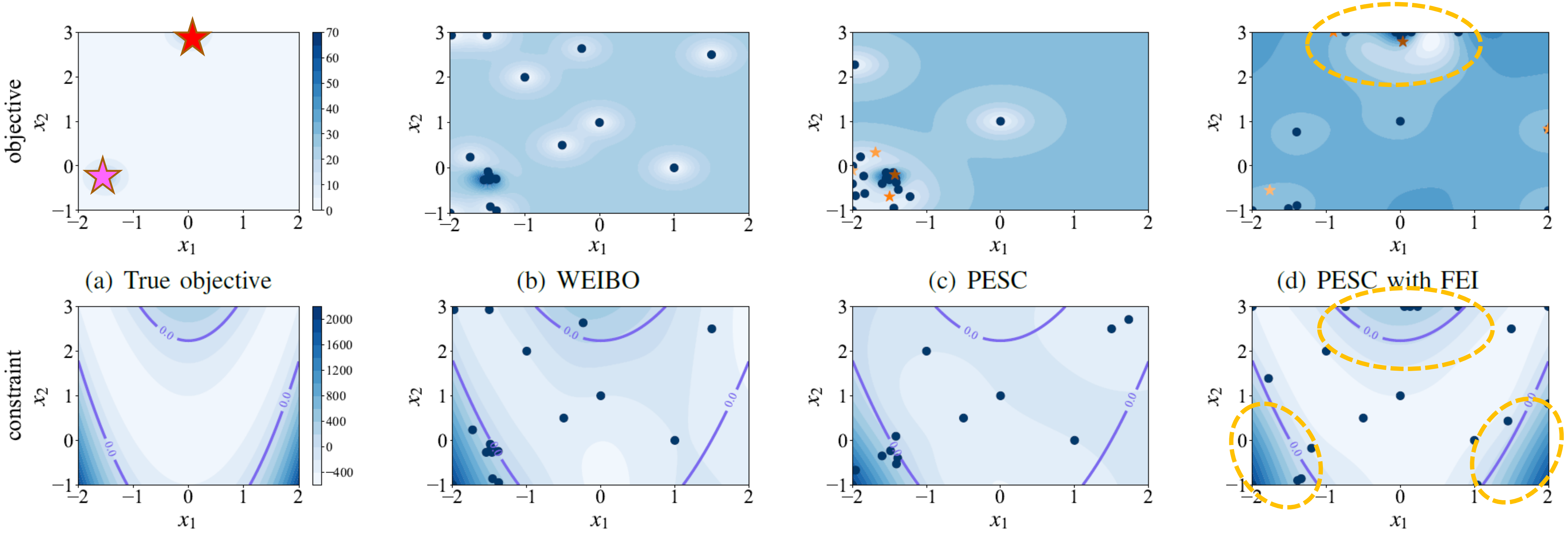


[5] J. E. Fieldsend, “Running up those hills: Multi-modal search with the niching migratory multi-swarm optimiser,” in CEC. IEEE, 2014, pp. 2593–2600.
[6] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” AAAI Press, 1996.

PESC With Feasibility Expected Improvement

- We apply PESC and FEI alternately in the iterations to combine the merits of both acquisition functions





(a) True objective (b) WEIBO (c) PESC (d) PESC with FEI

(e) True constraint (f) WEIBO (g) PESC (h) PESC with FEI

Time-Weighted PESC (wPESC)

- Time-weighted term is proposed to balance information gain and simulation time

$$\boldsymbol{\alpha}_{\mathrm{wPESC}}(\boldsymbol{x}_t^*, \boldsymbol{t}) = \boxed{\frac{\boldsymbol{\alpha}_t(\boldsymbol{x}_t^*)}{\boldsymbol{\alpha}}} + \boxed{\mathbf{w}\frac{\mathbf{1}/\boldsymbol{c}_t}{\boldsymbol{C}}} \qquad \boldsymbol{\alpha} = \sum_{t=1}^{|\mathcal{F}|} \boldsymbol{\alpha}_t(\boldsymbol{x}_t^*)\,, \boldsymbol{C} = \sum_{t=1}^{|\mathcal{F}|} \mathbf{1}/\boldsymbol{c}_t$$

Prefer large information gain **Prefer less simulation time**

- $\boldsymbol{c}_t$ is the empirical simulation time of the t-th testbench
- $\boldsymbol{w}$ is a customized weight coefficient
- The testbench with the maximal $\boldsymbol{\alpha}_{\mathrm{wPESC}}(\boldsymbol{x}_t^*, \boldsymbol{t})$ is selected to simulate at $\boldsymbol{x}_t^*$

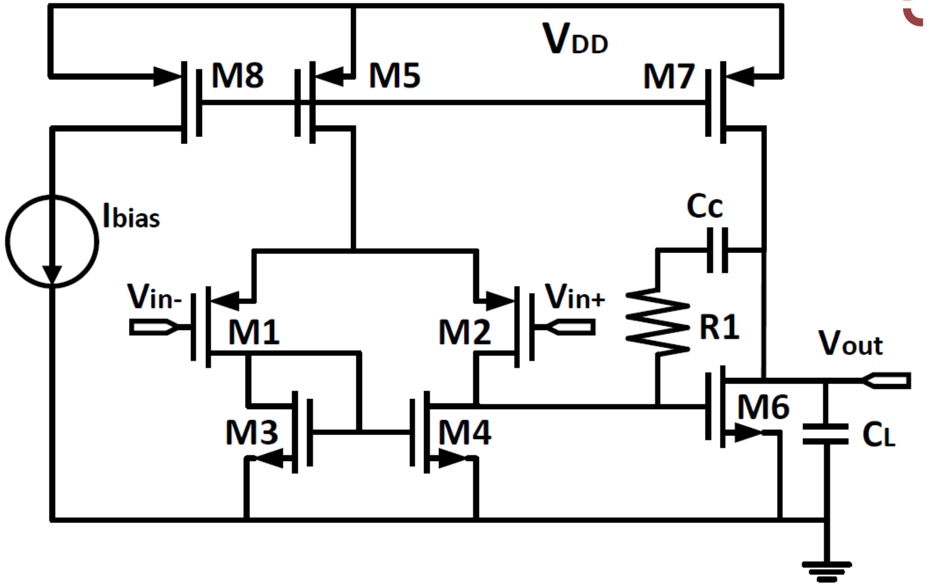
Experimental Results

- Circuits tested
  - Two-Stage Operational Amplifier
  - Low-Power Amplifier
- Simulation-based Algorithms compared
  - Differential evolutionary(DE)
  - GASPAD [7]
  - WEIBO [8]
  - MTBO [9]

[7] B. Liu, D. Zhao, P. Reynaert, and G. G. Gielen, “Gaspad: A general and efficient mm-wave integrated circuit synthesis method based on surrogate model assisted evolutionary algorithm,” IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 33, no. 2, pp. 169–182, 2014.
[8] W. Lyu, P. Xue, F. Yang, C. Yan, Z. Hong, X. Zeng et al., “An efficient bayesian optimization approach for automated optimization of analog circuits,” IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 65, no. 6, pp. 1954–1967, 2017.
[9] J. Huang, S. Zhang, C. Tao, F. Yang, C. Yan, D. Zhou, and X. Zeng, “Bayesian optimization approach for analog circuit design using multitask gaussian process,” in 2021 IEEE International Symposium on Circuits and Systems (ISCAS), 2021, pp. 1–5.

Experiment: Two-Stage Operational Amplifier

- 180nm CMOS process
- 10 Design variables
- 5 testbenches



- Design specifications:

maximize	$GAIN$	Tb1
s.t.	$PM > 60^{\circ}$	
High demand	$UGF > 40MHz$	Tb2
	$PSRR > 55dB$	
	$CMRR > 60dB$	Tb3
	$R_o > 1\Omega$	Tb4
	$SR > 200V/\mu s$	Tb5

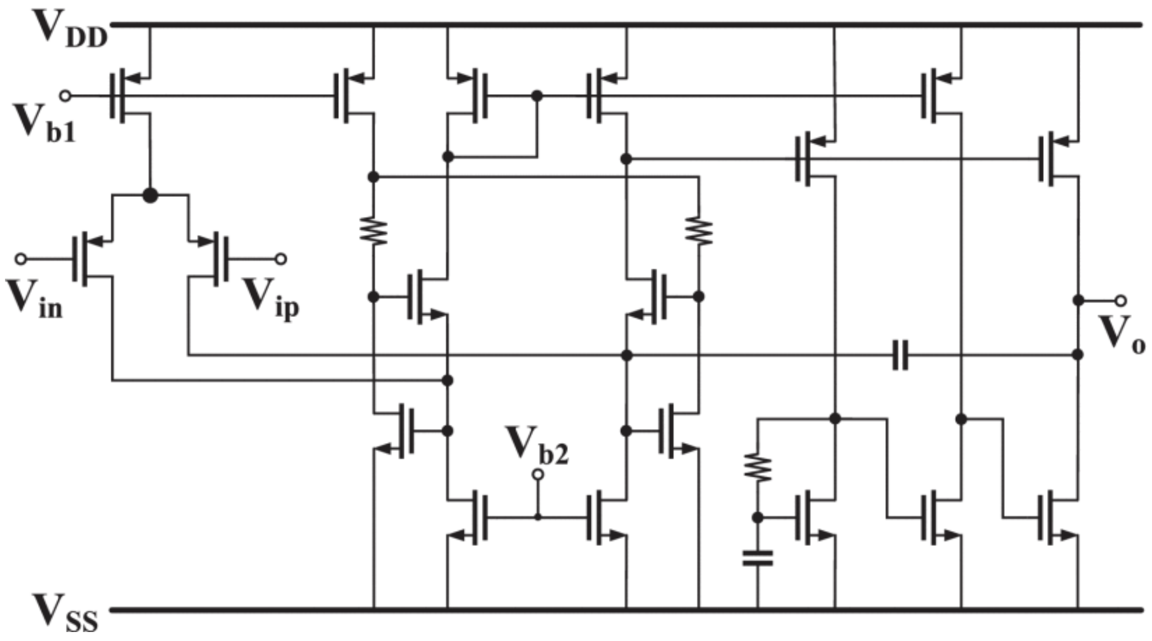
Experiment: Two-Stage Operational Amplifier

- Comparison
  - Our algorithm find the best ***GAIN***
  - The **wPESC+FEI+MM** method can gain 2.71x speedup compared with **WEIBO**, while the most recent method **MTBO** proposed in 2021 just gains 1.14x speedup
  - Both **PESC** and **wPESC** reasonably allocate more simulation points to Tb2, which is very difficult to reach its specifications

Spec/Algorithm	PM[2] (°)	UGF[2] (MHz)	PSRR[2] (dB)	CMRR[3] (dB)	R_o[4] (Ω)	SR[5] (μs)	GAIN[1] worst. (dB)	GAIN[1] mean. (dB)	GAIN[1] best. (dB)	Tb1 (22%)	Tb2 (26%)	Tb3 (16%)	Tb4 (12%)	Tb5 (24%)	Weighted Avg. #Sim	Speedup
Spec	>60	>40	>55	>60	<1	>200	Max	Max	Max	-	-	-	-	-	-	-
DE[5]	62.72	41.93	59.07	84.06	0.52	431.55	86.28	87.20	88.86	1867	1867	1867	1867	1867	1867	0.08×
GASPAD[18]	60.70	40.73	59.11	84.23	0.51	476.86	88.16	88.84	89.36	211	211	211	211	211	211	0.74×
WEIBO[8]	60.60	40.54	59.40	84.32	0.51	434.75	87.97	88.95	89.43	157	157	157	157	157	157	1×
MTBO[9]	61.27	40.19	59.94	84.77	0.47	383.75	88.51	89.20	89.69	138	138	138	138	138	138	1.14×
PESC	61.16	40.87	59.09	85.16	0.42	481.87	88.77	89.34	89.93	131	161	22	41	30	87	1.80×
wPESC	61.23	40.97	59.12	85.13	0.42	495.53	88.35	89.22	89.92	114	141	17	38	23	75	2.09×
wPESC+FEI	61.25	40.87	59.07	85.20	0.42	480.05	88.89	89.36	**89.96**	81	124	23	36	23	64	2.45×
wPESC+FEI+MM	61.21	40.64	59.42	85.19	0.42	443.35	**88.95**	**89.36**	89.93	**71**	**120**	**16**	**29**	**19**	**58**	**2.71×**

Experiment: Low-Power Amplifier

- 350nm CMOS process
- 11 Design variables
- 3 testbenches



- Design specifications:

maximize	IQ	Tb1
s.t.	$GAIN > 110dB$	Tb2
	$SRR > 0.18V/\mu s$ $SRF > 0.2V/\mu s$	Tb3

Loose constraints

Experiment: Low-Power Amplifier

- Comparison
  - Our algorithm find the best ***IQ***
  - Our proposed method gains 3.89x speedup over **WEIBO**, while **MTBO** just gains 1.52x speedup
  - **PESC** method significantly reduces the number of simulation points wasted on Tb3

Spec/Algorithm	mean_GAIN[2] (dB)	mean_SRR[3] (V/μs)	mean_SRF[3] (V/μs)	mean_IQ[1] worst. (μA)	mean_IQ[1] mean. (μA)	mean_IQ[1] best. (μA)	Tb1 (7%)	Tb2 (37%)	Tb3 (56%)	Weighted Avg. #Sim	Speedup
Spec	>110	>0.18	>0.2	Min	Min	Min	-	-	-	-	-
DE[5]	110.04	0.23	0.29	20.17	19.87	19.43	2730	2730	2730	2730	0.06×
GASPAD[18]	110.12	0.23	0.28	19.19	19.11	19.04	415	415	415	415	0.41×
WEIBO[8]	110.04	0.23	0.28	19.21	19.11	19.03	171	171	171	171	1×
MTBO[9]	110.13	0.23	0.28	19.22	19.06	19.02	116	116	116	116	1.52×
PESC	110.14	0.23	0.28	19.11	19.05	19.02	84	158	26	79	2.16×
wPESC	110.14	0.23	0.28	19.10	19.05	19.02	87	140	26	73	2.38×
wPESC+FEI	110.09	0.23	0.28	**19.05**	19.04	19.02	67	87	24	51	3.35×
wPESC+FEI+MM	110.13	0.23	0.28	19.06	**19.03**	**19.02**	**48**	**75**	**21**	**43**	**3.89×**

Thank you!

Questions