VPlace: A New Approach to Performance-Driven Analog IC Placement

ISPD 2026 Presentation

Donghao Fang[1], Hailiang Hu[1], Wuxi Li[2], Jiang Hu[1]

1. Texas A&M University, College Station, TX, USA
2. AMD, Inc., Santa Clara, CA, USA

March 15–18, 2026 | Bonn, Germany

Outline

Motivation: Analog IC Placement

Performance Gap

Analog IC performance (Gain, BW, UGF) is **highly sensitive** to physical layout parasitics.

Post-Layout Degradation

Schematic simulation is ideal, but physical placement **significantly degrades** performance.

Goal

Directly optimize **post-layout metrics** during the analog IC placement stage.

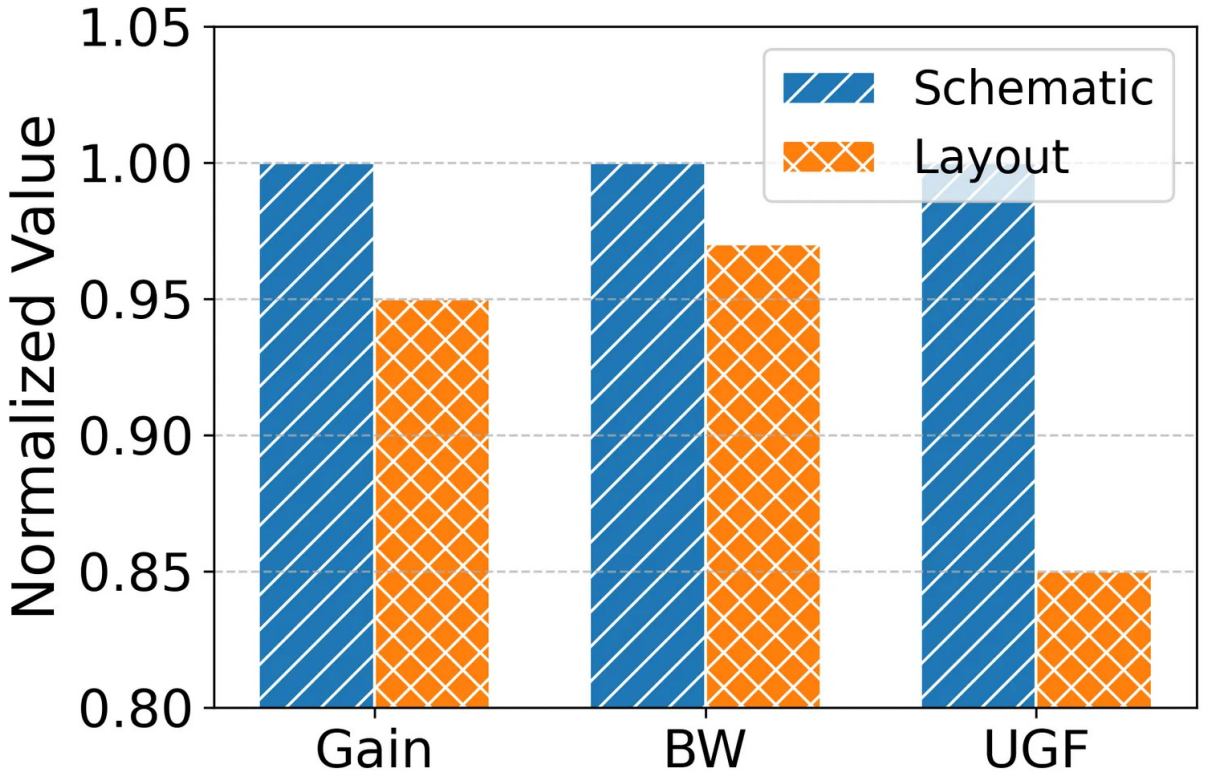

Fig. 1(a): Schematic vs. post-layout performance gap

Motivation: HPWL is Not Enough

Non-Monotonic Relationship

Minimizing HPWL **does not guarantee** better Gain or Bandwidth — peak performance occurs at intermediate HPWL values.

Poor Proxy Metric

HPWL ignores critical analog factors like **coupling capacitance** and device matching.

Need Direct Optimization

We must optimize **performance metrics** directly, not just wirelength.

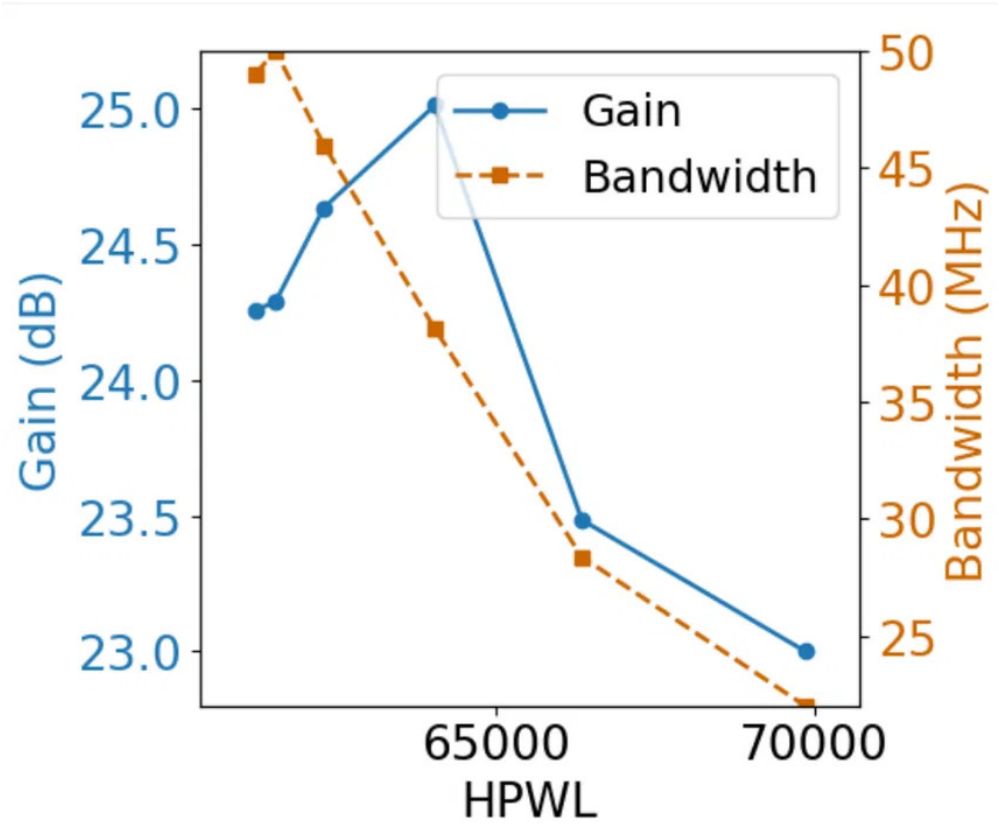

Fig. 1(b): Non-monotonic relationship between HPWL and Gain/BW

Challenge: Inefficient Search in a Sparse Solution Space

Optimization Difficulty

Standard algorithms (SA, BO) are **sample-inefficient** in this vast, invalid-dominant space.

Each evaluation requires a full SPICE simulation, making the optimization extremely costly.

Sparsity of Valid Solutions

Valid layouts (no overlaps, design-rule constraints) are **extremely sparse** in the raw coordinate space.

Random sampling yields mostly illegal solutions, making naive search strategies highly inefficient.

Challenge 2: Surrogate Model Limitations

1

Schematic-Only Focus

State-of-the-art models like **MMM [1]** (Machine learning-based Macro-Modeling) are designed for sizing. They **lack layout spatial information**.

2

Non-Differentiable Kernels

High-accuracy models often use Tree-based methods (XGBoost). **Gradient-based optimization is impossible**.

3

The Gap

We need a surrogate that is both **layout-aware** and compatible with **black-box optimization**.

[1] Y. Lin et al., "MMM: machine learning-based macro-modeling for linear analog ICs and ADC/DACs," IEEE TCAD, 2024.

Background: VAE (Variational Autoencoder) [2]

Generative Model

VAE learns a **probabilistic mapping** between high-dimensional data and a low-dimensional latent space.

Continuous Latent Space

Encodes data into a Gaussian distribution $z \sim \mathcal{N}(\mu, \sigma)$, enabling smooth interpolation.

Reconstruction

The decoder reconstructs the original input from sampled latent vectors.

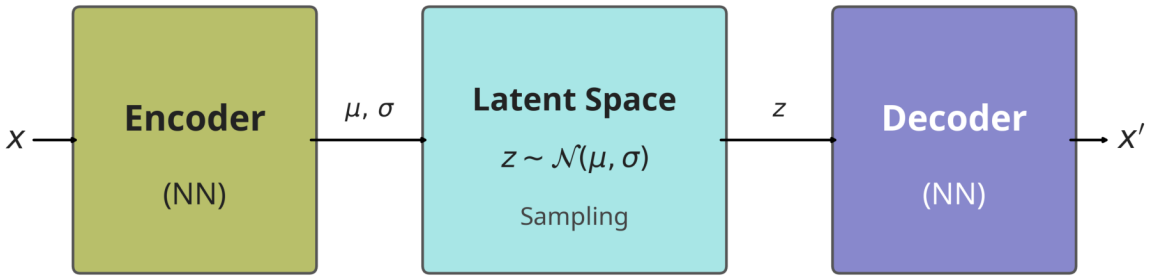

Fig. 2: VAE architecture — Encoder, Reparameterization, Decoder.

[2] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," ICLR, 2014.

Background: VQ-VAE (Vector Quantized VAE) [3]

Vector Quantization

Replaces continuous sampling with a **discrete codebook** look-up.

Discrete Latent Space

Maps input to a sequence of discrete indices z', capturing categorical structures.

High Fidelity

Maintains high reconstruction quality while compressing the search space.

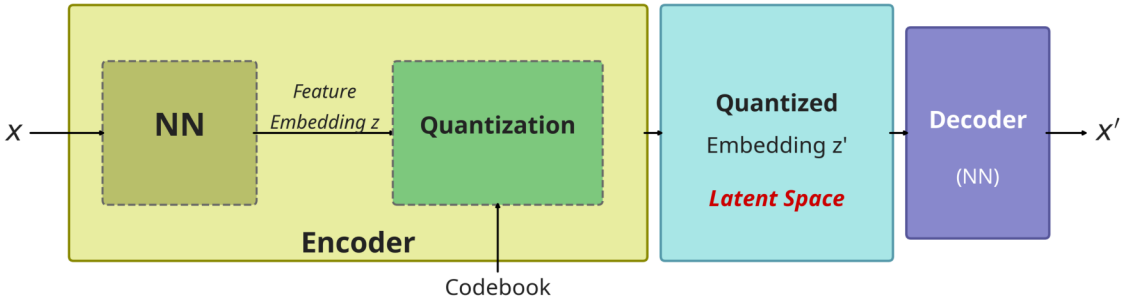

Fig. 3: VQ-VAE architecture — Encoder, Codebook Quantization, Decoder.

[3] A. van den Oord et al., "Neural Discrete Representation Learning," NeurIPS, 2017.

VQ-VAE Training: Three-Term Loss

Total Loss: $\mathscr{L} = \mathscr{L}_{rec} + \beta \cdot \mathscr{L}_{embed} + \mathscr{L}_{commit}$

β: weighting factor | sg[·]: stop-gradient operator

(I) RECONSTRUCTION LOSS

$\mathscr{L}_{rec}$

$$\mathscr{L}_{rec} = \|x - x'\|^2_2$$

Enforce **faithful reconstruction** of placement coordinates from quantized latents.

(II) EMBEDDING LOSS

$\mathscr{L}_{embed}$

$$\mathscr{L}_{embed} = \|sg[z] - e\|^2_2$$

Update **codebook vectors** toward the encoder's latent clusters.

(III) COMMITMENT LOSS

$\mathscr{L}_{commit}$

$$\mathscr{L}_{commit} = \|z - sg[e]\|^2_2$$

Prevent encoder drift; keep outputs close to assigned codewords for a stable discrete latent space.

Take-away: Together, the three terms jointly train the **encoder, decoder, and codebook**, producing a compact discrete latent space suitable for downstream optimization.

Key Mechanism: Why VQ-VAE?

1. **Handling Sparsity & Constraints**

   High-dimensional space is sparse due to strict design rules (non-overlap, symmetry). VQ-VAE maps valid layouts to a **compact manifold**.

2. **Flexible Information Compression**

   Unlike rigid topological representations (e.g., B*-tree, Sequence Pair), VQ-VAE offers a **data-driven, flexible compression ratio**.

3. **Optimization Friendliness**

   The latent space becomes **compact and structured**, enabling stable surrogate modeling and efficient Bayesian Optimization.

4. **Deterministic Decoding**

   Unlike VAE (stochastic sampling), VQ-VAE decoding is **fully deterministic**: the same latent code always produces the same layout, ensuring reproducibility and reliable surrogate evaluation during BO.

Related Work & Research Gap

Method	Pros	Cons
ALIGN [4] (Geometry-driven)	Automated flow, Constraints handling	Optimizes **geometry** only, Indirect proxy for performance
PEA + SA [5] (Performance-driven)	Performance-driven, GNN-based	Classification only (Pass/Fail), Low scalability
MMM [1] (Machine Learning-Based Macro-Modeling)	High accuracy (<1% error), Fast inference	**Schematic-only**, Ignores layout parasitics

The Missing Piece

We need a method that combines the **accuracy of MMM** with **layout awareness**, while efficiently exploring the vast placement space.

→ **Proposed: VPlace (VQ-VAE + Latent BO)**

[4] T. Dhar et al., "ALIGN: A system for automating analog layout," IEEE Design & Test, 2020.
[5] Y. Li et al., "A customized graph neural network model for guiding analog IC placement," ICCAD 2020.
[1] Y. Lin et al., "MMM: machine learning-based macro-modeling for linear analog ICs and ADC/DACs," IEEE TCAD, 2024.

Proposed Framework: VPlace

Stage 1: Data Preparation

Generate diverse layout samples to train the VQ-VAE model.

Stage 2: VV-MMM Training

Train the layout-aware performance estimator using latent embeddings.

Stage 3: Latent Optimization

Perform Bayesian Optimization in the latent space to find the best layout.

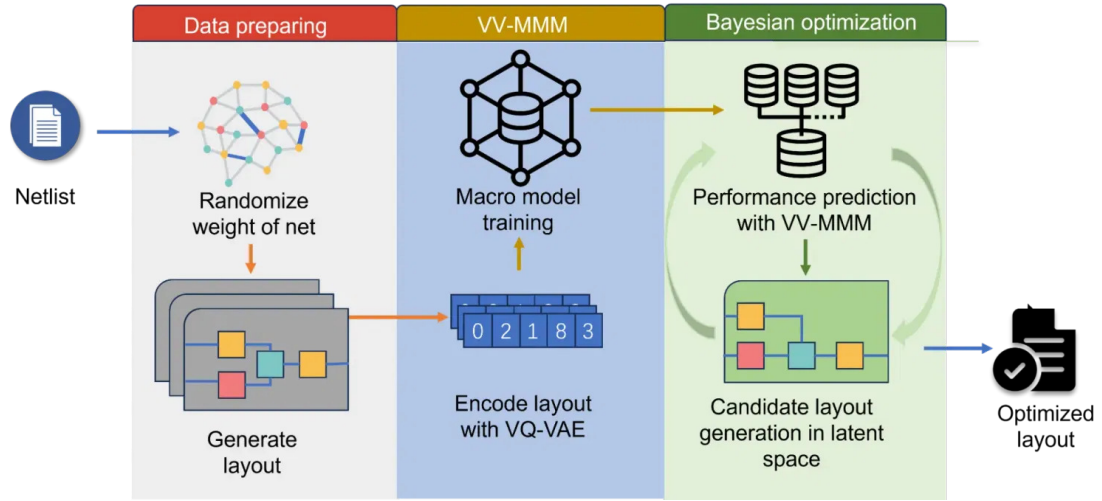

Fig. 4: VPlace framework overview

Method: MMM Limitations

1. Schematic-Only Prediction

MMM relies solely on **device information & netlist** to predict circuit performance, missing the critical physical context.

2. Missing Layout Information

It assumes an ideal layout, completely **ignoring parasitic effects** caused by placement and routing.

Training Data: Randomized Net Weights for Coverage

Problem: Biased Training Data

The original training data consists only of **wirelength-optimized layouts** obtained from different iterations of the optimization process.
This results in a **narrow data distribution** — when BO explores regions outside this distribution, the surrogate model lacks coverage and yields → inaccurate performance predictions.

Solution: Randomized Net Weights

Randomize net weights during placement data generation → diverse layout quality (good & poor).
Surrogate gains **broader layout coverage**, improving prediction accuracy across a richer and more diverse layout space.

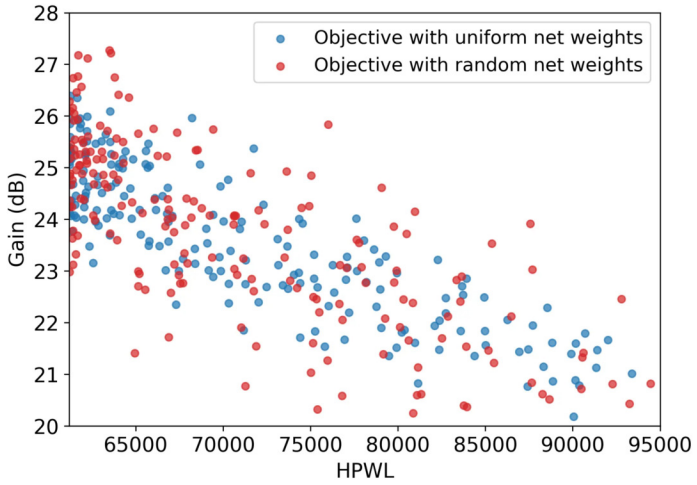

Fig. 5: Random net weights cover a wider HPWL–Gain space

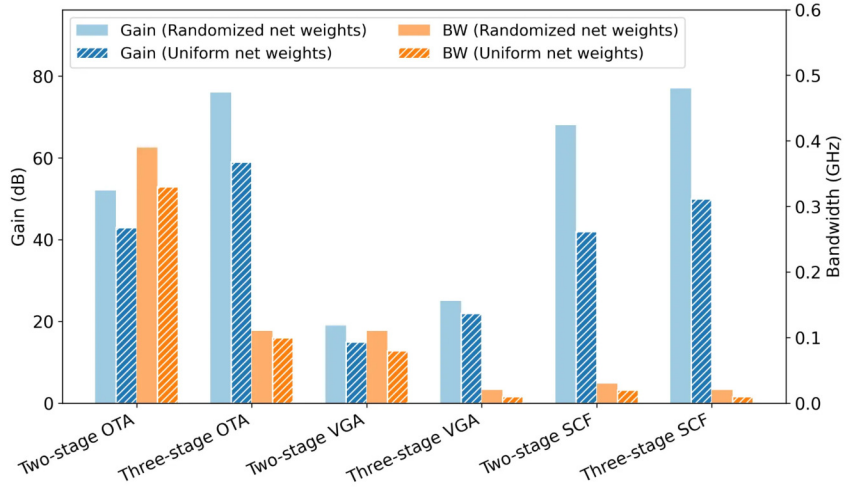

Fig. 6: Randomized training yields higher Gain & BW across all circuits

Proposed: VV-MMM (VQ-VAE enhanced MMM)

Layout-Aware Prediction

VV-MMM incorporates **layout information** into the performance estimation model.

Latent Embeddings

Uses VQ-VAE's latent vector **z'** as a compact layout representation input.

High Accuracy

Achieves high correlation ($R^2 > 0.9$) with post-layout simulation results.

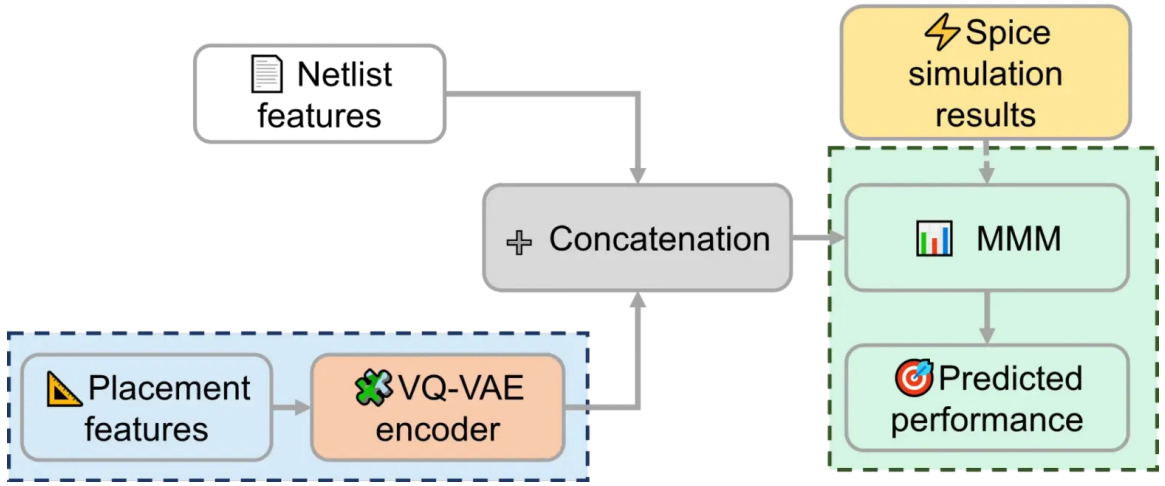

Fig. 7: VV-MMM incorporates layout embeddings

Why Optimization in Latent Space?

1. Improved Placement Feasibility

The latent manifold implicitly captures design constraints, making decoded layouts **more likely to be valid** — closer to realizable placements, though ILP legalization is still applied to remove remaining overlaps.

2. Low Dimensionality

Compresses high-dimensional coordinates into a **compact latent vector**, making Bayesian Optimization highly efficient.

Method: Latent BO Loop

1. Candidate Selection (z')

Acquisition function explores the latent space and proposes a new candidate vector z' likely to improve performance.

2. Decoding & Evaluation

The VQ-VAE decoder maps z' back to a physical layout (**deterministic**: same z' always yields the same layout), then evaluated by VV-MMM.

3. Model Update

The observed performance updates the Gaussian Process, refining the search for the next z.

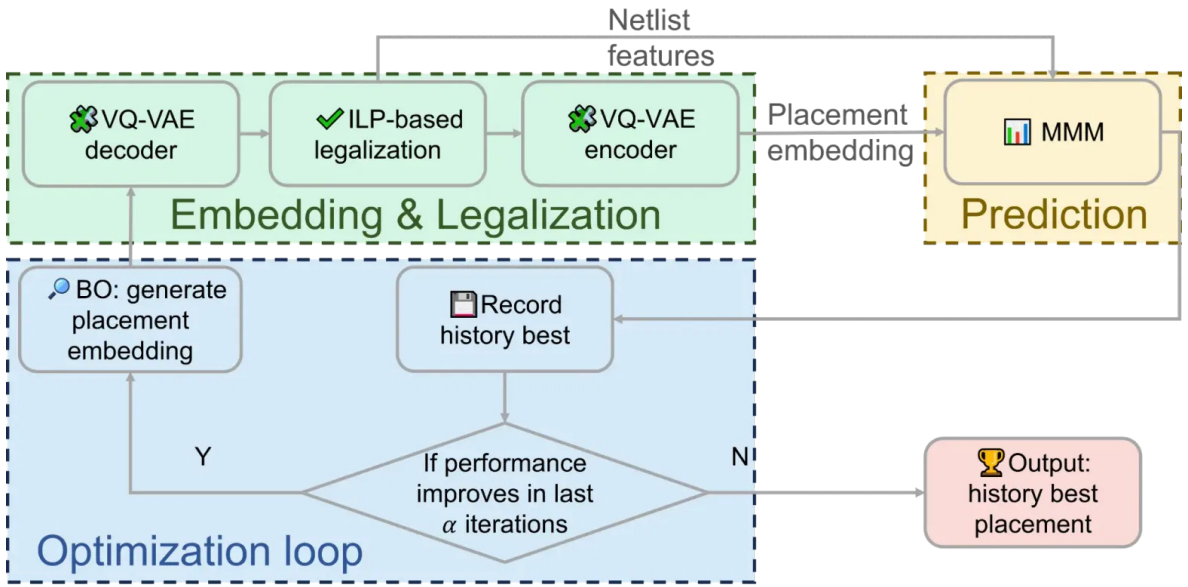

Fig. 8: Bayesian Optimization in Latent Space

> **Notation:** z is a vector in the Latent Space (compressed features), distinct from physical coordinates.

Experiment Setup

Technology

ASAP7 PDK (7nm FinFET predictive process).

Evaluation

Post-layout simulation via **HSPICE**.

Baselines

1. **ALIGN** (Geometry-driven)
2. **PEA+SA** (Performance-driven)

Table 1: Benchmark Circuit Specifications

Benchmark Circuit	# Transistors	Type
Two-stage OTA	9	Op-Amp
Three-stage OTA	13	Op-Amp
Two-stage VGA	15	Amplifier
Three-stage VGA	19	Amplifier
Two-stage SCF	29	Filter
Three-stage SCF	33	Filter

Main Results: Performance Gain

+22–26%

vs. ALIGN (Geometry-driven)

Significant improvement in Gain, UGF, and Phase Margin.

+10–16%

vs. PEA+SA (Performance-driven)

Outperforms SOTA ML methods across all metrics.

Circuit	Gain			Unity Gain Freq (UGF)			Bandwidth (BW)			Phase Margin (PM)		
	ALIGN	PEA	VPlace	ALIGN	PEA	VPlace	ALIGN	PEA	VPlace	ALIGN	PEA	VPlace
Two-stage OTA	0.86	0.92	**1.00**	0.85	0.94	**1.00**	0.87	0.95	**1.00**	0.88	0.93	**1.00**
Three-stage OTA	0.75	0.86	**1.00**	0.71	0.83	**1.00**	0.64	0.82	**1.00**	0.81	0.89	**1.00**
Two-stage VGA	0.68	0.79	**1.00**	0.75	0.86	**1.00**	0.72	0.72	**1.00**	0.77	0.85	**1.00**
Three-stage VGA	0.72	0.88	**1.00**	0.69	0.81	**1.00**	0.50	1.00	**1.00**	0.75	0.85	**1.00**
Two-stage SCF	0.69	0.81	**1.00**	0.69	0.81	**1.00**	0.67	1.00	**1.00**	0.71	0.82	**1.00**
Three-stage SCF	0.69	0.81	**1.00**	0.68	0.79	**1.00**	0.50	1.00	**1.00**	0.71	0.79	**1.00**
Average (Norm.)	**0.73**	**0.84**	**1.00**	**0.73**	**0.84**	**1.00**	**0.65**	**0.92**	**1.00**	**0.77**	**0.85**	**1.00**

Table 2: Normalized performance (Baseline: VPlace = 1.0). ALIGN achieves 0.73–0.77× and PEA achieves 0.84–0.92× of VPlace on average, confirming 22–26% and 10–16% gaps respectively. Average of per-circuit normalized ratios.

Ablation Study: Impact of VQ-VAE

1. Impact on Model Accuracy

Circuit	R^2 Score	
	With VQ-VAE	Without
5-transistor OTA	0.96	0.90
One-stage VGA	0.89	0.84

Table 3: VQ-VAE embedding significantly improves surrogate model accuracy (R2) by capturing latent topology features.

Conclusion:
Latent space embedding is critical for both prediction accuracy and final circuit performance.

2. Impact on Circuit Performance

Metric (Norm.)	VPlace Variant	
	With VQ-VAE	Without
Gain	1.00	0.55 (-45%)
UGF	1.00	0.53 (-47%)
Bandwidth	1.00	0.81 (-19%)
Phase Margin	1.00	0.97 (-3%)
HPWL	1.00	1.02 (+2%)

Table 4: Removing VQ-VAE leads to severe performance degradation (up to 47% drop), proving that direct coordinate optimization is insufficient.

Sensitivity: Codebook Size K

- **Codebook Size (K)** determines the resolution of the latent manifold.
- **Small K (<128)**: Poor reconstruction fidelity; limited layout diversity.
- **Large K (>512)**: Overfitting; search space becomes too sparse/complex.
- **Optimal K = 256**: Achieves the best trade-off for Gain and Bandwidth.

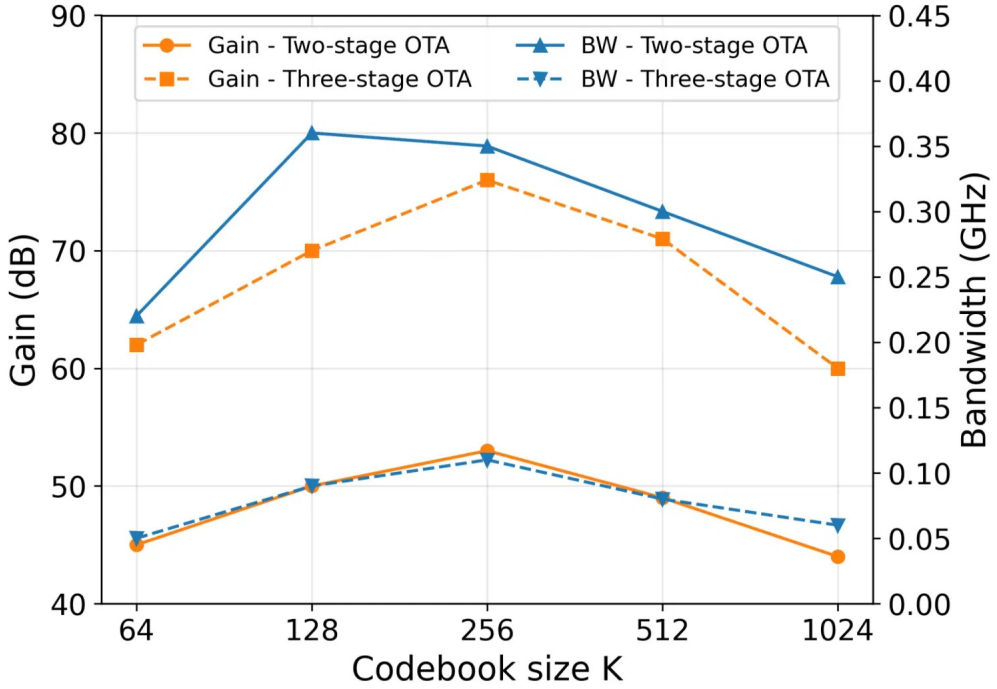

Fig. 9: Impact of Codebook Size K on Performance.

Sensitivity: Latent Length n

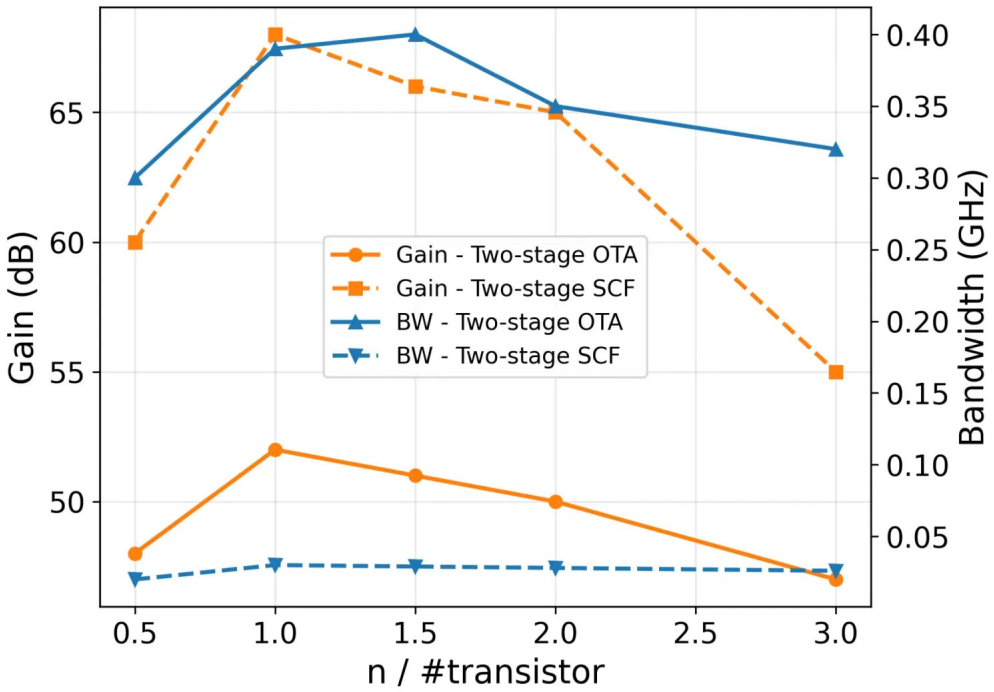

Fig. 10: Impact of Latent Length n on Performance.

- **Latent Length (n)** controls the information capacity of the encoding.
- **Too Short (n < 1.0×)**: Information loss; reconstruction quality degrades, leading to worse circuit performance.
- **Too Long (n > 2.0×)**: Redundancy increases search complexity without gain.
- **Optimal n ≈ 1.0–1.5×**: Matches the number of devices (#transistors).

Conclusion

01 Beyond Geometry-Driven Placement

Analog placement **cannot rely on HPWL alone**. Performance is non-monotonic with wirelength, requiring direct optimization of electrical metrics.

02 VPlace Framework

Proposed framework leverages **VV-MMM** for accurate performance prediction and **Latent BO** for efficient exploration of the design space.

03 Superior Performance

VPlace improves performance by **22–26% over ALIGN** and **10–16% over PEA+SA** across multiple analog circuits.

Thank You

Q & A

Texas A&M University & AMD Inc.

donghao@tamu.edu