ARRIVALFM-OTA: A LARGE-SCALE BENCHMARK FOR PHYSICAL REPRESENTATION LEARNING IN ANALOG CIRCUIT DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Surrogate modeling in analog integrated circuit design is critically obstructed by stiff nonlinear differential-algebraic equations, discrete device sizing, and a scarcity of open, standardized benchmarks and datasets. To bridge this gap, we introduce **ArrivalFM-OTA**, a **forward modeling (FM)** benchmark and dataset for physical representation learning in analog circuit design. ArrivalFM-OTA comprises over 2 million industrial-grade Cadence Spectre simulation datapoints across 28 operational transconductance amplifier (OTA) topologies and 10 FinFET technology nodes from 7 nm to 20 nm, evaluated under continuous environmental sweeps. On top of this data, ArrivalFM-OTA establishes a surrogate modeling challenge across 4 standardized distribution splits and 68 joint prediction targets spanning validity flags, global performance metrics, and internal transistor operating states. Evaluating representative neural baselines across extensive multi-seed initializations, we propose **DAWN (DC Operating Point Analog Waypoint Nodalization)**, a nodal-conditioned residual framework that improves global metric prediction across all backbones by up to $+0.055$ R^2, with our leading model reaching $R^2 \approx 0.793$ in-distribution and $R^2 \approx 0.764$ on an unseen technology node. However, performance drops to $R^2 \approx 0.281$ on unseen topologies, revealing that while neural models achieve competitive interpolation across continuous physical parameters, structural extrapolation across circuit graphs remains a fundamental open challenge for physical representation learning. Finally, we open-source all data, benchmark protocols, multi-seed baseline profiles, and our simulation orchestration pipeline to provide a standardized foundation for transistor-level machine learning and circuit design workflows.

1 INTRODUCTION

Modern chips fundamentally rely on analog integrated circuits for sensing, power, communication, and data conversion. However, nonlinear physical tradeoffs make manual sizing difficult, while design automation remains bottlenecked by simulation latency. Compounded by a critical scarcity of open source transistor-level data required to bridge individual device physics to macroscopic circuit behavior, fast, physically grounded neural surrogate models have emerged as an essential foundation to replace slow numerical solvers and enable scalable analog automation.

Neural surrogates like CktGNN (Dong et al., 2023), ParaGraph (Ren et al., 2020), and INSIGHT (Poddar et al., 2025) have explored graph and sequence formulations for circuit metric prediction, while open-source initiatives (Mehradfar et al., 2025; Yang et al., 2025) have scaled datasets to multimillion-datapoint scales. However, existing work remains confined to legacy planar CMOS and maps sizing parameters to a limited set of global metrics under nominal conditions. Recently, OSIRIS (Chiari et al., 2026) contributed 87,100 verified transistor-level layouts for placement and parasitic extraction (PEX), with its dataset spanning 4 amplifier topologies on planar 130nm CMOS generated with Netgen, Ngspice, and Magic open-source simulators. Crucially, existing datasets lack internal transistor-level DC operating points, the essential physical values that directly influence higher-order performance metrics (Razavi, 2017). Furthermore, to our knowledge, no standardized benchmark exists to evaluate surrogate generalization across complex nanoscale FinFET physics and continuous environmental voltage, temperature, and load variations.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

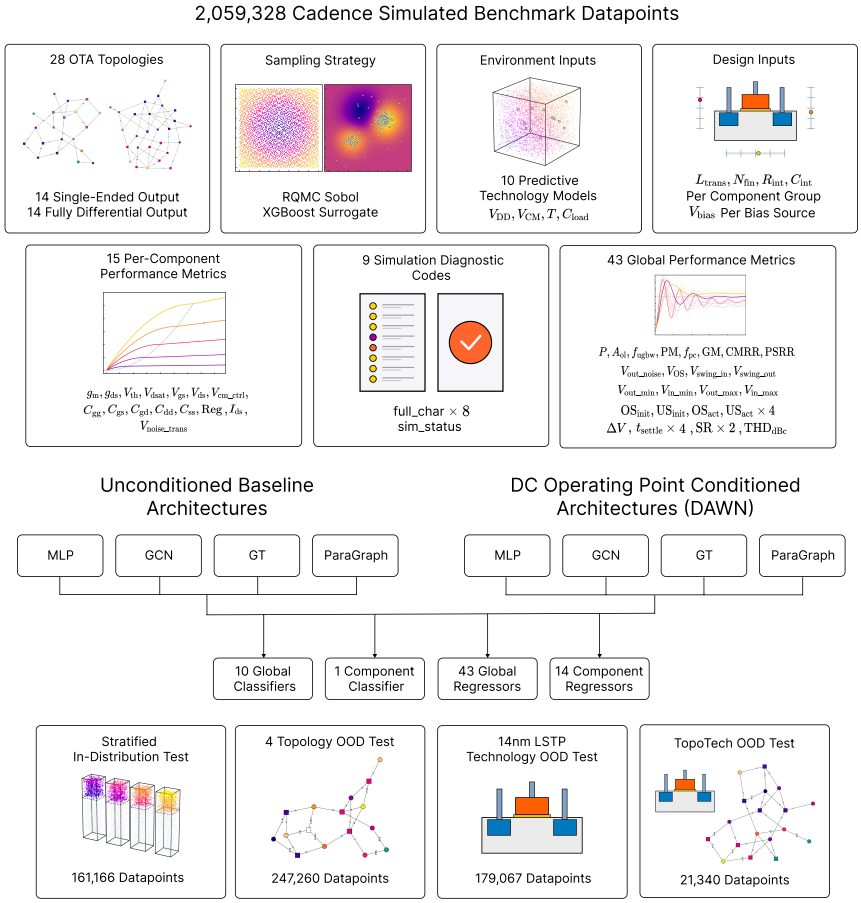

Figure 1: Overview of the **ArrivalFM-OTA** benchmark task. We utilize **Contact**, our data orchestration pipeline, to generate 2,649,573 simulation datapoints with Cadence Spectre (Cadence Design Systems, 2020) at ultra-strict convergence tolerances, designating 2,059,328 for benchmark use. We then evaluate 4 neural backbones on ArrivalFM-OTA alongside their nodal-conditioned (**DAWN**) variants on 68 joint classification and regression tasks across 4 test splits. For breakdowns of all data pipeline diagnostics and performance metric definitions, consult Appendix E.

To address these limitations, we introduce **ArrivalFM-OTA**, an open benchmark task as described in Figure 1. We focus on operational transconductance amplifiers (OTAs), the foundational atomic units of analog integrated circuits, which encapsulate nonlinear device biasing, feedback stability, and high-dimensional performance trade-offs within compact, self-contained topologies. While commercial Process Design Kits (PDKs) are locked behind proprietary nondisclosure agreements, open-source PDKs like SkyWater 130 nm, GF180MCU, and FreePDK45 remain restricted to legacy planar CMOS, and the sole open academic FinFET kit, ASAP7 (Clark et al., 2016), covers only an isolated 7nm node. Consequently, open-source Predictive Technology Models (PTM) (Sinha et al., 2012) provide the only avenue to model a spectrum of FinFET processes without legal restrictions. Free from proprietary layout constraints, PTMs emulate fundamental FinFET properties, providing an ideal foundation to study physical generalization across technology nodes. Spanning 10 PTM nodes from 7 nm to 20 nm across high-performance (HP) and low-standby-power (LSTP) variants, our benchmark directly evaluates surrogate generalization across multigenerational FinFET physics. Our core contributions are as follows:

- We introduce the **ArrivalFM-OTA** open benchmark task, establishing a standardized **forward modeling (FM)** challenge across 2,059,328 benchmark datapoints, 4 rigorous test splits, and 68 joint prediction targets spanning AC/DC performance, transient dynamics, transistor-level operating states, and simulation validity diagnostics.

- Underlying this benchmark, we release a foundational pretraining dataset of $2{,}649{,}573$ Cadence Spectre circuit simulations covering 28 OTA topologies and 10 nanoscale FinFET technology nodes alongside **Contact**, our open-source parallel simulation orchestration framework for community data generation.
- We propose **DAWN (DC Operating Point Analog Waypoint Nodalization)**, a nodal-conditioned residual cascade that routes intermediate operating point representations through zero-initialized projections, and conduct an exhaustive evaluation of 8 model architectures spanning 4 distribution splits and 10 random initializations each, uncovering a fundamental asymmetry between unseen technology node and topological generalization.

2 Related Work

2.1 Open Source Analog Circuit Datasets

Existing open-source datasets on analog circuits largely focus on either static netlist mining or data extraction from planar CMOS simulation sweeps. Efforts like GANA (Kunal et al., 2020), Masala-CHAI (Bhandari et al., 2025), AMSNet (Tao et al., 2024), and AnalogGenie (Gao et al., 2025) compile static circuit schematics for topology discovery, structural graph mining, and visual schematic recognition. On the simulation side, an abundance of prior works simulate designs with Ngspice on the open-source SkyWater 130 nm PDK for algorithmic sizing benchmarks (Li et al., 2024), scalar metric prediction (Dong et al., 2023), layout parasitics (Chiari et al., 2026), transformer-based zero-shot surrogate evaluation (Yang et al., 2025), and a wide variety of optimization, classification, and inverse design tasks (Leibl et al., 2026)[1], scaling from tens of thousands to multimillion-datapoint scales. Meanwhile, AICircuit (Mehradfar et al., 2024) and FALCON (Mehradfar et al., 2025) released datasets on 45 nm CMOS PDKs spanning more than $361{,}000$ and 1 million datapoints simulated with Cadence Spectre to predict 3 to 9 global metrics per design. However, existing simulation datasets focus primarily on circuit-level performance metrics, leaving internal transistor-level DC operating points (DCOP) unrecorded. In physical circuit theory, higher-order AC transfer functions, frequency response, and transient dynamics are nonlinearly governed by these localized device operating states (Razavi, 2017). In contrast, **ArrivalFM-OTA** explicitly captures the complete physical hierarchy: spanning 10 FinFET nodes from 7 nm to 20 nm under sign-off Spectre convergence ($\texttt{reltol} = 10^{-8}$), our benchmark establishes 68 joint prediction targets that systematically supervise internal transistor DC operating states alongside global performance specifications and numerical convergence flags. Because these internal states are tracked at every active component in the schematic graph, a single forward pass predicts between 100 and 338 physical values simultaneously, transforming analog surrogate modeling from black-box scalar regression into dense, physically grounded solver emulation. Consult Table 1 for an extensive comparison across existing work.

2.2 Neural Circuit Surrogates and Physical State Conditioning

Surrogate modeling in analog integrated circuit design spans diverse neural representations, including multi-layer perceptrons over flattened sizing vectors, sequence models (Poddar et al., 2025), continuous waveform neural operators (Ho et al., 2025), message-passing graph neural networks (Ren et al., 2020; Dong et al., 2023), and graph transformers (Fan et al., 2024). However, existing surrogate formulations directly map circuit design parameters to macroscopic global metrics in a single step, bypassing the intermediate transistor-level DC operating states that physically dictate analog circuit behavior. In physical simulation, Sanchez-Gonzalez et al. (2020) and Gilmer et al. (2017) demonstrated that learning localized intermediate state representations is essential for neural architectures to generalize across stiff dynamical systems. Analog integrated circuits adhere to an identical physical hierarchy: transfer functions, high-frequency AC dynamics, and transient settling are nonlinearly governed by the internal DC operating states of active transistors (Razavi, 2017). In deep learning architectures, stable auxiliary adaptation across multi-stage physical pathways is facilitated by zero-initialized projections, as formalized by Fixup initialization (Zhang et al., 2019), which guarantees that added residual branches preserve identity mappings at initialization to prevent optimization collapse.

[1] Concurrent Work

Table 1: Comprehensive comparison of released open source analog circuit datasets. **Environment Sweeps** denote whether voltage references $V_{\text{DD}}, V_{\text{CM}}$, temperature T, capacitive load C_L, or process variations are evaluated beyond nominal defaults. **DCOP** denotes whether internal per-transistor DC operating states are extracted from simulation for each datapoint.

Dataset	Topologies	Datapoints	Metrics	DCOP	Environment Sweeps	Simulator	Technology
GANA (2020)	1,507	1,507	N/A	N/A	N/A	None	None
AMSNet (2024)	824	824	N/A	N/A	N/A	None	None
AnalogGenie (2025)	3,350	3,350	N/A	N/A	N/A	None	None
Masala-CHAI (2025)	7,500	7,500	N/A	N/A	N/A	None	None
AnalogGym (2023)	30	30	16	No	Yes	Ngspice	130nm CMOS PDK
CktGNN (2023)	60,000	60,000	4	No	No	Ngspice	130nm CMOS PDK
OSIRIS (2026)	4	87,100	Layout FoM	No	No	Netgen / Magic / Ngspice	130nm CMOS PDK
ZeroSim (2025)	60	3,600,000	11	No	No	Ngspice	130nm CMOS PDK
AICircuit (2024)	9	361,000+	3–9	No	No	Cadence Spectre	45nm CMOS PDK
FALCON (2025)	20	1,000,000+	3–7	No	No	Cadence Spectre	45nm CMOS PDK
ArrivalFM-OTA	**28**	**2,649,573**	**100–338**	**Yes**	**Yes**	**Cadence Spectre (`reltol` $= 10^{-8}$)**	**10 × 7nm–20nm FinFET PTM**

3 Dataset and Benchmark Formulation

3.1 Adaptive Sampling Strategy

As formalized in Algorithm 1 and detailed in Appendices C.2 and C.3, our data generation pipeline utilizes a **RQMC Sobol XGBoost Surrogate** sampling approach to combine maximum uniform exploration of the parameter space with efficient sampling of high-yielding, high-performance designs. To quantify topological design complexity, we define an OTA's **empirical sample density** ρ_{EMP} over its active sampling dimension D_i. Derived via an initial batch of 16,384 quasi-Monte Carlo (RQMC) Sobol evaluations at seed 100, a candidate sizing vector $\mathbf{x} \in [0,1]^{D_i}$ is deemed electrically valid if it achieves full simulation convergence ($\texttt{sim_status} = 3$) while producing non-NaN values for input offset voltage V_{OS} and output voltage swing $V_{\text{out_range}}$:

$$Y_{\text{valid}}(\mathbf{x}) = \mathbb{I}\Big(\texttt{sim_status}(\mathbf{x}) = 3 \ \wedge \ V_{\text{OS}}(\mathbf{x}) \neq \text{NaN} \ \wedge \ V_{\text{out_range}}(\mathbf{x}) \neq \text{NaN}\Big) \tag{1}$$

Normalizing the total valid count $N_{\text{valid}} = \sum Y_{\text{valid}}(\mathbf{x})$ by the parameter dimension yields:

$$\rho_{\text{EMP}} = \frac{N_{\text{valid}}}{D_i} \tag{2}$$

We then assign a topological difficulty, exploration batch count N_{exp}, and precision batch count N_{prec} for each topology under an invariant $2:1$ exploration to precision ratio:

$$\big(\text{Tier}, N_{\text{exp}}, N_{\text{prec}}\big) = \begin{cases} (\textbf{Easy}, 2, 1), & \text{if } \rho_{\text{EMP}} > 512, \\ (\textbf{Medium}, 4, 2), & \text{if } 128 \leq \rho_{\text{EMP}} \leq 512, \\ (\textbf{Hard}, 6, 3), & \text{if } \rho_{\text{EMP}} < 128. \end{cases} \tag{3}$$

Algorithm 1 Adaptive RQMC Sobol XGBoost Surrogate Sampling Loop

Require: Circuit topology $\mathcal{T}$, batch sizes $B_{\text{exp}} = 16{,}384$ and $B_{\text{prec}} = 12{,}000$, dataset $\mathcal{D} = \emptyset$
1: $\mathcal{D} \leftarrow \text{Simulate}\big(\text{RQMCSobolSample}(\mathcal{T}, B_{\text{exp}})\big)$ {Pilot evaluation}
2: Compute empirical yield N_{valid} from $\mathcal{D}$, calculate ρ_{EMP}, and derive equivalent $(\text{Tier}, N_{\text{exp}}, N_{\text{prec}})$
3: **for** $k = 2$ **to** N_{exp} **do**
4: $\quad \mathcal{D} \leftarrow \mathcal{D} \cup \text{Simulate}\big(\text{RQMCSobolSample}(\mathcal{T}, B_{\text{exp}})\big)$ {Exploration phase}
5: **end for**
6: **for** $k = 1$ **to** N_{prec} **do**
7: $\quad \mathbf{X}_{\text{screened}} \leftarrow \text{XGBoostSurrogateScreen}(\mathcal{T}, \mathcal{D}, B_{\text{prec}})$ {Screen 2^{22} candidates; select B_{prec} designs}
8: $\quad \mathcal{D} \leftarrow \mathcal{D} \cup \text{Simulate}(\mathbf{X}_{\text{screened}})$ {Precision phase}
9: **end for**
10: **return** Dataset $\mathcal{D}$ with total simulation count $|\mathcal{D}| = N_{\text{exp}} B_{\text{exp}} + N_{\text{prec}} B_{\text{prec}}$

3.2 Dataset Overview and Benchmark Splits

We construct a foundational OTA performance dataset using **Contact**, our open-source high-throughput simulation orchestrator. Executing the sampling strategy in Section 3.1, we use Contact to queue Cadence Spectre simulations (Cadence Design Systems, 2020) in massively parallel batches across more than 140 CPU cores, spanning more than $250,000$ core-hours of compute while guaranteeing sign-off numerical precision by enforcing ultra-strict convergence tolerances ($\texttt{reltol} = 10^{-8}, \texttt{vabstol} = 10^{-12}, \texttt{iabstol} = 10^{-16}$). In total, ArrivalFM-OTA comprises $2{,}649{,}573$ fully characterized simulations spanning 28 expert-curated OTA topologies covering both single-ended and fully differential architectures, evaluated across 10 FinFET predictive technology models (PTMs) from 7 nm to 20 nm under continuous environmental sweeps. Crucially, this span captures analog scaling dynamics where output conductance g_{ds} and transconductance g_{m} drive higher open-loop gain in scaled sub-10 nm nodes and low-standby power (LSTP) variants, as seen in Appendix Figure 18. Each datapoint pairs an OTA Spectre `.scs` netlist with input design and environment parameters, capturing 15 per-component operating point metrics, 43 global performance specifications, and 9 simulation diagnostic codes. For complete testbench netlists, solver settings, pipeline mechanics, and formal metric definitions, consult Appendices D and E, with exhaustive dataset distributions detailed in Appendix F.

To evaluate forward performance prediction, we partition $2{,}059{,}328$ selected simulation datapoints into 6 standardized evaluation splits designed to test both in-distribution predictive accuracy and out-of-distribution generalization, as summarized in Table 2. To evaluate surrogate capacity to emulate complete physical solver passes without simulation in the loop, ArrivalFM-OTA formalizes 68 joint prediction targets across 3 task types:

- **10 Global Classification Tasks:** Supervising numerical simulation convergence and 9 functional stability flags across nominal bandwidth, phase crossover, swing range, offset voltage, transient settling, and periodic stability.
- **43 Global Regression Tasks:** Predicting circuit-level specifications spanning small-signal AC response, DC voltage swings, transient settling times, slew rates, overshoot and undershoot percentages, integrated noise, and harmonic distortion.
- **15 Component-Level Heads:** Supervising 1 categorical transistor operating region classification head alongside 14 continuous local operating states, encompassing 13 internal transistor DC operating point and noise metrics alongside active common-mode feedback (CMFB) control voltages.

ArrivalFM-OTA is designed as a representation-agnostic benchmark: our public release provides raw Spectre netlists to support sequence and tokenized language models, tabular sizing tables for flat continuous regressors, and structured graph encodings for geometric deep learning. For benchmark task definitions, units, and transformation rules across all targets, consult Appendix G.

Table 2: Standardized **ArrivalFM-OTA** benchmark splits across $2{,}059{,}328$ simulation datapoints. In-distribution subsets (Training, Validation, and In-Distribution Test) follow an $80/10/10$ randomized split stratified per topology. Out-of-distribution (OOD) test sets isolate structural and physical domain shifts by holding out 4 diverse amplifier topologies (`OTA_FD02`, `OTA_SE09`, `OTA_FD13`, `OTA_SE13`) and 1 unseen technology node (14nm LSTP). For information on the technology nodes and OTA topologies utilized in this benchmark, consult Appendices A and B.

Benchmark Split	Datapoints	Share	Evaluation Target
Training	1,289,328	62.6%	80% of seen topologies on seen technology nodes
Validation	161,167	7.8%	10% of seen topologies on seen technology nodes
In-Distribution Test	161,166	7.8%	10% of seen topologies on seen technology nodes
Topology OOD Test	247,260	12.0%	Unseen topologies on seen technology nodes
Technology OOD Test	179,067	8.7%	Seen topologies on unseen technology node
TopoTech OOD Test	21,340	1.0%	Unseen topologies on unseen technology node

Table 3: Taxonomy and sub-vector layout of circuit graph features and global context vector $\mathbf{u}$. Components and nets share a unified 16-dimensional node feature space with orthogonal zero-padding to isolate device classes from electrical net identities.

Element	Dim	Vector Layout & Sub-Vector Composition
Components $\mathcal{V}_{\text{comp}}$	16	**5D** type one-hot (`nfet, pfet, res, cap, cmfb`) + **9D** zeros + **2D** input parameters ($L_{\text{trans}}, N_{\text{fin}}$), ($R_{\text{int}}, 0$), ($C_{\text{int}}, 0$), or $(0, 0)$ from device type
Nets $\mathcal{V}_{\text{net}}$	16	**5D** zeros + **9D** type one-hot (`VDD, GND, VINP, VINN, VOUTP, VOUTN, VCM, VBIAS, VNODE`) + **2D** input parameters ($V_{\text{DD}}, 0$), ($V_{\text{CM}}, 0$), ($V_{\text{bias}}, 0$), or $(0, 0)$ from node type
Pins $\mathcal{E}$	9	**9D** type one-hot (`D, G, S, B, T, CMS+, CMS−, FB, REF`)
Context $\mathbf{u}$	21	**17D** PTM node parameters (h_{fin}, t_{fin}, EOT, $\Phi_{g,\text{nmos}}$, $\Phi_{g,\text{pmos}}$, N_{body}, L_{dg}, $\mu_{0,\text{nmos}}$, $\mu_{0,\text{pmos}}$, $V_{\text{sat,nmos}}$, $V_{\text{sat,pmos}}$, T_{sili}, $\rho_{c,\text{nmos}}$, $\rho_{c,\text{pmos}}$, F_{pitch}, $\text{ETA}_{0,\text{nmos}}$, $\text{ETA}_{0,\text{pmos}}$) + **4D** environmental conditions (V_{DD}, V_{CM}, C_{load}, T_{temp})

4 Representational Paradigms and the DAWN Framework

4.1 Circuit Representation and Architectural Baselines

While ArrivalFM-OTA supports diverse representational paradigms, we establish our primary baselines using geometric architectures. **Attributed Bipartite Graph:** We formalize each circuit as a directed bipartite graph:

$$\mathcal{G} = (\mathcal{V}_{\text{comp}} \cup \mathcal{V}_{\text{net}}, \mathcal{E}, \mathbf{X}, \mathbf{E}, \mathbf{u}) \tag{4}$$

where components $\mathcal{V}_{\text{comp}}$ and nets $\mathcal{V}_{\text{net}}$ form disjoint node partitions ($\mathcal{V}_{\text{comp}} \cap \mathcal{V}_{\text{net}} = \emptyset$) unified into a 16-dimensional node feature tensor $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times 16}$. To maintain a homogeneous tensor representation for message passing while strictly preventing semantic cross-talk between device classes and electrical nets, categorical indicators are orthogonally zero-padded as summarized in Table 3. Directed pin edges $e_{ij} \in \mathcal{E}$ carry 9-dimensional terminal attributes $\mathbf{e}_{ij}$, instantiated bidirectionally between devices and nets. Consult Appendix Figure 19 for a visual reference.

Global Physical Context Vector: A 21-dimensional vector $\mathbf{u} = [\mathbf{p}_{\text{PTM}} \,\|\, \mathbf{v}_{\text{env}}] \in \mathbb{R}^{21}$ concatenates 17 continuous PTM technology parameters across all 10 FinFET nodes with 4 continuous operating corners. The vector $\mathbf{u}$ is broadcast to graph readout and conditioning layers across all backbones.

Evaluated Neural Backbones: To benchmark surrogate modeling across distinct inductive biases under strictly matched parameter budgets ($L = 4$ layers or propagation steps, hidden width $d = 512$, and dropout $p = 0.1$), we implement 4 representative architectures:

(1) **MLP**: A DeepSets-style feedforward baseline processing node features $\mathbf{X}$ and global context $\mathbf{u}$ independently with topological edges disabled ($\mathcal{E} = \emptyset$), strictly isolating the empirical utility of graph message passing under matched network capacity.

(2) **GCN** (Kipf & Welling, 2017): A standard isotropic graph convolutional network propagating normalized symmetric neighborhood features over graph topology $\mathcal{G}$.

(3) **Graph Transformer (GT)** (Shi et al., 2021): A relational self-attention network deploying 4 attention heads via `TransformerConv`, fusing projected 9-dimensional directed pin edge attributes $\mathbf{h}_{ij}^{e} \in \mathbb{R}^{512}$ directly into the attention mechanism to capture fine-grained terminal interactions.

(4) **ParaGraph (R-GGNN)**: We adopt the multi-relational netlist formulation of ParaGraph (Ren et al., 2020), modeling terminal asymmetry via dedicated projection kernels $\mathbf{W}_r \in \mathbb{R}^{d \times d}$ for each of the 9 pin relations. Crucially, whereas the original ParaGraph design deploys untied relational-attention and GraphSAGE layers, we implement this multi-relational paradigm as a parameter-tied Relational Gated Graph Neural Network (R-GGNN). Node states are updated recurrently via a shared GRU cell across $T = 4$ steps, preventing severe parameter explosion at $d = 512$ and enforcing parameter parity (2.15M parameters) across baselines while emulating iterative numerical relaxation sweeps. Consult Appendix H for extensive architectural definitions.

Multi-Task Readout and Conditioning: All four backbones feed their final node embeddings $\mathbf{h}_v^{(L)}$ and global context $\mathbf{u}$ into an identical readout hierarchy. For graph-level predictions, permutation-invariant pooling concatenates mean and maximum node summaries with a linear projection of the global context vector $\mathbf{h}_u = \text{GELU}(\mathbf{W}_u\mathbf{u} + \mathbf{b}_u)$:

$$\mathbf{h}_{\text{circuit}} = \text{MLP}_{\text{global}}\left(\left[\text{MeanPool}(\{\mathbf{h}_v^{(L)}\}) \,\|\, \text{MaxPool}(\{\mathbf{h}_v^{(L)}\}) \,\|\, \mathbf{h}_u\right]\right) \in \mathbb{R}^{512}. \tag{5}$$

$\mathbf{h}_{\text{circuit}}$ dispatches to 10 binary classification heads predicting simulation validity and stability flags, alongside 41 regression heads for global circuit metrics. Simultaneously, to condition device operating states on environmental corners, each component node combines structural embeddings with global context ($\tilde{\mathbf{h}}_v = \mathbf{h}_v^{(L)} + \mathbf{h}_u$) to drive heads for transistor operating regions and 14 continuous DC operating point state variables.

4.2 DC Operating Point Analog Waypoint Nodalization (DAWN)

In analog circuit design, global-level performance specifications such as open-loop gain, unity-gain bandwidth, and transient dynamics are nonlinearly governed by localized transistor DC operating points (Razavi, 2017). However, existing surrogate formulations predominantly treat performance evaluation as a direct end-to-end regression from design parameters to global metrics, omitting intermediate physical states. Thus, we formulate **DAWN (DC Operating Point Analog Waypoint Nodalization)** as a physics-conditioned residual cascade operating in two hierarchical steps: **1. Residual Waypoint Cascade:** Structural node embeddings are conditioned on operating corners via $\tilde{\mathbf{h}}_v = \mathbf{h}_v^{(L)} + \mathbf{h}_u$. A compact bottleneck MLP ($512 \to 128 \to 128$) projects this conditioned representation into an additive physical residual:

$$\mathbf{z}_v = \text{Dropout}\left(\text{GELU}\left(\mathbf{W}_1\tilde{\mathbf{h}}_v + \mathbf{b}_1\right)\right), \quad \boldsymbol{\Delta}_v = \mathbf{W}_{\text{up}}(\mathbf{W}_2\mathbf{z}_v + \mathbf{b}_2) + \mathbf{b}_{\text{up}} \tag{6}$$

where $\mathbf{W}_{\text{up}} \in \mathbb{R}^{512\times128}$ (adding only $\sim 6\%$ parameter overhead to the base backbone). **2. Multi-Scale Dual Readout:** The learned residual $\boldsymbol{\Delta}_v$ augments representations at both device and circuit levels:

$$\mathbf{h}_v^{\text{dcop}} = \tilde{\mathbf{h}}_v + \boldsymbol{\Delta}_v, \qquad \mathbf{h}_v^{\text{phys}} = \mathbf{h}_v^{(L)} + \boldsymbol{\Delta}_v. \tag{7}$$

At the component level, $\mathbf{h}_v^{\text{dcop}}$ drives localized prediction heads for operating regions and continuous DCOP variables: $\hat{\mathbf{y}}_{\text{dcop},v} = \text{Head}_{\text{dcop}}(\mathbf{h}_v^{\text{dcop}})$. At the circuit level, global pooling operates on $\mathbf{h}_v^{\text{phys}}$ concatenated with $\mathbf{h}_u$ to predict circuit performance. Crucially, following Fixup initialization (Zhang et al., 2019), the up-projection parameters are initialized strictly to zero ($\mathbf{W}_{\text{up}} = \mathbf{0}$, $\mathbf{b}_{\text{up}} = \mathbf{0}$). This guarantees that at step 0, $\boldsymbol{\Delta}_v = \mathbf{0}$, ensuring $\mathbf{h}_v^{\text{phys}} \equiv \mathbf{h}_v^{(L)}$ and preserving the exact identity mapping of the base backbone. This formulation completely prevents auxiliary gradient interference from destabilizing early optimization. Because DAWN operates strictly on node embeddings after message passing, it serves as a universal drop-in cascade across our baseline models **(DAWN MLP, DAWN GCN, DAWN GT, DAWN ParaGraph)** without modifying core backbones.

4.3 Multi-Stage Multi-Task Objective

All models in the **ArrivalFM-OTA** benchmark are trained end-to-end under a unified multi-task objective combining global simulation classification and multi-target regression:

$$\mathcal{L}_{\text{total}} = \bar{\mathcal{L}}_{\text{BCE}} + \bar{\mathcal{L}}_{\text{reg}}, \tag{8}$$

where $\bar{\mathcal{L}}_{\text{BCE}}$ denotes the mean binary cross-entropy across all 10 simulation convergence (`sim_status == 3`) and functional stability flags. The multi-target continuous objective $\bar{\mathcal{L}}_{\text{reg}}$ normalizes over all active regression heads spanning global performance metrics and localized transistor operating states:

$$\bar{\mathcal{L}}_{\text{reg}} = \frac{1}{M_{\text{reg}}}\left(\sum_{k\in\mathcal{S}_{\text{global}}} \bar{\ell}_k(\hat{y}_k, y_k) + \sum_{j\in\mathcal{S}_{\text{local}}} \bar{\ell}_j(\hat{\mathbf{y}}_j, \mathbf{y}_j) + \bar{\ell}_{\text{region}}(\hat{\mathbf{y}}_{\text{reg}}, \mathbf{y}_{\text{reg}})\right) \tag{9}$$

where each $\bar{\ell}(\cdot)$ denotes the batch- and node-averaged loss evaluated strictly over finite masks ($\mathbb{I}[\text{finite}(y)]$) to exclude non-converging simulations. We apply Huber loss ($\delta = 1.0$) for heavy-tailed dynamic metrics, MSE for continuous specs, and cross-entropy for 5-class transistor operating regions. Jointly supervising local and global targets enforces internal representations that satisfy device-level physics while optimizing circuit-level predictions.

Table 4: ArrivalFM-OTA benchmark results on **global performance metrics** with mean ± standard deviation across **10** random initializations each, evaluating 8 neural surrogate architectures across 4 standardized evaluation splits as detailed in Table 2. Bold denotes the best performing baseline per split.

Architecture	In-Distribution Test (161,166)		Technology OOD Test (179,067)	
	Global R^2	Classification Accuracy (%)	Global R^2	Classification Accuracy (%)
DAWN GCN	**0.7929 ± 0.0041**	**94.76% ± 0.04%**	**0.7639 ± 0.0047**	**94.20% ± 0.14%**
DAWN GT	0.7580 ± 0.0246	93.44% ± 1.11%	0.7221 ± 0.0317	92.82% ± 1.20%
DAWN ParaGraph	0.6786 ± 0.0073	91.64% ± 0.15%	0.6537 ± 0.0072	91.53% ± 0.23%
DAWN MLP	0.5753 ± 0.0065	80.26% ± 0.06%	0.5245 ± 0.0063	79.60% ± 0.13%
GCN	0.7492 ± 0.0039	93.74% ± 0.07%	0.7160 ± 0.0035	93.13% ± 0.15%
GT	0.7116 ± 0.0059	92.13% ± 0.46%	0.6741 ± 0.0087	91.46% ± 0.52%
ParaGraph	0.6254 ± 0.0085	89.59% ± 0.12%	0.5988 ± 0.0107	89.41% ± 0.25%
MLP	0.5600 ± 0.0044	80.16% ± 0.04%	0.5121 ± 0.0048	79.47% ± 0.08%
Architecture	**Topology OOD Test (247,260)**		**TopoTech OOD Test (21,340)**	
	Global R^2	Classification Accuracy (%)	Global R^2	Classification Accuracy (%)
DAWN GCN	0.2708 ± 0.0544	72.36% ± 0.42%	0.2023 ± 0.0498	69.94% ± 0.81%
DAWN GT	0.2298 ± 0.0937	70.02% ± 1.59%	0.1699 ± 0.1060	67.42% ± 2.10%
DAWN ParaGraph	0.2331 ± 0.0685	72.13% ± 1.23%	0.1433 ± 0.1182	69.45% ± 1.80%
DAWN MLP	-1.1337 ± 1.3308	64.29% ± 1.03%	-1.5440 ± 1.4246	63.71% ± 0.91%
GCN	**0.2807 ± 0.0434**	**72.37% ± 0.44%**	**0.2093 ± 0.0578**	**70.00% ± 0.50%**
GT	0.1706 ± 0.0978	68.35% ± 2.32%	0.1240 ± 0.1020	65.91% ± 2.54%
ParaGraph	0.1738 ± 0.1261	71.36% ± 1.48%	0.0380 ± 0.2271	68.43% ± 2.04%
MLP	-2.5865 ± 3.2237	65.16% ± 1.38%	-2.6745 ± 3.1459	64.30% ± 1.20%

Table 5: **ArrivalFM-OTA** component-level performance metrics on the **stratified in-distribution** test split, reported as mean ± standard deviation across **10** random initializations. Bold denotes the superior configuration within each paired baseline. Consult Appendix Table 16 for per-device metric definitions.

Model	$\mathbf{g}_{ds}$ (R^2)	$\mathbf{g}_{m}$ (R^2)	$\mathbf{I}_{ds}$ (R^2)	$\mathbf{V}_{ds}$ (R^2)	$\mathbf{V}_{noise_trans}$ (R^2)	$\mathbf{V}_{cm_ctrl}$ (R^2)	Reg (%)
DAWN GCN	**0.8216 ± 0.0016**	**0.7626 ± 0.0019**	**0.7726 ± 0.0025**	**0.7617 ± 0.0020**	**0.7953 ± 0.0059**	**0.6481 ± 0.0037**	**78.42% ± 0.29%**
GCN	0.7772 ± 0.0010	0.7038 ± 0.0019	0.7256 ± 0.0011	0.7387 ± 0.0027	0.7723 ± 0.0042	0.5885 ± 0.0033	76.15% ± 0.23%
DAWN GT	**0.6428 ± 0.0975**	**0.5395 ± 0.1137**	**0.5524 ± 0.1133**	**0.4770 ± 0.1556**	**0.3177 ± 0.1767**	**0.3137 ± 0.1721**	**63.30% ± 8.11%**
GT	0.5286 ± 0.0588	0.4348 ± 0.0622	0.4525 ± 0.0630	0.3069 ± 0.1036	0.2047 ± 0.0821	0.2871 ± 0.1512	54.34% ± 4.59%
DAWN ParaGraph	**0.7181 ± 0.0226**	**0.6134 ± 0.0330**	**0.6330 ± 0.0322**	0.5524 ± 0.0456	0.2313 ± 0.2129	0.4507 ± 0.0520	**68.00% ± 2.06%**
ParaGraph	0.6964 ± 0.0060	0.6064 ± 0.0093	0.6324 ± 0.0088	**0.5560 ± 0.0115**	**0.5704 ± 0.0488**	**0.5145 ± 0.0107**	67.98% ± 0.86%
DAWN MLP	**0.3039 ± 0.0012**	**0.3045 ± 0.0013**	**0.3209 ± 0.0009**	**-0.0117 ± 0.0013**	**0.0795 ± 0.0024**	0.0610 ± 0.0010	**33.05% ± 0.48%**
MLP	0.2979 ± 0.0006	0.2996 ± 0.0012	0.3170 ± 0.0006	-0.0180 ± 0.0011	0.0696 ± 0.0020	0.0610 ± 0.0006	32.48% ± 0.28%

5 Experiments and Benchmark Results

To protect against random initialization noise, we follow the multi-seed evaluation standard established in the Open Graph Benchmark (Hu et al., 2020), training each baseline architecture detailed in Section 4 across 10 independent random initializations (`seed42` to `seed51`); Table 4 reports average continuous regression R^2 and classification accuracy across 43 continuous specifications and 10 discrete stability flags. Across both **In-Distribution** and **Technology OOD** splits, augmenting relational backbones with DAWN yields consistent $+0.04$ to $+0.05$ R^2 gains. Under **Topology OOD** and **TopoTech OOD**, however, the impact of DAWN is mixed: while producing gains of up to $+0.105$ R^2 on **GT** and **ParaGraph**, **DAWN GCN** and **GCN** perform identically within standard deviation. Meanwhile, non-relational **MLP** and **DAWN MLP** surrogates collapse entirely on unseen topologies, proving that local operating features cannot compensate for the absence of circuit connectivity. Without per-architecture hyperparameter tuning, we avoid broad cross-model generalizations; overall, while DAWN reliably aids predictions on known circuits, generalizing to unseen topologies remains a fundamental hurdle for all baselines.

Beyond global specifications, Table 5 displays baseline per-component metric prediction results on the **In-Distribution** test split. Across relational backbones, **DAWN GCN** achieves the highest component fidelity across all categories, notably delivering a $+0.060$ R^2 gain on common-mode control voltage $V_{\text{cm_ctrl}}$, which is critical for common-mode feedback (CMFB) stability. While **DAWN GT** ($+0.114$ on g_{ds}, $+8.96\%$ on operating region classification) and **DAWN ParaGraph**

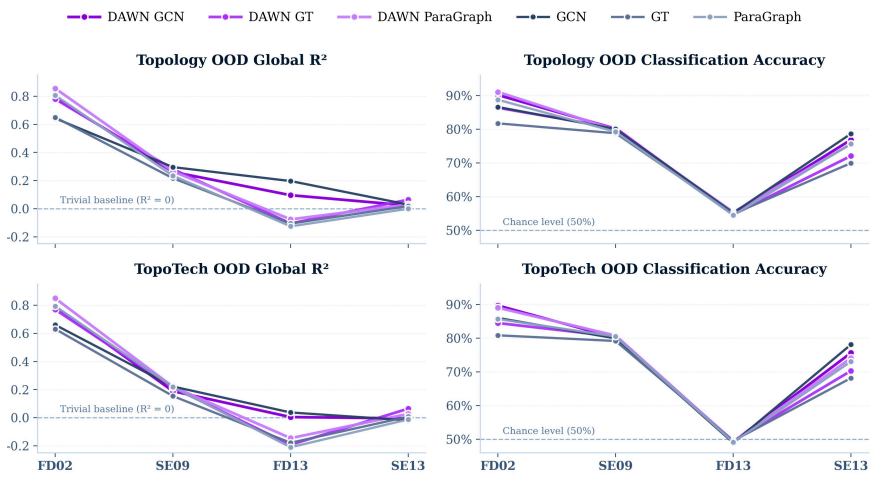

Figure 2: ArrivalFM-OTA **per-topology performance breakdown** across **Topology OOD** and **TopoTech OOD** holdout splits. We report mean global metric R^2 and mean global classifier accuracy across **10** random initializations each for 6 neural surrogate architectures, evaluated across held-out topologies: `OTA_FD02` and `OTA_SE09` (44,768 datapoints each), and `OTA_FD13` and `OTA_SE13` (89,556 datapoints each), omitting error distributions for visual clarity. Consult Appendix B for extensive circuit topology details.

substantially improve conductances, both exhibit higher initialization variance ($\sigma \approx 0.10$–0.21) on localized voltage and noise heads. Overall, explicit DC operating point residual supervision via DAWN is shown to consistently elevate transistor-level physical fidelity across relational backbones.

Dissecting the structural extrapolation bottleneck, Figure 2 breaks down surrogate performance across individual held-out circuits under **Topology OOD** and **TopoTech OOD**. On the 5-transistor single-stage `OTA_FD02`, models generalize with $R^2 > 0.78$ and classification $> 85\%$, where DAWN yields massive gains of $+0.13$ to $+0.15$ R^2 on **GCN** and **GT**. However, prediction fidelity degrades as topological divergence grows. On `OTA_FD13`, a differential inverter-based Nauta architecture whose transconductance mechanism fundamentally differs from training motifs, continuous regression collapses entirely with $R^2 \leq 0$ and classification plummets to $\approx 50\%$. Similarly, on the 13-transistor gain-boosted telescopic `OTA_SE13`, deep cascoding and auxiliary feedback loops suppress regression fidelity to $R^2 \leq 0.06$. Crucially, this breakdown explains why aggregate benchmark performance plateaus at $R^2 \approx 0.281$: while standard models accurately navigate continuous sizing spaces on known topologies, transferring across novel circuit connectivity causes qualitative collapse across all tested model families, highlighting structural generalization as an open frontier for neural circuit performance surrogates. Additionally, exhaustive evaluation profiles covering all 53 global and 15 per-device metrics across baselines are serialized per-topology and overall as structured `.csv` summaries in the supplementary material.

6 Conclusion

We introduce **ArrivalFM-OTA**, an open benchmark spanning 2,649,573 industrial-grade Spectre simulations across 28 OTA topologies and 10 FinFET nodes (7 nm to 20 nm) under continuous environmental sweeps. We also propose **DAWN**, a nodal residual framework grounding global circuit metrics in intermediate transistor operating points to improve surrogate fidelity across backbones by up to $+0.055$ R^2. Our multi-seed evaluations reveal a sharp representational asymmetry: while models navigate continuous technology sweeps ($R^2 \approx 0.764$), extrapolating to unseen circuit topologies hits a persistent bottleneck at $R^2 \approx 0.281$, establishing structural generalization as the central open frontier for physical circuit learning. To support future research, our dataset includes additional stochastic and layout metrics detailed in Appendix E.3 ($A_{\text{total}}, \sigma_{V_{\text{th}}}, \Xi_{\text{fin}}, \Omega_{\text{jitter}}, \Psi_{\text{stress}}$). Open-sourcing our simulation pipeline and pre-trained checkpoints provides an accessible community foundation to scale surrogate modeling across broader circuit classes and advance inverse sizing and generative topology synthesis.

AI USE STATEMENT

In this work, we used generative AI tools for drafting sections of the manuscript, as well as research ideation and execution, specifically implementing code boilerplate, formatting documentation, and brainstorming candidate OTA architecture classes to ensure broad topological coverage. We have not used generative AI tools for generating synthetic datasets, running circuit simulations, or calculating ground-truth performance metrics, and automated mathematical derivations or proofs are not applicable to this work. Additionally, we used AI tools for related work retrieval and discovery, as well as language polishing to tighten prose and eliminate wordy phrasing. We have reviewed all AI-assisted work. Specifically, all brainstormed OTA topology classes were verified against established analog integrated circuit literature (Razavi, 2017; Carusone et al., 2012), manually schematized, and validated through extensive Cadence Spectre convergence runs to establish physical realizability before entry into our dataset. All software boilerplate was inspected, edited, tested, and integrated into our data collection, training, and inference pipelines by the authors. All drafted text and literature references were manually cross-checked against primary sources and raw data logs to ensure complete factual precision. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work complies with the ICLR Code of Ethics. Our research does not involve human subjects, personal data, or private user records; therefore, Institutional Review Board (IRB) review was not required. To democratize semiconductor research, a field historically bottlenecked by proprietary non-disclosure agreements and commercial tool costs, we plan to release our full dataset, simulation orchestration pipeline (**Contact**), and multi-seed baseline checkpoints under open-source licenses (CDLA-Permissive-2.0, MIT). All simulations utilize open-access Predictive Technology Models (PTMs) developed for public research, ensuring that our release contains no proprietary foundry trade secrets, intellectual property, or commercial process design kits (PDKs).

As fundamental building blocks of analog integrated circuits, operational transconductance amplifiers (OTAs) are general-purpose electronic components used ubiquitously across a wide variety of industries. Our benchmark focuses strictly on basic forward surrogate modeling and physical representation learning without specialized application towards weaponized or harmful implementations. We encourage responsible community adoption of our benchmark, dataset, and simulation orchestration pipeline to advance transparent electronic design automation, open-source EDA tooling, and reproducible physical machine learning.

REPRODUCIBILITY STATEMENT

To ensure unconditional scientific reproducibility and support community benchmark extension, complete protocols, code, and data assets are organized across this manuscript, its appendices, and the accompanying supplementary material:

(1) Simulation Pipeline and Dataset Generation: Technology node constants and OTA topology catalogues are formalized in Appendices A and B. Design and input parameter sampling methodologies are detailed in Section 3.1 and Appendix C. Cadence Spectre testbenches, simulation tolerances, simulation diagnostics, and full performance metric definitions are formalized in Appendices D and E. Full dataset statistics and distributions are documented in Appendix F.

(2) Baseline Training and Benchmark Protocols: Universal training hyperparameters, layer-wise message-passing equations for all baseline backbones (MLP, GCN, GT, ParaGraph), and the DAWN zero-initialized residual cascade are detailed in Section 4 and Appendix H. Official benchmark protocols are detailed in Appendix G.

(3) Code and Artifact Release: The attached supplementary `.pdf` document provides direct links to four anonymized repositories hosting the complete ArrivalFM-OTA ecosystem: (i) the full **ArrivalFM-OTA Dataset**, (ii) the **Contact** simulation orchestration and data collection pipeline, (iii) **Benchmark Baselines** containing all source code, Conda environment specifications, pre-trained model checkpoints/inference results, and (iv) **Standardized Benchmark Protocols** and templates.

ACKNOWLEDGMENTS

Acknowledgments will be made following peer review.

REFERENCES

Chris Auth, A. Aliyarukunju, M. Asoro, Daniel B. Bergstrom, Vishal B. Bhagwat, J. Birdsall, Nabhendra Bisnik, Marc Buehler, Vinay Chikarmane, G. Ding, Qiang Fu, Harry Gomez, Weimin Han, Dennis G. Hanken, M. Haran, M. Hattendorf, R. W. Heussner, Hideyo Susumu Hiramatsu, B. Ho, S. Jaloviar, Ilsup Jin, S. Joshi, S. Kirby, Sravya Kosaraju, H. Kothari, Gerald S. Leatherman, K. S. Lee, Juergen Leib, A. Madhavan, K. Marla, H. Meyer, Tony Mule, C. Parker, Srivatsan Parthasarathy, Chris Pelto, Leonard C. Pipes, I. Post, M. Prince, A. Rahman, S. Rajamani, A. R. Saha, J. D. Santos, Manisha Sharma, V. Sharma, J. Shin, P. Alladi Sinha, P. R. Smith, Matthew Sprinkle, A. St. Amour, C. Staus, Rahul Suri, D. J. Towner, A. Tripathi, Ahmet Tura, C. Ward, and Andrew Yeoh. A 10nm high performance and low-power CMOS technology featuring 3rd generation FinFET transistors, Self-Aligned Quad Patterning, contact over active gate and cobalt local interconnects. In *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017.

Jitendra Bhandari, Vineet Bhat, Yuheng He, Hamed Rahmani, Siddharth Garg, and Ramesh Karri. Masala-CHAI: A large-scale spice netlist dataset for analog circuits by harnessing ai, 2025. URL `https://arxiv.org/abs/2411.14299`.

Cadence Design Systems. *Spectre Circuit Simulator Reference, Version 19.1*. San Jose, CA, 2020.

Tony Chan Carusone, David A. Johns, and Kenneth W. Martin. *Analog Integrated Circuit Design*. John Wiley & Sons, 2nd edition, 2012. ISBN 978-0470770108.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.

Giuseppe Chiari, Michele Piccoli, and Davide Zoni. OSIRIS: Bridging analog circuit design and machine learning with scalable dataset generation. In *International Conference on Learning Representations (ICLR)*, 2026.

Lawrence T. Clark and Vinay Vashishtha. Design with sub-10 nm FinFET technologies. In *2017 IEEE Custom Integrated Circuits Conference (CICC)*, 2017.

Lawrence T. Clark, Vinay Vashishtha, Lucian Shifren, Aditya Gujja, Saurabh Sinha, Brian Cline, Chandarasekaran Ramamurthy, and Greg Yeric. ASAP7: A 7-nm FinFET Predictive Process Design Kit. *Microelectronics Journal*, 2016.

Zehao Dong, Weidong Cao, Muhan Zhang, Dacheng Tao, Yixin Chen, and Xuan Zhang. Cktgnn: Circuit graph neural network for electronic design automation. In *International Conference on Learning Representations, ICLR*, 2023.

David Eriksson, Michael Pearce, Jacob R. Gardner, Ryan Turner, and Matthias Poloczek. Scalable global optimization via local Bayesian optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, pp. 5497–5508, 2019.

Shaoze Fan, Haoshu Lu, Shun Zhang, Ningyuan Cao, Xin Zhang, and Jing Li. Graph-transformer-based surrogate model for accelerated converter circuit topology design. In *Proceedings of the 61st ACM/IEEE Design Automation Conference DAC*, 2024.

Jian Gao, Weidong Cao, Junyi Yang, and Xuan Zhang. AnalogGenie: A Generative Engine for Automatic Discovery of Analog Circuit Topologies. In *International Conference on Learning Representations (ICLR)*, 2025.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning (ICML)*, 2017.

Jason Ho, James A. Boyle, Linshen Liu, and Andreas Gerstlauer. Lasana: Large-scale surrogate modeling for analog neuromorphic architecture exploration. In *2025 ACM/IEEE 7th Symposium on Machine Learning for CAD (MLCAD)*, pp. 1–8, 2025. doi: 10.1109/MLCAD65511.2025.11189133.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR*, 2017.

Joshua D. Knowles. ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 2006.

Kishor Kunal, Tonmoy Dhar, Meghna Madhusudan, Jitesh Poojary, Arvind Sharma, Wenbin Xu, Steven M. Burns, Jiang Hu, Ramesh Harjani, and Sachin S. Sapatnekar. Gana: Graph convolutional network based automated netlist annotation for analog circuits. In *2020 Design, Automation Test in Europe Conference Exhibition (DATE)*, pp. 55–60, 2020.

Markus Leibl, Filipe Azevedo, Stefan Uhlich, Andrea Bonetti, Arun Venkitaraman, Chia-Yu Hsieh, Phuoc Pham, Lorenzo Servadei, Helmut Graeb, and Ricardo Martins. LOADBench: A Large Open Analog Dataset and Benchmark for Machine Learning in IC Design. In *ACM/IEEE International Symposium on Machine Learning for CAD (MLCAD)*, 2026.

Jintao Li, Haochang Zhi, Ruiyu Lyu, Wangzhen Li, Zhaori Bi, Keren Zhu, Yanhan Zeng, Weiwei Shan, Changhao Yan, Fan Yang, Yun Li, and Xuan Zeng. AnalogGym: An open and practical testing suite for analog circuit synthesis. In *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design*, 2024.

Asal Mehradfar, Xuzhe Zhao, Yue Niu, Sara Babakniya, Mahdi Alesheikh, Hamidreza Aghasi, and Salman Avestimehr. AICircuit: A multi-level dataset and benchmark for AI-driven analog integrated circuit design. In *NeurIPS Machine Learning and the Physical Sciences Workshop*, 2024.

Asal Mehradfar, Xuzhe Zhao, Yilun Huang, Emir Ceyani, Yankai Yang, Shihao Han, Hamidreza Aghasi, and Salman Avestimehr. FALCON: An ML Framework for Fully Automated Layout-Constrained Analog Circuit Design. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

S. Natarajan, M. Agostinelli, S. Akbar, M. Bost, A. Bowonder, V. Chikarmane, S. Chouksey, A. Dasgupta, K. Fischer, Q. Fu, T. Ghani, M. Giles, S. Govindaraju, R. Grover, W. Han, D. Hanken, E. Haralson, M. Haran, M. Heckscher, R. Heussner, P. Jain, R. James, R. Jhaveri, I. Jin, H. Kam, E. Karl, C. Kenyon, M. Liu, Y. Luo, R. Mehandru, S. Morarka, L. Neiberg, P. Packan, A. Paliwal, C. Parker, P. Patel, R. Patel, C. Pelto, L. Pipes, P. Plekhanov, M. Prince, S. Rajamani, J. Sandford, B. Sell, S. Sivakumar, P. Smith, B. Song, K. Tone, T. Troeger, J. Wiedemer, M. Yang, and K. Zhang. A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size. In *2014 IEEE International Electron Devices Meeting (IEDM)*, 2014.

A.B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, 1995.

Souradip Poddar, Youngmin Oh, Yao Lai, Hanqing Zhu, Bosun Hwang, and David Z. Pan. INSIGHT: A Universal Neural Simulator Framework for Analog Circuits with Autoregressive Transformers. In *ACM/IEEE Design Automation Conference (DAC)*, 2025.

Behzad Razavi. *Design of Analog CMOS Integrated Circuits*. McGraw-Hill Education, 2nd edition, 2017.

Haoxing Ren, George F. Kokai, Walker J. Turner, and Ting-Sheng Ku. ParaGraph: Layout parasitics and device parameter prediction using graph neural networks. In *57th ACM/IEEE Design Automation Conference (DAC)*, 2020.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter W. Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning (ICML)*, 2020.

Thomas J. Santner, Brian J. Williams, and William I. Notz. *The Design and Analysis of Computer Experiments*. Springer, 2003.

Huiling Shang, S. Jain, Emmanuel Josse, E. Alptekin, M. H. Nam, S. W. Kim, K. H. Cho, I. Kim, Y. Liu, X. Yang, X. Wu, J. Ciavatti, N. S. Kim, R. Vega, L. Kang, H. V. Meer, S. Samavedam, Menşure Çelik, Steven Soss, and R. Sampson. High performance bulk planar 20nm CMOS technology for low power mobile applications. In *Symposium on VLSI Technology (VLSIT)*, 2012.

Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjin Wang, and Yu Sun. Masked label prediction: Unified message passing model for semi-supervised classification. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI*, 2021.

Teja Singh, Sundar Rangarajan, Deepesh John, Russell Schreiber, Spence Oliver, Rajit Seahra, and Alex Schaefer. 2.1 Zen 2: The AMD 7nm Energy-Efficient High-Performance x86-64 Microprocessor Core. In *2020 IEEE International Solid- State Circuits Conference (ISSCC)*.

Saurabh Sinha, Greg Yeric, Vikas Chandra, Brian Cline, and Yu Cao. Exploring sub-20nm FinFET design with predictive technology models. In *ACM/IEEE Design Automation Conference (DAC)*, 2012.

I.M Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 1967.

Zhuofu Tao, Yichen Shi, Yiru Huo, Rui Ye, Zonghang Li, Li Huang, Chen Wu, Na Bai, Zhiping Yu, Ting-Jung Lin, and Lei He. Amsnet: Netlist dataset for ams circuits. In *2024 IEEE International LLM-Aided Design Workshop (LAD)*, 2024.

Shien-Yang Michael Wu, Chih-Yung Lin, Mu-Chi Chiang, J. J. Liaw, J.-Y. Cheng, S.-H. Yang, Mong-Song Liang, Takumi Miyashita, C.-H. Tsai, B.-C. Hsu, H.-Y. Chen, T. Yamamoto, S. Y. S. Chang, Vincent S. Chang, Chang-Yun Chang, J.-H. Chen, Hsin-Chi Chen, K.-C. Ting, Y.-K. Wu, Kuo-Hua Pan, R.-F. Tsui, C.-H. Yao, Paul Chang, Hanh Lien, Tzyh-Cheang Lee, H. M. Lee, W. H. Chang, T. Chang, R. Chen, Mu-Shih Yeh, Chen-Shien Chen, Yuan-Hung Chiu, Y.-H. Chen, H.-C. Huang, Y.-C. Lu, Chih-Sheng Chang, Ming-Hsing Tsai, C.-C. Liu, Kuei-Shun Chen, C.-C. Kuo, H. T. Lin, S. M. Jang, and Y. Ku. A 16nm FinFET CMOS technology for mobile SoC and computing applications. In *2013 IEEE International Electron Devices Meeting (IEDM)*, 2013.

Xiaomeng Yang, Jian Gao, Yanzhi Wang, and Xuan Zhang. ZeroSim: Zero-shot analog circuit evaluation with unified transformer embeddings. In *IEEE/ACM International Conference On Computer Aided Design*, 2025.

Hongyi Zhang, Yann N. Dauphin, and Tengyu Ma. Fixup Initialization: Residual Learning Without Normalization. In *International Conference on Learning Representations (ICLR)*, 2019.

A TECHNOLOGY NODES AND CONSTANT DERIVATIONS

All design candidates in our dataset are parameterized across a spectrum of FinFET technology nodes. Table 6 outlines foundational device-level constants and derived parasitic parameters utilized to model performance scaling in this work from 20nm down to 7nm high performance (HP) and low standby power (LSTP) nodes. Transistor gate length bounds $L_{\min}$ and $L_{\max}$ in addition to nominal supply voltage $V_{\mathrm{DD_nom}}$ are extracted from corresponding predictive technology model (PTM) device cards (Sinha et al., 2012). To accommodate future research into statistical mismatch perturbations without extrapolating beyond the simulator's physical limits, sampled operational bounds for $L_{\min}$ and $L_{\max}$ are guardbanded by 5% inward from absolute PTM boundaries.

To abstract physical variance across PTM nodes, a 17-dimensional global environment vector as detailed in Table 3 is extracted directly from corresponding model cards, isolating principal physical, electrostatic, and material properties of the FinFET process. Parameters are divided into six core categories:

- **Geometry:** Fin height h_{fin}, fin thickness t_{fin}, fin pitch F_{pitch}, drain spacer length L_{dg}, and silicide thickness T_{sili}.
- **Dielectrics:** Equivalent oxide thickness EOT and gate work functions $\Phi_{\mathrm{g_nmos}}$, $\Phi_{\mathrm{g_pmos}}$.
- **Doping:** The body and source-to-drain doping concentrations N_{body} and N_{sd}.
- **Transport:** Low-field carrier mobility μ_0 and the saturation velocities $V_{\mathrm{sat_nmos}}$ and $V_{\mathrm{sat_pmos}}$.
- **Parasitics:** Specific contact resistivities $\rho_{\mathrm{c_nmos}}$, $\rho_{\mathrm{c_pmos}}$.
- **Short-Channel Effects:** Drain-induced barrier lowering coefficients η_{0_nmos}, η_{0_pmos}.

The contacted poly pitch CPP and metal pitch P_{metal} are synthesized from published ASAP7 (Clark et al., 2016), Intel 10nm (Auth et al., 2017), AMD Zen (Singh et al.), TSMC 16nm (Wu et al., 2013), and IBM 20nm (Shang et al., 2012) process reports. Additionally, the Pelgrom coefficient $A_{V_{\mathrm{th}}}$ which characterizes threshold voltage mismatch is extracted from literature covering 22nm to sub-10nm process nodes (Natarajan et al., 2014; Clark & Vashishtha, 2017). Because only PTM-derived parameters are utilized for prediction modeling in this publication, all literature derived physical constraints are structurally isolated. They are included solely to serve as modular proxy metrics for future area and mismatch-constrained design tasks.

To ensure consistency across benchmarked nodes, we derive the capacitance density ρ_{cap} as a function of the metal pitch P_{metal}, assuming an effective interconnect height $H_{\mathrm{metal}} \approx 2 \cdot P_{\mathrm{metal}}$ and an inter-line spacing $S \approx 0.5 \cdot P_{\mathrm{metal}}$. The relative permittivity of the metal ϵ_{r} is tuned per node reflecting the shift towards advanced ULK dielectrics in advanced processes to mitigate parasitic capacitance.

$$\rho_{\mathrm{cap}} \approx \frac{\epsilon_0 \cdot \epsilon_{\mathrm{r}} \cdot H_{\mathrm{metal}}}{S \cdot P_{\mathrm{metal}}} = \frac{\epsilon_0 \cdot \epsilon_{\mathrm{r}} \cdot (2 \cdot P_{\mathrm{metal}})}{(0.5 \cdot P_{\mathrm{metal}}) \cdot P_{\mathrm{metal}}} = \frac{4 \cdot \epsilon_0 \cdot \epsilon_{\mathrm{r}}}{P_{\mathrm{metal}}} \tag{10}$$

We then derive the resistive coefficient κ_{res} in units of m^2/Ω given the minimum metal width $W_{\min} \approx 0.5 \cdot P_{\mathrm{metal}}$ and the sheet resistance $R_{\mathrm{s}} \approx \rho_{\mathrm{m}}/H_{\mathrm{metal}}$ where interconnect height $H_{\mathrm{metal}} \approx 2 \cdot P_{\mathrm{metal}}$. The effective resistivity of the metal ρ_{m} is calibrated per node to account for transition from pure copper to more advanced metallization schemes like cobalt as feature sizes scale below 16 nm, reflecting increased electron scattering at grain boundaries.

$$\kappa_{\mathrm{res}} \approx \frac{W_{\min}^2}{R_{\mathrm{s}}} = \frac{(0.5 \cdot P_{\mathrm{metal}})^2}{\rho_{\mathrm{m}}/(2 \cdot P_{\mathrm{metal}})} = \frac{0.25 \cdot P_{\mathrm{metal}}^2 \cdot 2 \cdot P_{\mathrm{metal}}}{\rho_{\mathrm{m}}} = \frac{0.5 \cdot P_{\mathrm{metal}}^3}{\rho_{\mathrm{m}}} \tag{11}$$

Table 6: Technology node constants utilized for performance modeling.

	7 nm HP	**7 nm LSTP**	**10 nm HP**	**10 nm LSTP**	**14 nm HP**	**14 nm LSTP**	**16 nm HP**	**16 nm LSTP**	**20 nm HP**	**20 nm LSTP**
L_{min} (nm)	11	11	11	11	11	11	11	11	11	11
L_{max} (nm)	28	28	28	28	28	28	28	28	22	22
$V_{\text{DD_nom}}$ (V)	0.7	0.7	0.75	0.75	0.8	0.8	0.85	0.85	0.9	0.9
h_{fin} (nm)	18	18	21	21	23	23	26	26	28	28
t_{fin} (nm)	6.5	6.5	8	8	10	10	12	12	15	15
EOT (nm)	0.62	0.84	0.68	0.88	0.75	0.92	0.8	1	0.84	1.1
$\Phi_{\text{g_nmos}}$ (eV)	4.4243	4.6136	4.4212	4.6043	4.4187	4.603	4.4061	4.5887	4.3752	4.5568
$\Phi_{\text{g_pmos}}$ (eV)	4.7467	4.5645	4.7509	4.5662	4.7518	4.5743	4.7642	4.5885	4.7976	4.6215
N_{body} (cm^{-3})	1×10^{16}	1×10^{16}	2.5×10^{16}	2.5×10^{16}	5×10^{16}	5×10^{16}	1×10^{17}	1×10^{17}	5×10^{17}	5×10^{17}
L_{dg} (nm)	2	2	2	2	2.5	2.5	3	3	3	3
$\mu_{\text{0_nmos}}$ ($\text{cm}^2/\text{V} \cdot \text{s}$)	650	650	568	568	520	520	448	448	380	380
$\mu_{\text{0_pmos}}$ ($\text{cm}^2/\text{V} \cdot \text{s}$)	420	400	376	360	336	330	300	300	260	260
$V_{\text{sat_nmos}}$ (cm/s)	1.6×10^7	1.6×10^7	1.32×10^7	1.32×10^7	1.1×10^7	1.1×10^7	9×10^6	9×10^6	8.5×10^6	8.5×10^6
$V_{\text{sat_pmos}}$ (cm/s)	1.53×10^7	1.53×10^7	1.29×10^7	1.29×10^7	1.05×10^7	1.05×10^7	8.8×10^6	8.8×10^6	8.4×10^6	8.4×10^6
T_{sili} (nm)	7	7	7	7	8	8	9	9	10	10
$\rho_{\text{c_nmos}}$ ($\Omega \cdot \text{cm}^2$)	4×10^{-9}	4×10^{-9}	6×10^{-9}	6×10^{-9}	8×10^{-9}	8×10^{-9}	1.5×10^{-8}	1.5×10^{-8}	4×10^{-8}	4×10^{-8}
$\rho_{\text{c_pmos}}$ ($\Omega \cdot \text{cm}^2$)	5×10^{-9}	5×10^{-9}	7×10^{-9}	7×10^{-9}	9×10^{-9}	9×10^{-9}	2×10^{-8}	2×10^{-8}	5×10^{-8}	5×10^{-8}
F_{pitch} (nm)	22	22	28	28	32	32	42	42	60	60
$\text{ETA}_{\text{0_nmos}}$	0.6785	0.39	0.6778	0.4079	0.5063	0.5141	0.5261	0.7358	0.5063	0.6656
$\text{ETA}_{\text{0_pmos}}$	0.97	0.54	0.93	0.57	0.767	0.74	0.778	0.9925	0.71	0.85
CPP (nm)	57	54	54	54	78	78	90	90	86	86
P_{metal} (nm)	57	36	36	36	64	64	64	64	64	64
ρ_{cap} (F/m^2)	3.84×10^{-3}	2.42×10^{-3}	3.84×10^{-3}	3.84×10^{-3}	2.16×10^{-3}	2.16×10^{-3}	2.16×10^{-3}	2.16×10^{-3}	2.16×10^{-3}	2.16×10^{-3}
κ_{res} (m^2/Ω)	1.94×10^{-19}	7.72×10^{-19}	2.48×10^{-19}	2.48×10^{-19}	2.18×10^{-19}	2.18×10^{-19}	2.91×10^{-19}	2.91×10^{-19}	3.97×10^{-19}	3.97×10^{-19}
$A_{V_{\text{th}}}$ ($\text{V} \cdot \text{m}$)	1.1×10^{-9}	1.1×10^{-9}	1.1×10^{-9}	1.1×10^{-9}	1.07×10^{-9}	1.07×10^{-9}	1.07×10^{-9}	1.07×10^{-9}	2.08×10^{-9}	2.08×10^{-9}

B CIRCUIT TOPOLOGY CATALOG

This section details the 28 OTA topologies found in this work, categorized by output stage, difficulty tier, and empirical sample density ρ_{EMP} as defined in Section 3.1. Prioritizing core signal routing characteristics, visual schematics in this section utilize standard three-terminal device abstractions, with default body-to-rail bulk connections omitted. We directly incorporate 13 OTA topologies from GANA (Kunal et al., 2020), adopting an annotation scheme informed by both GANA's taxonomy and canonical analog circuit literature (Razavi, 2017).

B.1 EASY SINGLE-ENDED OUTPUT

10 OTA topologies detailed in Table 7 serve as the foundational tier of the single-ended output design space, defined by minimal design parameters and straightforward single-stage routing networks. These topologies achieve an empirical sample density of $\rho_{\text{EMP}} > 512$ and contribute $44{,}768$ design datapoints each to our ArrivalFM-OTA benchmark task, providing a robust foundational understanding of OTA dynamics before scaling to higher complexities. Representative topologies in this tier are illustrated in Figures 3, 4, and 5.

Table 7: Fundamental summary of the 10 easy single-ended ouptut topologies in our dataset. Transistors, resistors (R), and capacitors (C) denote component counts for each circuit.

Topology ID	Description	Transistors	R	C	ρ_{EMP}	Design Challenges
OTA_SE01	NMOS Input Single Stage	5	0	0	998	PMOS load mirror pole constraints
OTA_SE02	PMOS Input Single Stage	5	0	0	998	PMOS input low transconductance penalties
OTA_SE03	NMOS Input Telescopic	7	0	0	593	Restricted output voltage swing
OTA_SE04	PMOS Input Telescopic	7	0	0	569	Cascode node parasitic pole degradation
OTA_SE05	NMOS Input Low Voltage Cascode Current Mirror	7	0	0	895	Wide swing mirror saturation balancing
OTA_SE07	NMOS Input Recycling Folded Cascode	13	0	0	521	Recycling mirror mismatch sensitivity
OTA_SE08	PMOS Input Recycling Folded Cascode	11	0	0	523	Multipath pole-zero doublet formation
OTA_SE09	NMOS Input Miller Compensated	7	0	1	556	Unnulled RHP zero phase degradation
OTA_SE11	NMOS Input Class AB	9	0	0	650	Push-pull quiescent bias sensitivity
OTA_SE12	NMOS Input Class AB Push Pull Inverter	7	0	1	624	Inverter stage PVT bias sensitivity

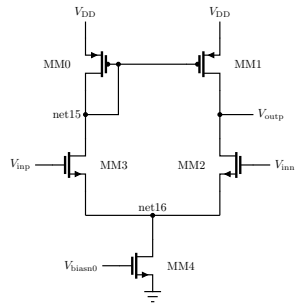

Figure 3: OTA_SE01, a 5-transistor single-stage OTA topology utilizing a standard NMOS-input and active-loaded differential pair. In single-ended feedback, differential branch currents are converted into a singular output at V_{outp}, which stabilizes its operating point via negative feedback routed directly to the inverting input terminal V_{inn}.

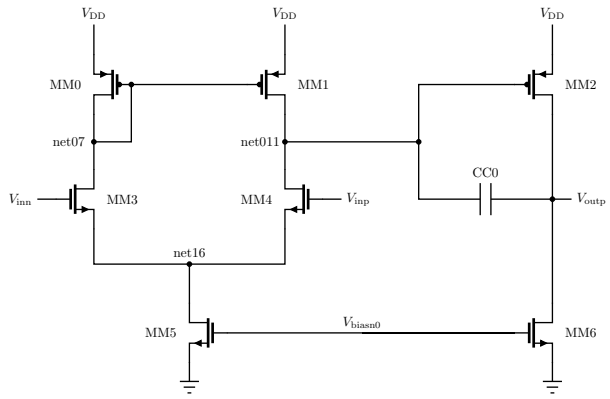

Figure 4: OTA_SE09, a 7-transistor OTA featuring an NMOS differential input pair, a PMOS active load, and a PMOS common-source second stage with Miller compensation capacitor CC0. Serving as a held-out evaluation circuit for **Topology OOD** and **TopoTech OOD** evaluation splits in our ArrivalFM-OTA benchmark, this circuit provides a generalization test coupling many subcircuit primitives including differential pairs, active current mirrors, and a miller capacitor. This demands surrogate models to accurately predict performance metrics with training exposure to miller compensation techniques (OTA_SE10, OTA_FD08, OTA_FD09).

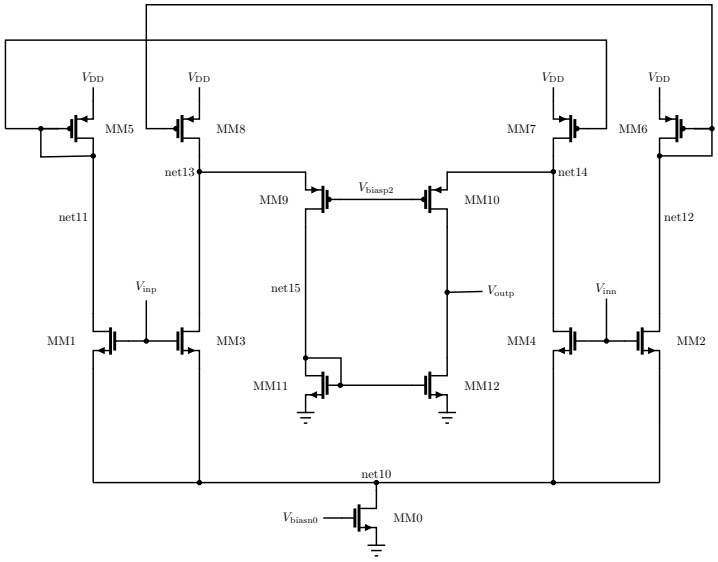

Figure 5: OTA_SE07, an NMOS input recycling folded cascode amplifier. This topology serves as a rigorous benchmark for sampling algorithms due to its 14+ dimensional parameter space and highly coupled, nonlinear operating regions. Stacked cascode devices and nested current recycling paths create narrow valid saturation hypervolumes where naive random sampling fails.

B.2 Easy Fully Differential Output

Transitioning into the fully differential domain, 3 OTA topologies detailed in Table 8 establish a low-dimensional baseline defined by an empirical sample density of $\rho_{\text{EMP}} > 512$. These topologies contribute $44{,}768$ design datapoints each to our ArrivalFM-OTA benchmark task, and provide an environment to isolate and evaluate differential dynamics such as common-mode feedback (CMFB) behaviour before introducing the complexities of multi-stage architectures. A representative baseline for this tier is illustrated in Figure 6.

Table 8: Fundamental summary of the 3 easy fully differential output topologies in the ArrivalFM-OTA dataset. Transistors, resistors (R), and capacitors (C) denote component counts for each circuit.

Topology ID	Description	Transistors	R	C	ρ_{EMP}	Design Challenges
OTA_FD01	NMOS Input Single Stage	5	0	0	942	CMFB sensing network output loading
OTA_FD02	PMOS Input Single Stage	5	0	0	924	Low transconductance CMFB loop stabilization
OTA_FD06	Complementary Input Class AB Inverter	4	0	0	1638	Tail-less quiescent bias sensitivity

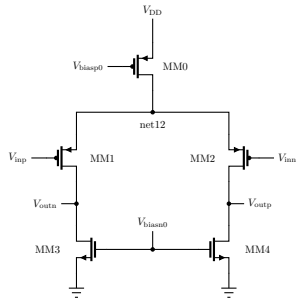

Figure 6: OTA_FD02, a fully-differential Class II OTA topology utilizing a standard PMOS-input, active-loaded differential pair, and a CMFB loop across the output terminals. Unlike single-ended designs, fully differential operation requires 2 decoupled feedback mechanisms. External differential feedback governs closed-loop signal gain across V_{inp} and V_{inn}, while an auxiliary Common-Mode Feedback (CMFB) loop actively senses the output average between the V_{outp} and V_{outn} terminals, injecting a control voltage $V_{\text{cm_ctrl}}$ to maintain the common-mode reference V_{CM}. Additionally, OTA_FD02 is a held-out test topology in our **Topology OOD** and **TopoTech OOD** evaluation splits to evaluate structural transferability, testing whether surrogate models trained on an equivalent NMOS-input architecture (OTA_FD01) can successfully generalize zero-shot to complementary PMOS-input dynamics.

B.3 Medium Single-Ended Output

4 OTA topologies depicted in Table 9 represent an escalation in complexity within the single-ended output design space, introducing advanced structures such as auxiliary gain-boosters and complementary rail-to-rail input pairs. As transistor counts expand, these topologies are defined by an empirical sample density of $128 < \rho_{\text{EMP}} < 512$, and contribute $89,536$ design datapoints each to our ArrivalFM-OTA benchmark. This tier serves as a critical intermediate benchmark for models to learn intricate multi-loop dependencies. Representative baseline topologies for this tier are illustrated in Figures 7 and 8.

Table 9: Fundamental summary of the 4 medium single-ended output topologies in the ArrivalFM-OTA dataset. Transistors, resistors (R), and capacitors (C) denote component counts for each circuit.

Topology ID	Description	Transistors	Resistors	Capacitors	ρ_{EMP}	Design Challenges
OTA_SE06	NMOS Input Cascode Current Mirror	15	0	0	305	Cascode mirror signal path pole degradation
OTA_SE10	PMOS Input Nulling Miller Compensated	7	1	1	366	RHP zero nulling resistor tracking
OTA_SE13	NMOS Input Gain Boosted Telescopic	13	0	0	207	Auxiliary loop pole-zero doublet formation
OTA_SE15	Complementary Rail-to-Rail Input Folded Cascode	14	0	0	237	Crossover region transconductance variation

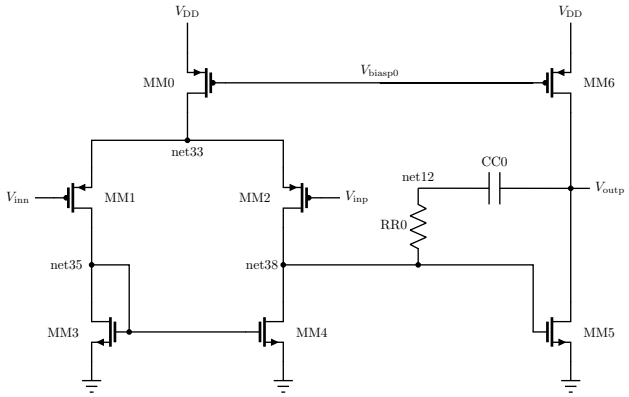

Figure 7: `OTA_SE10`, a medium single-ended output topology utilizing a nulling-based compensation network. This configuration requires a strict balance of the nulling resistor RR0 against the second-stage transconductance to prevent phase degradation.

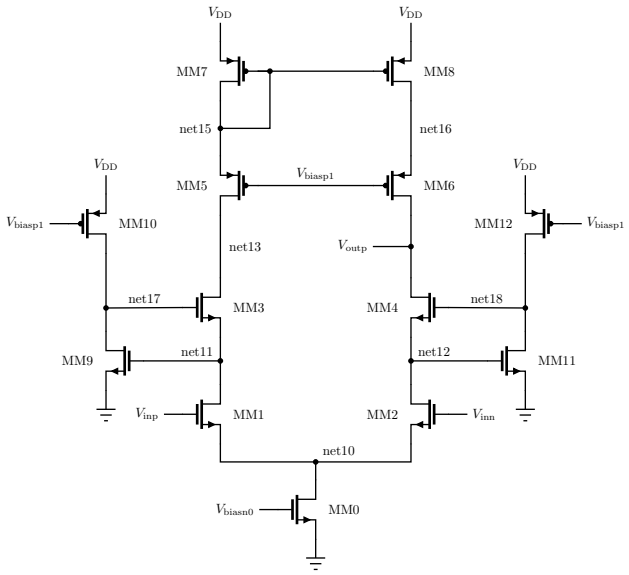

Figure 8: `OTA_SE13`, a 13-transistor gain-boosted telescopic cascode OTA with single-ended output. Designated as a held-out test topology in our **Topology OOD** and **TopoTech OOD** benchmark evaluation splits, `OTA_SE13` presents a challenging structural regime: unlike canonical multi-stage OTAs, the active auxiliary feedback loops introduce high-order frequency-domain complexities, specifically auxiliary-loop pole-zero doublet formation. Accurately predicting AC and transient metrics for this architecture requires surrogate models to infer without an abundant exposure to gain-boosted graph structures.

B.4 Medium Fully Differential Output

There are 8 medium fully differential output topologies detailed in Table 10, defined by an empirical sample density of $128 < \rho_{\text{EMP}} < 512$ contributing $89,536$ design datapoints each to our ArrivalFM-OTA benchmark. These topologies require the coordination of common-mode feedback (CMFB) networks across cascaded stages, or feature highly constrained single-stage structures such as intricate nulling, recycling, or telescopic networks, in addition to specialized inverter-based topologies as shown in Figure 9. This provides a robust environment for models to learn global differential symmetry and local node stability within a highly coupled parameter space.

Table 10: Summary of the 8 medium fully differential output topologies in the ArrivalFM-OTA dataset. Transistors, resistors (R), and capacitors (C) denote component counts for each circuit.

Topology ID	Description	Transistors	Resistors	Capacitors	ρ_{EMP}	Design Challenges
OTA_FD03	NMOS Input Telescopic	9	0	0	346	Differential cascode headroom constraint
OTA_FD04	PMOS Input Folded Cascode	11	0	0	307	Folding node non-dominant pole degradation
OTA_FD05	NMOS Input Recycling Folded Cascode	15	0	0	259	CMFB coupled recycling mirror mismatch
OTA_FD07	NMOS Input Wide Swing Current Mirror	13	0	0	282	CMFB driven wide-swing saturation margin
OTA_FD08	PMOS Input Two Stage Miller	9	0	2	489	Miller coupled multi-loop CMFB stability
OTA_FD09	NMOS Input Nulling Two Stage Miller	9	2	2	381	CMFB modulated RHP zero nulling
OTA_FD10	Complementary Rail-to-Rail Input Current Mirror	18	0	2	249	Rail-to-rail crossover transconductance variation
OTA_FD13	Complementary Input Nauta Inverter	12	0	0	366	Cross-coupled internal CMFB stability

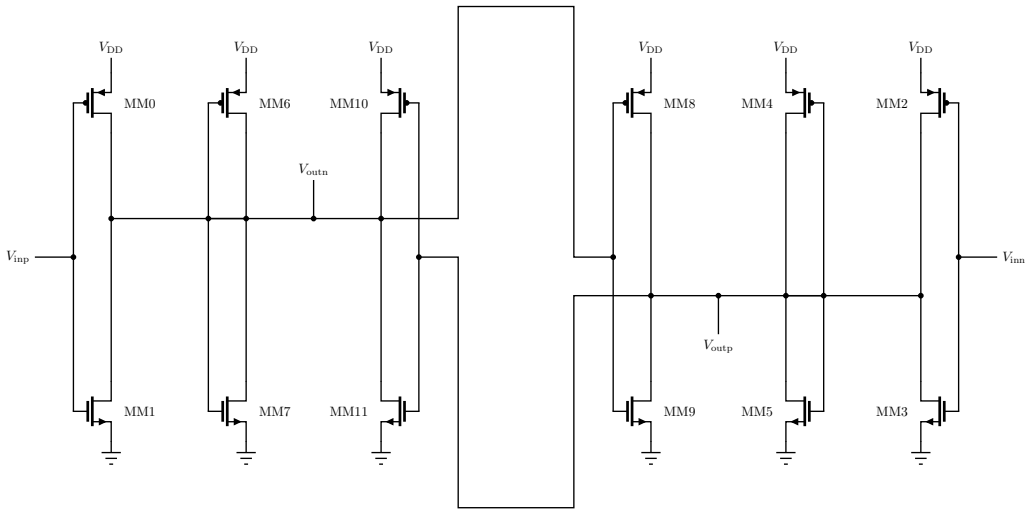

Figure 9: OTA_FD13, a Nauta inverter OTA topology. Designated as a held-out test topology in our **Topology OOD** and **TopoTech OOD** benchmark evaluation splits, it represents the most formidable architecture in our holdout split. As the only fully differential output topology operating in the complete absence of an explicit Common-Mode Feedback (CMFB) loop, surrogate models must rely entirely on inductive biases learned from a subset of inverter-based training topologies (OTA_SE12, OTA_FD06 to successfully infer performance metrics.

B.5 HARD FULLY DIFFERENTIAL TOPOLOGIES

As detailed in Table 11, 3 hard fully differential topologies are defined by an empirical sample density of $\rho_{\mathrm{EMP}} < 128$ and contribute $134,304$ design points each to the ArrivalFM-OTA benchmark. The primary design bottleneck at this scale is the severe entanglement of global and local feedback loops. Models must understand differential multi-pole dynamics in relation to intricate capacitor and resistor networks while coordinating multiple interdependent common-mode feedback (CMFB) loops across stages. A representative baseline for this tier is illustrated in Figure 10.

Table 11: Fundamental summary of the four Class VI OTA topologies in the ArrivalFM-OTA dataset.

Topology ID	Description	Transistors	Resistors	Capacitors	ρ_{EMP}	Design Challenges
OTA_FD11	Complementary Rail-to-Rail Input Two Stage Class AB	18	0	4	157	Class AB floating bias and CMFB multi-loop contention
OTA_FD12	NMOS Input Two Stage Cascode Mirror	13	2	2	109	Cascode mirror pole degradation under CMFB regulation
OTA_FD14	NMOS Input Gain Boosted Telescopic	21	0	0	57	Auxiliary loop pole-zero doublet and CMFB headroom collision

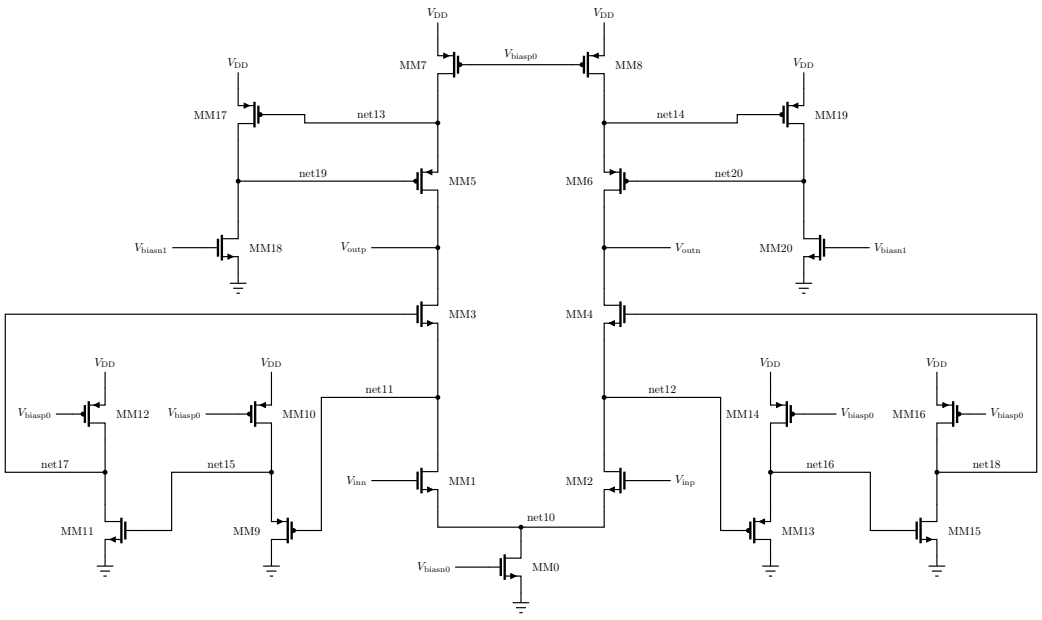

Figure 10: `OTA_FD14`, a 21-transistor hard fully differential topology in a gain-boosting configuration, containing highly nested feedback dependencies and high-order pole-zero doublets generated by its dual active regulation loops, which uniquely combine two-stage level-shifted NMOS with single-stage PMOS boosters.

C SAMPLING METHODOLOGIES

C.1 INPUT PARAMETER BOUNDS

To ensure the generation of physically realizable OTA designs, the input parameter search space must be rigorously constrained. Table 12 defines the strict boundaries applied in all sampled input design and environment dimensions, for all topologies in our ArrivalFM-OTA dataset. By bounding variables such as device sizing and passive component values alongside environmental stresses like supply voltage and temperature, we enable systematic exploration of the analog design space while simultaneously preventing the simulation of non-physical circuit states.

Table 12: Sampling bounds for input design and environment parameters. N_{fin} and L_{trans} are sampled per transistor grouping, C_{int} is sampled per passive capacitor grouping, R_{int} is sampled per passive resistor grouping, P_{err} is sampled per component, and V_{bias} is sampled per bias source. All other input parameters are sampled once per OTA design.

Parameter	Name	Method	Lower Bound	Upper Bound
$\mathbb{B}_{\text{hp}}$	Technology Node Type	Uniform categorical sampling between high performance (HP) with a value of 1 and low standby power (LSTP) with a value of 0	0	1
$\#_{\text{node}}$	Technology Node Size	Uniform categorical sampling between 7, 10, 14, 16, and 20 nm sizes	7	20
$(n_{\text{p}}, n_{\text{n}})$	Transistor Process State (Inactive)	Kept constant at the nominal TT corner represented by $(0, 0)$	N/A	N/A
L_{trans}	Transistor Channel Length	Uniform discrete sampling in 1 nm increments; Bounds dictated by length tolerances of target technology from Table 6	L_{min}	L_{max}
N_{fin}	Number of Transistor Fins	Uniform integer sampling across bounds	1	256
C_{int}	Internal Compensation Capacitance	Uniform logarithmic E12 series sampling across bounds	100 fF	10 pF
R_{int}	Internal Nulling Resistance	Uniform logarithmic E24 series sampling across bounds	500 Ω	500 kΩ
P_{err}	Component Mismatch Error	Kept constant at ideal conditions represented by 0	N/A	N/A
V_{DD}	Supply Voltage	Uniform continuous linear sampling across bounds with $\mathbb{P} = 0.7$; Fixed at $V_{\text{DD}_{\text{nom}}}$ with $\mathbb{P} = 0.3$	$0.85V_{\text{DD_nom}}$	$1.15V_{\text{DD_nom}}$
V_{CM}	Common Mode Voltage	Uniform continuous linear sampling across bounds	$0.1V_{\text{DD}}$	$0.9V_{\text{DD}}$
V_{bias}	Bias Voltage	Uniform continuous linear sampling across bounds	$0.1V_{\text{DD}}$	$0.9V_{\text{DD}}$
T	Temperature	Uniform integer sampling across bounds with $\mathbb{P} = 0.7$; Fixed at 27 °C with $\mathbb{P} = 0.3$	−40 °C	125 °C
C_{load}	Load Capacitance	Uniform logarithmic E12 series sampling across bounds	10 fF	100 pF

C.2 Initial Design Space Exploration

To establish a comprehensive dataset for OTA circuit performance, the exploratory continuous sizing and bias space must be sampled with maximum space-filling uniformity. We employ **Randomized Quasi-Monte Carlo (RQMC) Sobol Sampling** and evaluate against a **Random Uniform Sampling** baseline, generating standardized exploratory batches of size $N = 16,384 = 2^{14}$ across distinct random seeds.

C.2.1 Random Uniform Sampling

In uniform random sampling, candidate normalized design vectors $\mathbf{X}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,D}) \in [0,1]^D$, where D denotes the active circuit parameter dimensionality, are drawn independently and identically distributed from a continuous uniform distribution:

$$\mathbf{X}_i \sim \text{Unif}\left([0,1]^D\right), \quad i = 1, 2, \ldots, N \tag{12}$$

Sampled coordinates are subsequently mapped onto physical transistor dimensions and bias voltages according to the technology bounds in Appendix C.1. Because independent uniform draws enforce no mutual spatial repulsion, random uniform sampling suffers from Poisson clumping (Santner et al., 2003). In high-dimensional spaces, independent random draws form dense localized clusters while leaving expansive subregions of the physical parameter space entirely unsampled. This produces severe local oversampling alongside persistent coverage blind spots across valid design regimes.

C.2.2 Randomized Quasi-Monte Carlo (RQMC) Sobol Sampling

To eliminate localized clustering and ensure uniform space-filling coverage across coupled device parameters, we employ Randomized Quasi-Monte Carlo (RQMC) sampling via Sobol low-discrepancy sequences (Sobol', 1967) augmented with nested digital Owen scrambling (Owen, 1995). Deterministic Sobol sequences construct (t, D)-sequences in base 2 using primitive direction polynomials. Setting the exploratory batch size to an exact power of two, $N = 2^{14} = 16,384$, guarantees that each batch forms a closed (t, m, D)-net in base 2 (with $m = 14$). Under this construction, every elementary dyadic partition of volume 2^{t-m} contains exactly 2^t sample points, guaranteeing that all lower-dimensional parameter projections remain uniformly populated without axial stratification. To obtain statistically independent batches while strictly preserving the net structure, we apply digital Owen scrambling. This applies a hierarchy of randomized digit permutations:

$$\mathbf{X}_i = \pi_{\text{Owen}}\left(\mathbf{S}_i; s\right) \quad i = 1, 2, \ldots, N \tag{13}$$

where $\mathbf{S}_i \in [0,1)^D$ denotes the ith deterministic Sobol vector and $\pi_{\text{Owen}}(\cdot; s)$ is the uniform scrambling permutation parameterized by random seed s. Across distinct seeds, RQMC generates mutually independent batches while maintaining the optimal deterministic star discrepancy bound:

$$D_N^*(\mathbf{X}) = \mathcal{O}\left(\frac{(\log N)^D}{N}\right) \tag{14}$$

Whereas standard uniform sampling exhibits a slow Monte Carlo coverage convergence rate of $\mathcal{O}(N^{-1/2})$, scrambled Sobol sequences accelerate the effective discrepancy rate toward $\mathcal{O}(N^{-1})$. In exploratory batches of $N = 16,384$, this guarantees dense, cluster-free coverage across stiff, coupled analog operating regions prior to downstream simulation.

C.3 Targeted High-Performance Design Space Exploration

While low-discrepancy uniform sampling provides broad coverage across the global input parameter space, it suffers from low sample efficiency when targeting high-performing circuit configurations. OTA circuit performance tradeoffs across open-loop gain A_{ol}, unity-gain bandwidth f_{ugbw}, total harmonic distortion THD_{dBc}, and power dissipation P exist on narrow, high-dimensional Pareto fronts. Furthermore, valid circuit operation occupies a small fraction of the raw parameter space in complex topologies due to tight DC bias and device saturation constraints. To efficiently populate the dataset with high-yielding, high-performance designs, we employ **Sobol XGBoost Surrogate Sampling** and compare against a Bayesian **Sobol TuRBO-M-Auto Sampling** baseline, both generating targeted candidate batches of size $N = 12{,}000$ across distinct random seeds.

C.3.1 RQMC SOBOL TURBO-M-AUTO SAMPLING

The RQMC Sobol TuRBO-M-Auto method adapts the Trust Region Bayesian Optimization (TuRBO) algorithm (Eriksson et al., 2019) for high-performance OTA design exploration by maintaining $M = 60$ concurrent local trust regions $(\text{TR}_1, \ldots, \text{TR}_M)$ over $N = 200$ iterations, producing $M \times N = 12{,}000$ datapoints. The workflow operates through 4 sequential phases, illustrated in Figure 11:

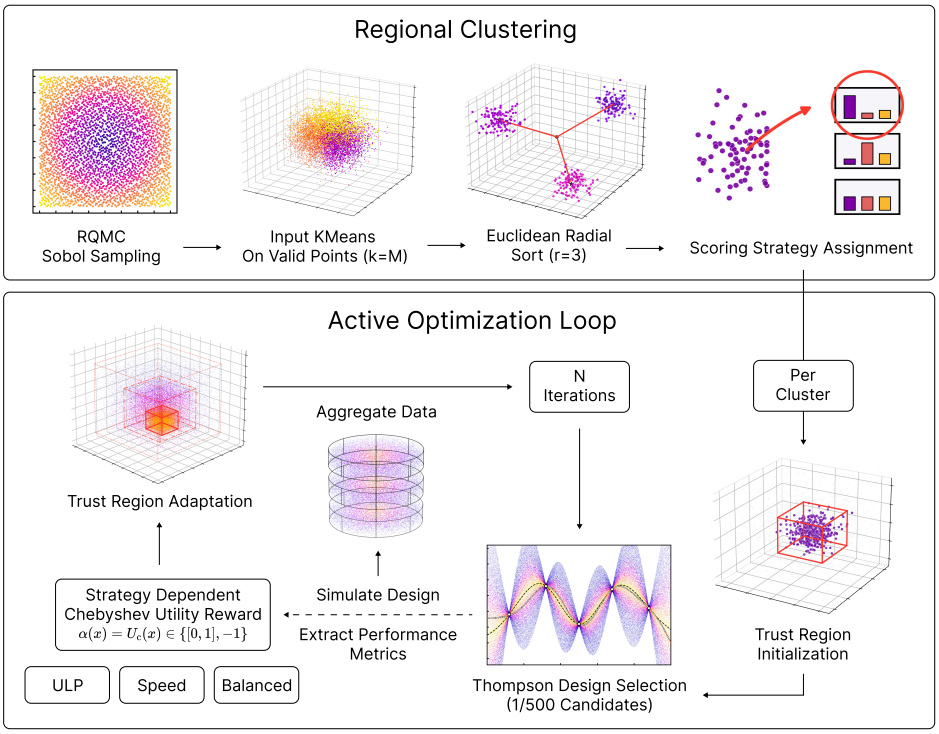

Figure 11: Overview of the RQMC Sobol TuRBO-M-Auto sampling strategy. A radial sort over MiniBatchKMeans centroids partitions initial seed points into M trust regions, each assigned one of three scalarization strategies, requiring M to be a multiple of 3. In the active Bayesian optimization loop, local Gaussian process surrogates perform discrete Thompson sampling over 500 candidate draws per step, updating trust region bounds dynamically based on strategy-dependent Chebyshev utility rewards, with non-convergent simulations assigned a penalty score of -1.

1. Radial Clustering and Initialization: Given an initial seed dataset generated with RQMC Sobol sampling (Appendix C.2.2), input parameters are normalized to the unit hypercube $[0, 1]^D$. The design space is partitioned into M distinct local neighborhoods using MiniBatchKMeans. To enforce spatial diversity across varying parameter norms, cluster centroids $\mathbf{c}_m$ are sorted radially by their Euclidean norm $d_m = \|\mathbf{c}_m\|_2 \in [d_{\min}, d_{\max}]$, seeding the initial centers $\mathbf{x}_{\text{center}}^{(m)}$.

2. Dynamic Scoring Strategy Allocation: To adequately map distinct regions of the Pareto optimal front, the radially sorted centroids are partitioned into consecutive triplets $(\text{TR}_{3j+1}, \text{TR}_{3j+2}, \text{TR}_{3j+3})$ for $j = 0, \ldots, \frac{M}{3} - 1$. Within each triplet, 3 multi-objective scalarization strategies are randomly permuted and assigned without replacement:

- **Ultra-Low Power (ULP):** Minimizes power dissipation P while maximizing open-loop gain A_{ol}.
- **Speed:** Maximizes unity-gain bandwidth f_{ugbw} and open-loop gain A_{ol}.
- **Balanced:** Minimizes P while simultaneously maximizing A_{ol} and f_{ugbw}.

Multi-objective metrics $\mathbf{y} = (y_1, \ldots, y_K)$ are scalarized into an augmented Chebyshev utility score following ParEGO (Knowles, 2006). Weights are randomized per iteration via $\mathbf{w} \sim \text{Dirichlet}(\mathbf{1})$ relative to an empirical utopian reference point $\mathbf{z}^*$:

$$d_k(\mathbf{x}) = \frac{|y_k(\mathbf{x}) - z_k^*|}{\Delta_k}, \qquad S_{\text{aug}}(\mathbf{x}) = \max_{k \in \{1,\ldots,K\}} \big(w_k \cdot d_k(\mathbf{x})\big) + \rho \sum_{k=1}^{K} w_k \cdot d_k(\mathbf{x}), \tag{15}$$

where $\Delta_k = \max(y_k) - \min(y_k)$ normalizes target scale disparities across metrics, and $\rho = 0.01$ is the augmentation coefficient ensuring strict Pareto monotonicity. To construct the final reward for Gaussian process surrogate optimization, $S_{\text{aug}}(\mathbf{x})$ is mapped to a bounded utility $U_{\text{c}}(\mathbf{x}) \in [0, 1]$:

$$U_{\text{c}}(\mathbf{x}) = 1 - \frac{S_{\text{aug}}(\mathbf{x})}{\max_{\mathbf{x}'} S_{\text{aug}}(\mathbf{x}')}. \tag{16}$$

Simulations that fail electrical convergence checks receive a default penalty score of $U_{\text{c}}(\mathbf{x}) = -1$.

3. Thompson Sampling: Within each trust region m, an independent Gaussian process surrogate is fitted over a sliding window of localized historical evaluations:

$$f^{(m)}(\mathbf{x}) \sim \mathcal{GP}\left(\mu_m(\mathbf{x}), k_m(\mathbf{x}, \mathbf{x}')\right), \tag{17}$$

where $k_m(\mathbf{x}, \mathbf{x}')$ is a Matérn-5/2 kernel with automatic relevance determination (ARD). Rather than solving an expensive continuous acquisition optimization over $[0, 1]^D$, $N_{\text{cand}} = 500$ candidate locations are randomly sampled inside the hyper-rectangular trust region:

$$\mathcal{B}_m = \left[\mathbf{x}_{\text{center}}^{(m)} - \frac{L_m}{2},\ \mathbf{x}_{\text{center}}^{(m)} + \frac{L_m}{2}\right] \cap [0, 1]^D, \tag{18}$$

where L_m denotes the current side length. Exactly 1 realization is drawn from the posterior GP distribution across all 500 points, and the candidate with the highest predicted utility is mapped to the operational parameter bounds specified in Appendix C.1 and dispatched for simulation.

4. Adaptive Trust Region Management: Each trust region dynamically updates its side length L_m according to simulation outcomes:

- **Expansion (Success):** If the evaluated candidate improves upon the current best score, $U_{\text{c}}(\mathbf{x}_{\text{cand}}) > U_{\text{c}}(\mathbf{x}_{\text{center}}^{(m)})$, the center is moved to $\mathbf{x}_{\text{cand}}$. Following $\tau_{\text{succ}} = 3$ consecutive improvements, the domain expands: $L_m \leftarrow \min(2L_m, L_{\max})$.
- **Contraction (Failure):** If candidate evaluations fail to yield an improvement for $\tau_{\text{fail}} = 5$ consecutive iterations, the domain contracts: $L_m \leftarrow L_m/2$.
- **Restart:** When L_m contracts below $L_{\min} = 2^{-7}$, the trust region is considered converged and reinitialized at a new under-explored location.

This iterative TuRBO-M loop runs across all $M = 60$ trust regions until terminating after each trust region has recommended $N = 200$ evaluations.

C.3.2 RQMC Sobol XGBoost Surrogate Sampling

The RQMC Sobol XGBoost Surrogate sampling framework is a 2-stage global screening pipeline that filters millions of candidates prior to simulation, leveraging XGBoost models (Chen & Guestrin, 2016) trained on historical simulation data to screen an expansive candidate pool of $N = 4,194,304 = 2^{22}$ points generated with RQMC Sobol sampling (Appendix C.2.2). The workflow operates through 4 sequential stages, as shown in Figure 12:

1. Binary Validity Classification Gate: To avoid sampling unviable or unconverged circuits, candidates first pass through a global XGBoost binary classifier $g_\phi(\mathbf{x}) \in [0, 1]$ trained on all initial datapoints, which outputs the probability of simulation validity $P_{\text{v}}(\mathbf{x})$. The target validity label $Y_{\text{valid}} \in \{0, 1\}$ is defined as:

$$Y_{\text{valid}} = \mathbb{I}\Big(\texttt{sim_status} = 3 \ \wedge\ V_{\text{OS}} \neq \text{NaN} \ \wedge\ V_{\text{swing_out}} \neq \text{NaN}\Big), \tag{19}$$

where $\mathbb{I}(\cdot)$ is the indicator function, `sim_status` is the simulation validity flag, V_{OS} is the input-referred offset voltage, and $V_{\text{swing_out}}$ is the output voltage swing range. The classifier is trained by minimizing binary cross-entropy loss over all historical evaluations. Candidates failing the probability threshold, $P_{\text{v}}(\mathbf{x}) < \tau_{\text{valid}}$ where $\tau_{\text{valid}} = 0.50$, are removed from the screening pool.

2. Multi-Output Performance Metric Prediction: For all candidate points passing Stage 1, a multi-output XGBoost regressor ensemble $\mathbf{h}_\theta(\mathbf{x}) = \left(h^{(1)}(\mathbf{x}), \ldots, h^{(K)}(\mathbf{x})\right)$ predicts continuous performance metrics $\hat{\mathbf{y}}(\mathbf{x}) = (\hat{y}_1, \ldots, \hat{y}_K)$ detailed in Table 13. Each estimator $h^{(k)}(\mathbf{x})$ consists

Table 13: OTA performance metrics predicted by the XGBoost regressor ensemble $\mathbf{h}_\theta$ during Sobol XGBoost Surrogate sampling. Detailed breakdowns of all metrics are provided in Appendix E.3.

Metric	Description	Units
A_{ol}	Open-Loop Gain	dB
f_{ugbw}	Unity-Gain Bandwidth	Hz
P	Power Dissipation	W
$V_{\text{swing_out}}$	Output Voltage Swing Range	V
V_{OS}	Offset Voltage	V
CMRR	Common-Mode Rejection Ratio	dB
PSRR	Power Supply Rejection Ratio	dB
THD_{dBc}	Total Harmonic Distortion	dB
$V_{\text{noise_out}}$	Total Output Integrated Noise	V_{rms}

of 300 gradient-boosted trees ($d_{\max} = 6, \eta = 0.1$) trained independently using mean squared error (MSE) loss:

$$\mathcal{L}_{\text{reg}}(\theta_k) = \frac{1}{|\mathcal{D}_k|} \sum_{\mathbf{x}_i \in \mathcal{D}_k} \left(y_{i,k} - h^{(k)}(\mathbf{x}_i; \theta_k)\right)^2, \tag{20}$$

where $\mathcal{D}_k = \{\mathbf{x}_i \mid \texttt{sim_status}_i = 3 \ \wedge \ y_{i,k} \neq \text{NaN}\}$ restricts training for metric k to fully characterizable runs containing valid non-NaN observations.

3. Subcluster Stratification: To ensure complete coverage across different technology nodes, designs, and environmental conditions, valid candidates are partitioned through a 3-level hierarchical stratification scheme (Appendix C.1):

- **Technology Node Partitioning:** The candidate pool is first grouped by technology node into 10 distinct processes (Appendix A).
- **Design Input Clustering:** Within each technology group, candidates are clustered over normalized active input design variables (L_{trans}, N_{fin}, C_{int}, R_{int}, V_{bias}) using MiniBatchKMeans into $M = 120$ primary design groups.
- **Environment Input Clustering:** Within each design group, candidates are further clustered over environmental operating conditions (V_{DD}, V_{CM}, T, C_{load}) into $N = 10$ environmental sub-clusters, yielding 1,200 total subclusters per technology node.

Sub-cluster centroids are radially sorted by Euclidean norm $d_m = \|\mathbf{c}_m\|_2$ and assigned to one of 6 strategic optimization profiles:

$$\text{Strategy} \in \left\{ \begin{array}{ll} \text{Risk-Averse ULP}, & \text{Risk-Neutral ULP}, \\ \text{Risk-Averse Speed}, & \text{Risk-Neutral Speed}, \\ \text{Risk-Averse Balanced}, & \text{Risk-Neutral Balanced} \end{array} \right\}. \tag{21}$$

Predicted metrics $\hat{\mathbf{y}}(\mathbf{x}) = (\hat{y}_1, \ldots, \hat{y}_K)$ are converted to an augmented Chebyshev utility score $U_{\text{c}}(\mathbf{x}) \in [0, 1]$ following the identical ParEGO scalarization defined in Section C.3.1, applying randomized Dirichlet weights $\mathbf{w} \sim \text{Dirichlet}(\mathbf{1})$ relative to utopian target vectors $\mathbf{z}^*$. To account for classifier uncertainty, the final strategic reward $\alpha(\mathbf{x})$ applies risk-weighted scaling:

$$\alpha(\mathbf{x}) = \begin{cases} P_{\text{v}}(\mathbf{x}) \cdot U_{\text{c}}(\mathbf{x}) & \text{if Risk-Averse,} \\ U_{\text{c}}(\mathbf{x}) & \text{if Risk-Neutral.} \end{cases} \tag{22}$$

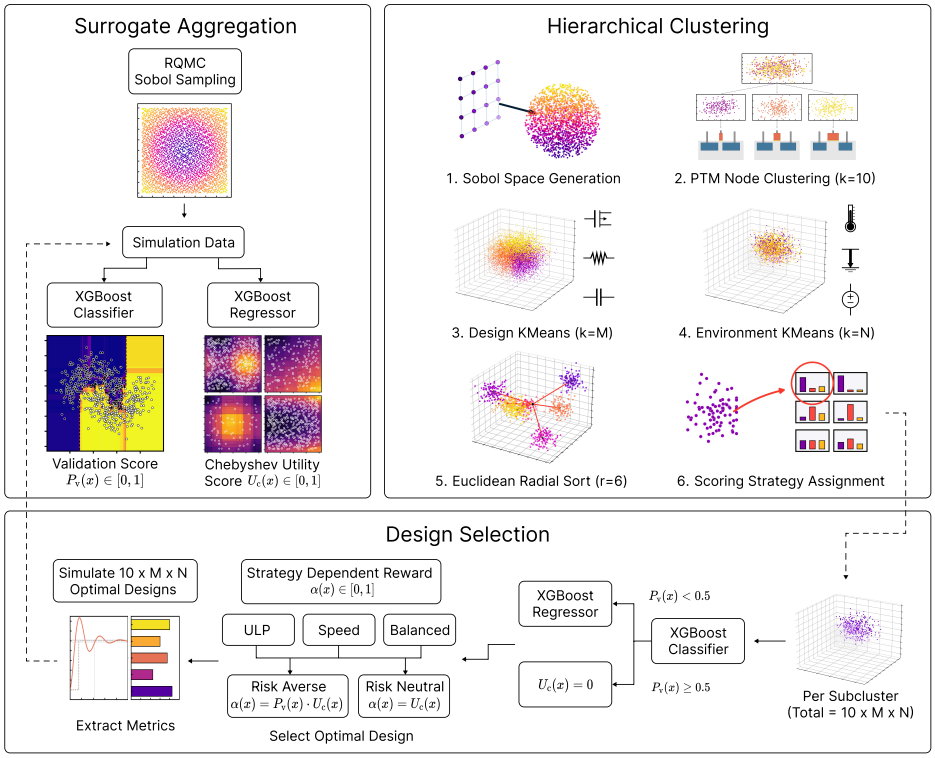

Figure 12: Overview of **RQMC Sobol XGBoost Surrogate** sampling. Initial RQMC Sobol datapoints train a dual XGBoost classifier and multi-output regressor. Un-simulated candidates are partitioned across a 3-level hierarchy (10 technology node groups $\times$ $M = 120$ design clusters $\times$ $N = 10$ environment clusters) and radially sorted into 6 strategic scoring profiles to select $10 \times M \times N = 12{,}000$ optimal candidate designs. Crucially, this creates a self-improving recursive cycle: newly simulated optimal designs have their extracted performance metrics appended back as data after every iteration, which retrains the XGBoost surrogates with high-yield data refining predictive accuracy in high-performing design regimes.

4. Stratified Tournament Selection: To prevent surrogate-guided sampling from clustering around a single high-performing design point, the candidate pool is partitioned across technology, physical dimensions, and operating environments:

1. **Technology Node Balance:** The sample budget is allocated equally across all 10 FinFET predictive technology models, assigning 1,200 candidate points to each technology node.
2. **Sizing and Bias Clustering:** Within each technology partition, MiniBatchKMeans clustering partitions candidate proposals into $M = 120$ spatial subclusters defined over the entire transistor sizing and voltage bias parameter space. This enforces geometric dispersion across the physical design manifold.
3. **Environmental Stratification:** For each sizing subcluster, candidates are further grouped into $N = 10$ environmental subclusters spanning operating conditions such as supply voltage, load capacitance, common-mode voltage, and temperature.
4. **Subcluster Tournament:** Within each resulting sub-cell, candidates compete in a tournament to maximize the acquisition function $\alpha(\mathbf{x})$ under the assigned strategic profile. The single highest-scoring candidate from each sub-cell is selected for execution.

Selecting the winner from each partition yields a final evaluation batch of $10 \times M \times N = 12{,}000$ designs per topology. This hierarchy mathematically guarantees uniform representation across all 10 technology nodes and dense dispersion across 120 distinct sizing regions, while simultaneously covering 10 distinct environmental operating corners per sizing cluster.

C.3.3 SAMPLING METHODOLOGY JUSTIFICATION

Training neural surrogates across the OTA design space requires balancing global space-filling coverage with high-performance yield. We systematically evaluate four sampling methodologies: pure **Uniform Random** detailed in Appendix C.2.1, pure **RQMC Sobol** in Appendix C.2.2, **RQMC Sobol TuRBO-M-Auto** in Appendix C.3.1, and **RQMC Sobol XGBoost Surrogate** in Appendix C.3.2. While unguided methods suffer from low valid yield and Bayesian optimization induces severe spatial clustering, we ultimately select the **RQMC Sobol XGBoost Surrogate** approach for its unique ability to maximize high-performance yield without sacrificing global space-filling diversity.

Table 14 and Figure 13 evaluate parameter coverage and performance yield across the four sampling strategies on `OTA_SE07`. Pure **RQMC Sobol** increases mean nearest-neighbor distance from 0.4599 to 0.4606 over **Uniform Random** sampling, improving spatial dispersion. However, both unguided methods show low efficiency for high-performance circuits. Because operating constraints restrict viable amplifiers to thin physical sub-manifolds, random search rarely encounters high-gain or high-bandwidth points by chance. In the UMAP projections of Figure 13, **Uniform Random** and **RQMC Sobol** distributions consist predominantly of designs in the bottom quartile of Figure of Merit, rendered in purple and blue. Conversely, **TuRBO-M-Auto** targets high-performing designs, shown by yellow clusters in Figure 13, but induces severe spatial distortion. Its mean nearest-neighbor distance drops to 0.3890, coverage entropy falls to 6.7613, and normalized entropy declines to 0.5824, causing the UMAP projection to fracture into disconnected filaments. The **XGBoost Surrogate** approach eliminates this penalty. With $44{,}768$ evaluation samples versus $49{,}152$ for unguided baselines, the surrogate achieves 291 occupied bins, a peak coverage entropy of 6.9373, and a normalized entropy of 0.5975. It maintains a mean nearest-neighbor distance of 0.4591, matching the global dispersion of random search while enriching the dataset with high-performance configurations.

These empirical findings define our rationale for selecting the **RQMC Sobol XGBoost Surrogate** approach as the primary data collection pipeline for ArrivalFM-OTA. Training robust neural surrogates requires simultaneously covering the global parameter space for interpolation and sampling high-performance regimes to capture optimal designs. Unguided uniform and Sobol sampling satisfy global coverage but yield predominantly low-performing circuits. Bayesian optimization with TuRBO-M-Auto samples high-performance points but collapses into localized trajectories, leaving wide parameter voids that cripple generalization. The **RQMC Sobol XGBoost Surrogate** approach, however, populates Pareto regimes without sacrificing coverage entropy or manifold continuity, establishing the surrogate screening pipeline as the optimal methodology for large-scale circuit benchmark curation.

Table 14: Quantitative comparison of space-filling and diversity metrics on `OTA_SE07`, an NMOS input recycling folded cascode amplifier evaluated across 4 evaluated sampling methodologies. **Uniform Random** sampling was conducted in 3 batches of $16{,}384$ with (`seed500`, `seed600`, `seed700`) initializations. **RQMC Sobol** sampling was conducted in 3 batches of $16{,}384$ with (`seed100`, `seed200`, `seed400`) initializations. **RQMC Sobol XGBoost Surrogate** and **RQMC Sobol TuRBO-M-Auto** both sampled a batch of $12{,}000$ at initializations of `seed300` and `seed800` respectively with $32{,}768$ Sobol samples generated with (`seed100`, `seed200`) initializations. A schematic representation of OTA_SE07 is detailed in Figure 5.

Sampling Method	Total Samples	Mean NN Distance	Occupied Bins	Coverage Entropy	Normalized Entropy
Uniform Random	49,152	0.4599	289	6.9202	0.5961
RQMC Sobol	49,152	0.4606	264	6.8497	0.5900
RQMC Sobol XGBoost Surrogate	32,768 + 12,000	0.4591	**291**	**6.9373**	**0.5975**
RQMC Sobol TuRBO-M-Auto	32,768 + 12,000	0.3890	282	6.7613	0.5824

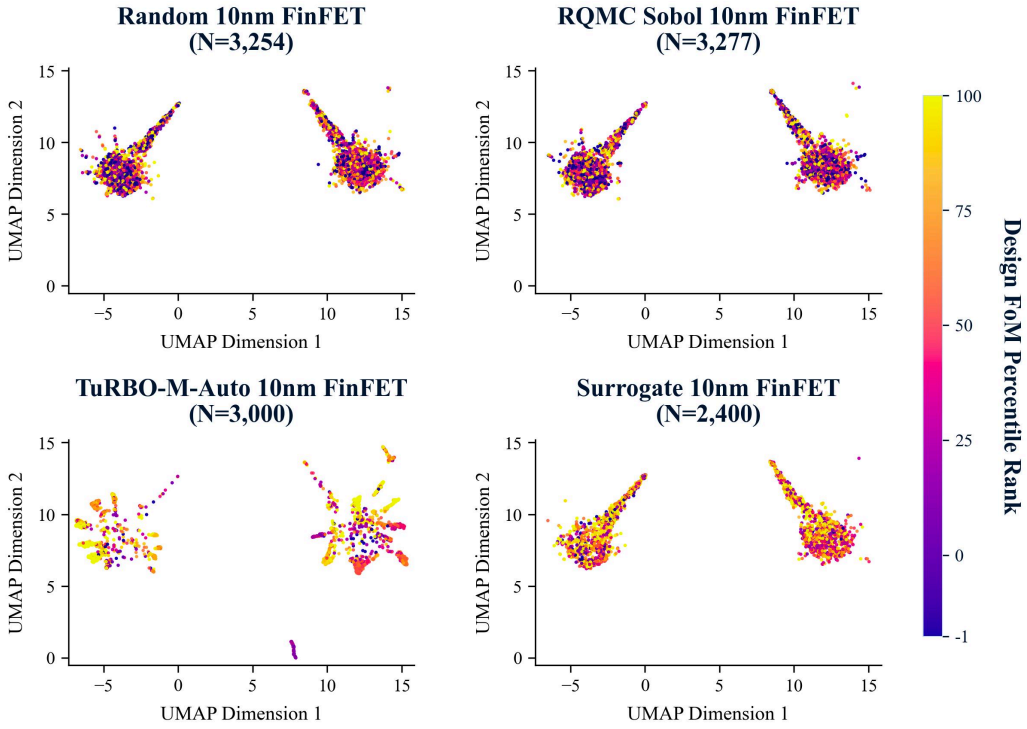

Figure 13: UMAP projection plots of OTA designs clustered by input parameter space from the experiment detailed in Table 14, showcasing 10nm technology node subsets of `seed500`, `seed100`, `seed800`, and `seed300` sampling runs. Design points are colored by their Figure of Merit (FoM) percentile, where FoM is calculated as the simulated open-loop gain A_{ol} multiplied by the unity-gain bandwidth f_{ugbw} divided by power consumption P. Design points without a valid f_{ugbw}, A_{ol}, or P are assigned a score of -1.

D CADENCE TESTBENCH AND SIMULATION SETUP

D.1 TESTBENCH CONFIGURATIONS

To ensure methodological consistency across different OTA topologies and designs, we utilize universal single-ended and fully differential testbenches in a unity-gain configuration. Reference schematics are depicted in Figures 14 and 15.

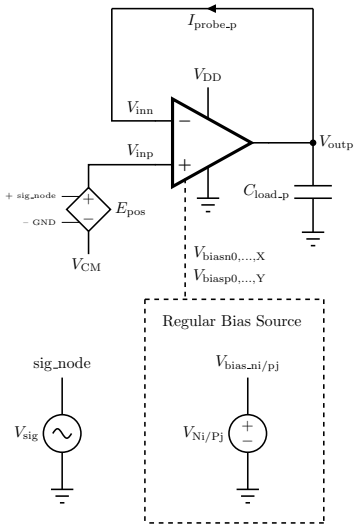

Figure 14: Universal testbench for single-ended OTAs, utilizing ideal bias sources.

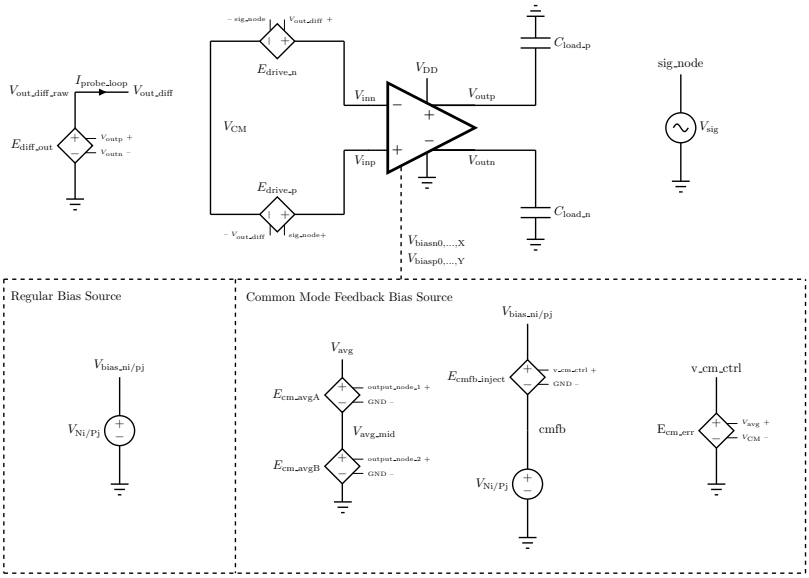

Figure 15: Universal testbench for fully differential OTAs, utilizing ideal bias sources and ideal common-mode feedback (CMFB) loops with voltage-controlled voltage sources at stage outputs. For architectures where common-mode regulation modulates an auxiliary or common-mode control terminal rather than explicit tail bias gates, such as `OTA_FD04`, $V_{\text{bias_ni/pj}}$ maps directly to the designated common-mode control node.

D.2 CADENCE SPECTRE SIMULATOR SETTINGS

All Cadence Spectre OTA simulations are conducted with precision heavy settings. Table 15 defines the global solver configurations applied to all topologies, establishing strict convergence criteria and error tolerances required to maintain simulation accuracy across the high-dimensional sampling space. Performance metrics are truncated at numerical noise floors to filter artifacts introduced by floating point precision limitations, as detailed in Appendix E.

Table 15: Standardized Cadence Spectre solver configurations utilized for dataset generation.

Parameter	Value	Purpose
`reltol`	10^{-8}	Relative proportional error
`vabstol`	10^{-12}	Absolute voltage noise floor
`iabstol`	10^{-16}	Absolute current noise floor
`tnom`	27 °C	Baseline calibration temperature
`scale`	1	Component instance scaling factor
`scalem`	1	Component model scaling factor
`cmin`	10^{-16}	Minimum parasitic capacitance floor
`gmin`	10^{-16}	Minimum parasitic conductance floor
`homotopy`	`gmin`	Path based DC convergence for initial state
`method`	`trapgear2`	Hybrid transient integration preserving accuracy without artificial dampening

D.3 DC OPERATING POINT ANALYSIS (DCOP)

```
dcOp dc
```
(23)

Performed to establish steady-state conditions and extract small-signal metrics for all active circuit components. Extracted performance metrics from `dcOp` analysis are detailed in Table 16.

D.4 STABILITY ANALYSIS (STB)

```
stb_sim stb probe=I_probe start=1u stop=1T dec=100
```
(24)

Characterizes the equivalent open-loop AC response through a current probe in a feedback loop. The frequency sweep is performed with a resolution of 100 points per decade across the [1 μHz, 1 THz] frequency range.

D.5 TRANSFER FUNCTION ANALYSIS (XF)

```
xf_sim (Vout 0) xf values=[1u]
```
(25)

Calculates the small-signal transfer function from various voltage sources like V_{DD} and V_{CM} to an output reference at 1 μHz. Used to calculate common-mode and power supply leakage.

D.6 NOISE ANALYSIS (NOISE)

```
noise_sim (Vout 0) noise start=1 stop=(10 * ugbw)
          dec=100 iprobe=Vsig
```
(26)

Calculates spectral density graphs $S(f)$ for all components in a circuit, providing total output noise in $\mathrm{V}/\sqrt{\mathrm{Hz}}$ and individual transistor noise contributions in V^2/Hz. The frequency sweep utilizes a resolution of 100 points per decade from 1Hz to 10 times the measured unity-gain bandwidth f_{ugbw}. `noise` analysis is only run on designs with an existing $f_{\mathrm{ugbw}} \geq 1000$ Hz.

Table 16: Operating point metrics extracted during `dcOp` analysis.

Parameter	**Symbol**	**Units**	**Frequency**
Transconductance	g_{m}	S	Per Transistor
Output Conductance	g_{ds}	S	Per Transistor
Threshold Voltage	V_{th}	V	Per Transistor
Saturation Voltage	V_{dsat}	V	Per Transistor
Gate-Source Voltage	V_{gs}	V	Per Transistor
Drain-Source Voltage	V_{ds}	V	Per Transistor
Common-Mode Control Voltage	$V_{\mathrm{cm_ctrl}}$	V	Per CMFB Loop in Fully Differential OTAs; See Figure 15
Drain Current	I_{ds}	A	Per Transistor
Total Gate Capacitance	C_{gg}	F	Per Transistor
Gate-Source Capacitance	C_{gs}	F	Per Transistor
Gate-Drain Capacitance	C_{gd}	F	Per Transistor
Total Drain Capacitance	C_{dd}	F	Per Transistor
Source-Bulk Capacitance	C_{ss}	F	Per Transistor
Operating Region	Reg	See Table 17	Per Transistor

Table 17: Transistor operating region classification.

Index	**Classification**	**Physical Condition**
0	Cutoff	$V_{\mathrm{gs}} < V_{\mathrm{th}}$
1	Triode	$V_{\mathrm{gs}} > V_{\mathrm{th}}, V_{\mathrm{ds}} < V_{\mathrm{dsat}}$
2	Saturation	$V_{\mathrm{gs}} > V_{\mathrm{th}}, V_{\mathrm{ds}} \geq V_{\mathrm{dsat}}$
3	Subthreshold	$V_{\mathrm{gs}} \approx V_{\mathrm{th}}$
4	Breakdown	$V_{\mathrm{ds}} \gg V_{\mathrm{DD}}$

D.7 DC SWEEP (`DC`)

```
swing_sweep dc param=dc_offset start=𝒟[0] stop=𝒟[1] step=100u
```
(27)

Evaluates the DC transfer characteristic $V_{\mathrm{out}}(V_{\mathrm{in}})$ across the topology-dependent domain $\mathcal{D}$ with a $100\mu\mathrm{m}$ step size to capture rail-to-rail saturation behavior:

$$\mathcal{D} = \begin{cases} [-V_{\mathrm{CM}}, V_{\mathrm{DD}} - V_{\mathrm{CM}}] & \text{(Single-Ended OTA)} \\ [-V_{\mathrm{DD}}, V_{\mathrm{DD}}] & \text{(Fully Differential OTA)} \end{cases} \tag{28}$$

PCHIP interpolation post-processes resulting data to ensure smooth transitions.

D.8 Periodic Steady-State Analysis (pss)

```
thd_pss pss fund=(0.1 * ugbw) harms=10 tstab=(100 / ugbw)
          steadyratio=10 maxperiods=100
```
(29)

Computes the periodic steady-state response of a circuit. To ensure reliable harmonic extraction across all OTA speeds, the input source is configured as a sinusoidal signal with an amplitude set to 10% of the supply voltage V_{DD}, while the fundamental frequency is dynamically scaled to 10% of the measured unity-gain bandwidth f_{ugbw}.

A bandwidth-dependent stabilization time `tstab` is employed to ensure sufficient transient settling before steady-state evaluation. To accurately compute nonlinear distortion products, 10 harmonics are resolved with solver convergence bounded by a steady ratio of 10 and a maximum of 100 periods.

D.9 Transient Analysis (tran)

```
tran tran stop=(1.2 * stop_time)
```
(30)

Four transient simulations are conducted that analyze large-signal and small-signal pulse response across rising and falling edges, defined by a `stop_time` and initial voltage `step_time` of:

$$\texttt{stop_time} = \frac{1}{2\pi \cdot f_{\mathrm{ugbw}}} \tag{31}$$

$$\texttt{step_time} = \max(0.001 \cdot \texttt{stop_time}, 1\ \mathrm{ps}) \tag{32}$$

Both `stop_time` and `step_time` are dependent upon the measured unity-gain bandwidth f_{ugbw} to ensure that all designs are measured given their ability to settle within 15 time constants. The voltage step is dictated by the topology dependent input pulse domain $\mathcal{S}$:

$$\mathcal{S} = \begin{cases} [0.01V_{\mathrm{DD}} - V_{\mathrm{CM}}, 0.99V_{\mathrm{DD}} - V_{\mathrm{CM}}] & \text{(Single-Ended Large-Signal Rise)} \\ [0.99V_{\mathrm{DD}} - V_{\mathrm{CM}}, 0.01V_{\mathrm{DD}} - V_{\mathrm{CM}}] & \text{(Single-Ended Large-Signal Fall)} \\ [-0.49V_{\mathrm{DD}}, 0.49V_{\mathrm{DD}}] & \text{(Differential Large-Signal Rise)} \\ [0.49V_{\mathrm{DD}}, -0.49V_{\mathrm{DD}}] & \text{(Differential Large-Signal Fall)} \\ [-0.05V_{\mathrm{DD}}, 0.05V_{\mathrm{DD}}] & \text{(Universal Small-Signal Rise)} \\ [0.05V_{\mathrm{DD}}, -0.05V_{\mathrm{DD}}] & \text{(Universal Small-Signal Fall)} \end{cases} \tag{33}$$

E Data Generation and Extraction Infrastructure

E.1 System Architecture and Pipeline Overview

To manage and abstract the complexities of mass-scale analog evaluation, we utilize a custom-built Cadence Spectre simulation and data generation framework called **Contact** to extract performance metrics across massive amounts of OTA simulations.

To support algorithmic benchmarking and dataset generation, Contact operates across six specialized execution modes. **Random**, **Sobol**, **TuRBO-M-Auto**, and **Surrogate** modes drive foundational design space sampling across uniform, quasi-random, bayesian optimized, and model recommended parameter sequences according to the sampling bounds defined in Appendix C.1. Further details regarding these sampling methodologies are detailed in Appendices C.2 and C.3. **Injection** mode is used to debug specific designs, bypassing the active sampler to directly ingest a reduced Parquet dataframe of targeted input parameters. Since large-scale distributed simulations are susceptible to transient network drops or cluster timeouts, **Repair** mode acts as the pipeline's fail-safe, simulating dropped designs to guarantee a complete dataset. The deterministic logging system used to flag these specific infrastructure failures is detailed in Appendix E.2.

Once an execution mode is selected, Contact bridges native Cadence Spectre execution with seamless performance metric extraction. This architecture scales across hundreds of CPU cores while ensuring zero-loss data translation, as depicted in Figure 16.

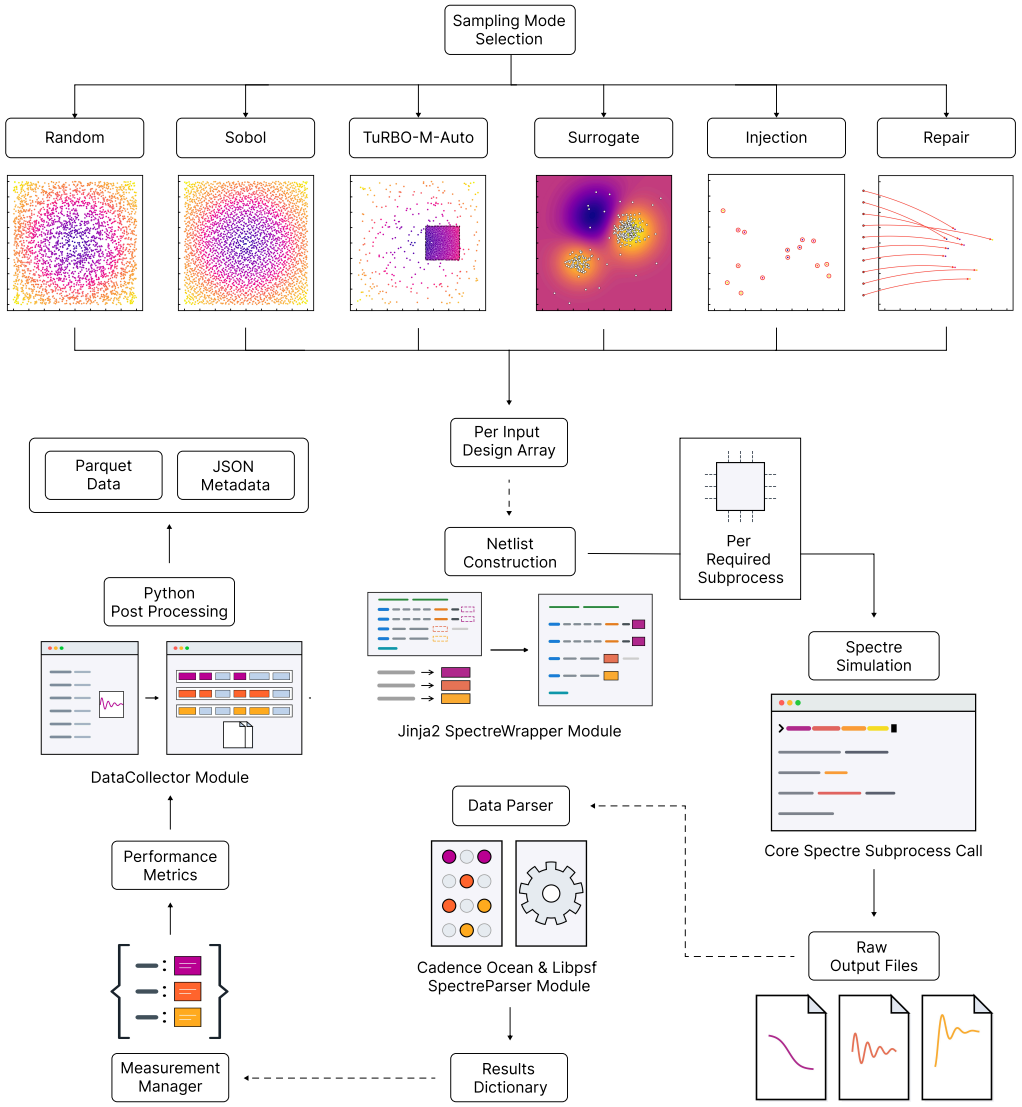

Figure 16: The Contact framework, utilizing a strictly decoupled software stack to bridge distributed simulation with data sanitization per required simulation process as detailed in Appendix E.2. Input design arrays are dynamically mapped into Cadence netlists via the Python `jinja2` templating package, ensuring flexible parameter insertion across varying topologies. Physical evaluation is then triggered via asynchronous Spectre subprocess calls. Upon simulation completion, a custom parsing module built on Cadence Ocean and the Python `libpsf` extraction package scrapes the raw output database, writing results to a Python dictionary. To manage high-concurrency execution without resource bottlenecking, a 12-token semaphore queue regulates the compute load, with tokens dynamically shared across the Spectre subprocesses, the Cadence Ocean parser, and the `libpsf` binary parser. The output results dictionary is subsequently ingested by a measurement manager, which calls specialized evaluation functions to ensure only sanitized, performance metrics advance into the final Parquet dataset.

E.2 Simulation Status and Error Logging

Evaluating a single candidate topology is not a monolithic execution block. Because OTAs require multidimensional performance characterization, a single candidate design triggers a cascading evaluation sequence comprising 2 to 10 distinct Cadence simulations chained across 1 to 9 Spectre subprocess calls per design, with some dependent upon the success of previous ones or the value of the measured unity-gain bandwidth f_{ugbw}.

This decentralized execution protocol ensures that a nonconvergence in a highly nonlinear large-signal transient simulation does not stall the execution queue or corrupt the DC biasing extraction. The nine distinct simulation stages are detailed in Table 18.

Table 18: The nine independent simulation subprocesses executed per OTA candidate design.

Subprocess	**Simulations**	**Conditions For Execution**
`gatekeeper`	`dcOp`, `stb`	Guaranteed
`full_char_1`	`xf`	`gatekeeper` stage success
`full_char_2`	`noise`	$f_{\text{ugbw}} > 1000$ Hz and `gatekeeper` stage success
`full_char_3`	`dc`	`gatekeeper` stage success
`full_char_4`	`pss`	$f_{\text{ugbw}} > 1000$ Hz and `gatekeeper` stage success
`full_char_5`	`tran` Small-Signal Rise	`gatekeeper` stage success
`full_char_6`	`tran` Small-Signal Fall	`gatekeeper` stage success
`full_char_7`	`tran` Large-Signal Rise	`gatekeeper` stage success
`full_char_8`	`tran` Large-Signal Fall	`gatekeeper` stage success

To prevent the allocation of compute resources to non-viable designs, the Contact execution pipeline enforces a strictly evaluated `gatekeeper` stage as outlined in Table 18. A design only successfully passes the gatekeeper stage if two distinct conditions are satisfied:

- All DC operating point metrics from `dcOp` analysis must be successfully extracted for every active transistor in the topology.
- `stb` analysis must be sucessful and yield a valid f_{ugbw} and a phase margin $\text{PM} > 0°$.

If a candidate fails to satisfy these baseline criteria, the evaluation sequence is immediately halted, and the subsequent eight `full_char` subprocesses are bypassed entirely to preserve overall pipeline throughput.

For characterization stages `full_char_1` through `full_char_8`, We utilize a set of diagnostic codes to precisely track the execution state of each subprocess, as detailed in Table 19 below:

Table 19: Diagnostic codes for tracking `full_char` subprocess execution stages.

Diagnostic Code	**Indicator**
3	Nominal success; simulation completed with all targeted metrics extracted
2	Stage was intentionally skipped either due to a `gatekeeper` failure or unsatisfied prerequisite condition
1	Timeout constraint violation strictly enforced at 120 s for Spectre execution and 60 s for binary parsing
0	Native Cadence Spectre convergence error
-1	Catch-all for transient infrastructure failures, including network drops, Python subprocess crashes, or Cadence license checkout limits

To aggregate the Spectre subprocess telemetry into a single, actionable dataset filtering metric, each candidate design is assigned an overarching `sim_status` code, as detailed in Table 20:

Table 20: `sim_status` codes for tracking overall design validity.

Status Code	Indicator
3	Design passes the `gatekeeper` stage with all subsequent `full_char` subprocesses returning a `diag_code` of 3, with exceptions granted for `full_char_2` and `full_char_4` which may hold a `diag_code` of 2 if bypassed via the $f_{\text{ugbw}} \leq 1000$ Hz conditional constraint
2	Design passes the `gatekeeper` stage, but one or more `full_char` subprocesses suffer a `diag_code` of 0 or 1
1	Design fails the `gatekeeper` baseline where `dcOp` metrics were successfully extracted for all transistors, but `stb` analysis fails to yield a valid f_{ugbw} or $\text{PM} > 0°$
0	Design fails the `gatekeeper` baseline, aborting entirely due to an unresolvable `dcOp` extraction
-1	Catch-all for transient infrastructure failures, including network drops, Python subprocess crashes, or Cadence license checkout limits

E.3 Performance Metric Extraction Methodologies

Following Cadence simulations and the subsequent translation of raw binaries with dedicated data parsers, We algorithmically evaluate exactly 46 global performance metrics and 5 local metrics per transistor to construct a comprehensive, continuous performance vector for each candidate topology.

The 46 global metrics capture the macro-level systemic behavior of the amplifier, while the 5 local metrics computed in addition to the baseline per-transistor metrics extracted from `dcOp` analysis track noise and stochastic behaviour.

The following subsections detail the specific mathematical formulations and measurement protocols utilized to calculate these target metrics.

E.3.1 Power Consumption (P)

Static power dissipation P is derived from `dcOp` analysis. Under nominal DC bias conditions with zero differential input excitation, the total quiescent supply current I_Q drawn from the positive voltage rail is extracted directly from the circuit simulator. The resulting power consumption in Watts is calculated as:

$$P = V_{\text{DD}} \cdot I_Q \tag{34}$$

where V_{DD} is the specified supply rail voltage for the design, and $I_Q = |I(V_{\text{DD}})|$ denotes the net quiescent current exiting the primary supply terminal.

E.3.2 DC Open Loop Gain (A_{ol})

Because all our OTA simulations are conducted in unity-gain, the magnitude of the return ratio output $|T(1\ \mu\text{Hz})|$ from `stb` analysis serves as a direct proxy for the DC open loop gain, and is bounded to a minimum floor of 10^{-10} V/V prior to decibel conversion. This is done to prevent logarithmic singularities and safely mask sub-tolerance numerical artifacts.

$$A_{\text{ol}} = 20 \log_{10} \left(\max \left(|T(1\ \mu\text{Hz})|, 10^{-10} \right) \right) \tag{35}$$

E.3.3 Unity-Gain Bandwidth (f_{ugbw})

Defines the frequency where the return ratio magnitude from `stb` analysis equals 0 dB. To ensure grid-independent accuracy, we identify the frequency interval $[f_1, f_2]$ where $|T(f_1)| \geq 1 > |T(f_2)|$. The crossover point is then resolved via log-log linear interpolation:

$$\log_{10}(f_{\text{ugbw}}) = \log_{10}(f_1) - \log_{10}(|T(f_1)|) \cdot \frac{\log_{10}(f_2) - \log_{10}(f_1)}{\log_{10}(|T(f_2)|) - \log_{10}(|T(f_1)|)} \tag{36}$$

The extraction returns a null state if the DC open loop gain $A_{\text{ol}} < 0$ dB, indicating the system never achieves unity-gain.

E.3.4 PHASE MARGIN (PM)

Quantifies circuit stability in degrees at the unity-gain bandwidth f_{ugbw}. The unwrapped phase response $\angle T(f)$ is first computed from `stb` analysis. The branch offset Θ_{branch} is then determined by rounding $\angle T(1\ \mu\mathrm{Hz})$ to the nearest 180° multiple:

$$\Theta_{\mathrm{branch}} = 180^\circ \cdot \left\lfloor \frac{\angle T(1\ \mu\mathrm{Hz})}{180^\circ} \right\rceil \tag{37}$$

The phase $\angle T(f_{\mathrm{ugbw}})$ is then resolved with semi-log linear interpolation within the interval $[f_{\mathrm{a}}, f_{\mathrm{b}}]$ containing f_{ugbw}:

$$\angle T(f_{\mathrm{ugbw}}) = \angle T(f_{\mathrm{a}}) + \frac{\angle T(f_{\mathrm{b}}) - \angle T(f_{\mathrm{a}})}{\log_{10}(f_{\mathrm{b}}) - \log_{10}(f_{\mathrm{a}})} \cdot (\log_{10}(f_{\mathrm{ugbw}}) - \log_{10}(f_{\mathrm{a}})) \tag{38}$$

The phase margin is then computed relative to Θ_{branch} and normalized to $[-180^\circ, 180^\circ]$:

$$\mathrm{PM} = \Big((\angle T(f_{\mathrm{ugbw}}) - \Theta_{\mathrm{branch}} + 180^\circ) \bmod 360^\circ \Big) - 180^\circ \tag{39}$$

The extraction returns a null state if f_{ugbw} cannot be resolved or resides outside the spectral sweep.

E.3.5 PHASE CROSSOVER FREQUENCY (f_{pc})

The frequency where the unwrapped return ratio phase $\angle T(f)$ derived from `stb` analysis crosses the -180° instability threshold. Because simulation frequency data is discrete, we algorithmically isolate the crossover point by identifying the frequency interval $[f_1, f_2]$ satisfying:

$$\mathrm{sgn}(\angle T(f_1) + 180^\circ) \neq \mathrm{sgn}(\angle T(f_2) + 180^\circ) \tag{40}$$

Given the phase response within this interval, f_{pc} is resolved via semi-logarithmic interpolation:

$$\log_{10}(f_{\mathrm{pc}}) = \log_{10}(f_1) + \frac{-180^\circ - \angle T(f_1)}{\angle T(f_2) - \angle T(f_1)} \cdot \big(\log_{10}(f_2) - \log_{10}(f_1) \big) \tag{41}$$

The extraction returns a null state if the phase $\angle T(f)$ never crosses the -180° threshold within the simulation sweep.

E.3.6 GAIN MARGIN (GM)

The fundamental safety buffer for closed-loop stability, quantifying the amount of additional loop gain in dB that the system can tolerate at the phase crossover frequency f_{pc}. It is derived directly from the return ratio $T(f)$ from `stb` analysis and bounded to a minimum floor of 10^{-10} V/V prior to decibel conversion. This is done to prevent asymptotic evaluation errors and truncate numerical simulator noise.

$$\mathrm{GM} = -20 \log_{10} \left(\max \left(|T(f_{\mathrm{pc}})|, 10^{-10} \right) \right) \tag{42}$$

To achieve sub-grid accuracy, we interpolate magnitude and phase within the frequency interval $[f_1, f_2]$ where $\angle T(f_1) > -180^\circ \geq \angle T(f_2)$. Given the log-log slope γ of the loop gain magnitude:

$$\gamma = \frac{\log_{10}(|T(f_2)|) - \log_{10}(|T(f_1)|)}{\log_{10}(f_2) - \log_{10}(f_1)} \tag{43}$$

The magnitude at the interpolated crossover f_{pc} is resolved as:

$$\log_{10}(|T(f_{\mathrm{pc}})|) = \log_{10}(|T(f_1)|) + \gamma \cdot \log_{10} \left(\frac{f_{\mathrm{pc}}}{f_1} \right) \tag{44}$$

The extraction returns a null state if f_{pc} does not exist for the design.

E.3.7 COMMON-MODE REJECTION RATIO (CMRR)

Derived in dB from the common-mode leakage X_{cm} obtained from `xf` analysis at 1 μHz, bounded to a minimum floor of 10^{-10} V/V. This is done to suppress solver artifacts and prevent divide-by-zero singularities when evaluating topologies with ideal common-mode isolation. For single-ended OTAs where $V_{\mathrm{inn}} = V_{\mathrm{out}}$ and $V_{\mathrm{inp}} = V_{\mathrm{CM}}$, output voltage is the superposition of differential and common-mode responses with the DC open loop gain A_{ol}:

$$V_{\mathrm{out}} = A_{\mathrm{ol}}(V_{\mathrm{CM}} - V_{\mathrm{out}}) + A_{\mathrm{cm}} \left(\frac{V_{\mathrm{CM}} + V_{\mathrm{out}}}{2} \right) \tag{45}$$

Algebraically solving for the common-mode leakage gain A_{cm} yields:

$$A_{\mathrm{cm}} = 2 \cdot \frac{\max(X_{\mathrm{cm}}, 10^{-10})(1 + A_{\mathrm{ol}}) - A_{\mathrm{ol}}}{1 + \max(X_{\mathrm{cm}}, 10^{-10})} \tag{46}$$

For fully differential OTAs, symmetric probing cancels the baseline common-mode voltage:

$$A_{\mathrm{cm}} = \max(X_{\mathrm{cm}}, 10^{-10}) \cdot (1 + A_{\mathrm{ol}}) \tag{47}$$

The final metric is then calculated as such:

$$\mathrm{CMRR} = 20 \log_{10} \left(\frac{A_{\mathrm{ol}}}{|A_{\mathrm{cm}}|} \right) \tag{48}$$

E.3.8 POWER SUPPLY REJECTION RATIO (PSRR)

Derived in dB from the closed-loop supply leakage X_{ps} obtained from `xf` analysis at 1 μHz, bounded to a minimum floor of 10^{-10} V/V. This is done to suppress solver artifacts and prevent divide-by-zero singularities when evaluating topologies with ideal supply isolation. Algebraically solving for the supply leakage gain A_{ps} with the DC open loop gain A_{ol} yields:

$$A_{\mathrm{ps}} = \max(X_{\mathrm{ps}}, 10^{-10}) \cdot (1 + A_{\mathrm{ol}}) \tag{49}$$

The final metric is then calculated as:

$$\mathrm{PSRR} = 20 \log_{10} \left(\frac{A_{\mathrm{ol}}}{|A_{\mathrm{ps}}|} \right) \tag{50}$$

E.3.9 TOTAL INTEGRATED OUTPUT NOISE ($V_{\mathrm{noise_out}}$)

Evaluated in units of $\mathrm{V_{rms}}$ by integrating the power spectral density $S(f)$ derived from `noise` analysis over the frequency interval $[1\ \mathrm{Hz}, 10 \cdot f_{\mathrm{ugbw}}]$, where f_{ugbw} is the measured unity-gain bandwidth. Since total output spectral density is expressed in units of $\mathrm{V}/\sqrt{\mathrm{Hz}}$, we first square output values to obtain $S(f)$ with units of V^2/Hz:

$$S(f) = \mathrm{Output}(f)^2 \tag{51}$$

To mitigate artifacts associated with standard trapezoidal integration on logarithmic sweeps, we model $S(f)$ as a piecewise power-law $S(f) = af^b$ within each incremental interval $[f_1, f_2]$. The local spectral slope b is:

$$b = \frac{\ln(S_2/S_1)}{\ln(f_2/f_1)} \tag{52}$$

The noise power contribution ΔP for the frequency interval is then resolved analytically to maintain stability near the flicker noise singularity $b \to -1$:

$$\Delta P = \int_{f_1}^{f_2} af^b \, df = \begin{cases} S_1 f_1 \frac{(f_2/f_1)^{b+1} - 1}{b+1} & \text{if } b \neq -1 \\ S_1 f_1 \ln(f_2/f_1) & \text{if } b = -1 \end{cases} \tag{53}$$

All segments ΔP_i are then summed together, bounded to a minimum floor of $10^{-10}\ \mathrm{V_{rms}}$:

$$V_{\mathrm{noise_out}} = \max \left(\sqrt{\sum_{f_i \in [1\ \mathrm{Hz}, 10 \cdot f_{\mathrm{ugbw}}]} \Delta P_i},\ 10^{-10} \right) \tag{54}$$

$V_{\mathrm{noise_out}}$ is not evaluated for designs with $f_{\mathrm{ugbw}} < 1000$ Hz.

E.3.10 INPUT REFERRED OFFSET VOLTAGE (V_{OS})

Given a `dc` sweep transfer characteristic $V_{\mathrm{out}}(V_{\mathrm{in}})$, we define the input-referred offset voltage as the V_{in} satisfying the root-finding condition, where The target threshold V_{target} is defined as V_{CM} for single-ended OTAs or 0 V for fully differential OTAs to nullify differential output and isolate intrinsic mismatch from common-mode bias:

$$f(V_{\mathrm{in}}) = V_{\mathrm{out}}(V_{\mathrm{in}}) - V_{\mathrm{target}} = 0 \tag{55}$$

We resolve via PCHIP interpolation within the discrete interval $[V_{\mathrm{in},i}, V_{\mathrm{in},i+1}]$ that satisfies the sign-change condition $f(V_{\mathrm{in},i}) \cdot f(V_{\mathrm{in},i+1}) < 0$. The interpolation is defined by the cubic polynomial:

$$P(V_{\mathrm{in}}) = a_i(V_{\mathrm{in}} - V_{\mathrm{in},i})^3 + b_i(V_{\mathrm{in}} - V_{\mathrm{in},i})^2 + c_i(V_{\mathrm{in}} - V_{\mathrm{in},i}) + d_i \tag{56}$$

where coefficients $\{a_i, b_i, c_i, d_i\}$ are uniquely determined by the PCHIP algorithm to ensure continuity across the sampled domain. While this cubic interpolation resolves the root with picovolt-level precision (rendering denser sweep increments computationally redundant), the physical validity of the measurement is fundamentally bounded by the simulator's relative tolerance (`reltol` $= 10^{-8}$). Thus, the result is constrained to a magnitude floor of 10^{-8} V:

$$V_{\mathrm{OS}} = \max(|V_{\mathrm{OS_raw}}|, 10^{-8}\ \mathrm{V}) \cdot \mathrm{sgn}(V_{\mathrm{OS_raw}}) \tag{57}$$

The extraction returns a null state if the transfer function lacks a suitable crossing.

E.3.11 SWING BOUNDARIES ($V_{\mathrm{in_min}}, V_{\mathrm{in_max}}, V_{\mathrm{swing_in}}, V_{\mathrm{out_min}}, V_{\mathrm{out_max}}, V_{\mathrm{swing_out}}$)

Input and output swing boundaries define the region of a `dc` sweep transfer characteristic in V around the input-referred offset voltage V_{OS} where the gain magnitude $A_{\mathrm{v}}(V_{\mathrm{in}}) = |dV_{\mathrm{out}}/dV_{\mathrm{in}}|$ remains within a 10% allowed deviation tolerance of the reference gain $A_{\mathrm{v}}(V_{\mathrm{OS}})$:

$$0.9 \cdot A_{\mathrm{v}}(V_{\mathrm{OS}}) \leq \left| \frac{dV_{\mathrm{out}}}{dV_{\mathrm{in}}} \right| \leq 1.1 \cdot A_{\mathrm{v}}(V_{\mathrm{OS}}) \tag{58}$$

We resolve via PCHIP interpolation as shown in (56) to determine input bounds $[V_{\mathrm{L}}, V_{\mathrm{R}}]$ within the linear sweep domain $\mathcal{D}$ from (28). The swing boundaries and ranges are then defined as:

$$\begin{gathered} \{V_{\mathrm{in_min}}, V_{\mathrm{in_max}}\} = \{V_{\mathrm{L}}, V_{\mathrm{R}}\} \\ \{V_{\mathrm{out_min}}, V_{\mathrm{out_max}}\} = \{V_{\mathrm{out}}(V_{\mathrm{L}}), V_{\mathrm{out}}(V_{\mathrm{R}})\} \\ V_{\mathrm{swing_in}} = V_{\mathrm{in_max}} - V_{\mathrm{in_min}} \\ V_{\mathrm{swing_out}} = V_{\mathrm{out_max}} - V_{\mathrm{out_min}} \end{gathered} \tag{59}$$

The extraction for all six metrics returns a null state if:

- The `dc` transfer characteristic fails to intersect V_{CM} for single-ended designs or 0 V for fully differential designs.
- The gain profile fails to maintain the 10% deviation criterion within the sweep bounds.
- $V_{\mathrm{swing_in}}$ or $V_{\mathrm{swing_out}}$ are less than a voltage floor of 10^{-4} V, invalidating designs that are severely compressed, fully saturated, or lacking a physically viable dynamic range.

E.3.12 TOTAL HARMONIC DISTORTION ($\mathrm{THD}_{\mathrm{dBc}}$)

Quantifies the power ratio of harmonic content with the spectral magnitude $|V(f)|$ relative to a fundamental frequency f_0, defined from `pss` analysis. The signal is windowed via Blackman-Harris to minimize leakage and resampled to 2^N points before computing the Fast Fourier Transform and applying amplitude correction. The harmonic power is then the root-sum-square of the first nine harmonics $n \in [2, 10]$:

$$\mathrm{THD}_{\mathrm{lin}} = \frac{\sqrt{\sum_{n=2}^{10} |V(n \cdot f_0)|^2}}{|V(f_0)|} \tag{60}$$

To prevent logarithmic singularities and mask non-physical solver harmonics, the linear metric is then bounded to a floor of 10^{-7}: a strict mathematical boundary derived from the simulator's relative tolerance (`reltol` $= 10^{-8}$) and the `pss` covergence multiplier (`steadyratio` $= 10$). The metric is then reported in dB relative to the carrier:

$$\mathrm{THD}_{\mathrm{dBc}} = 20 \log_{10} \left(\max(\mathrm{THD}_{\mathrm{lin}}, 10^{-7}) \right) \tag{61}$$

$\mathrm{THD}_{\mathrm{dBc}}$ is not evaluated for designs with a measured unity-gain bandwidth $f_{\mathrm{ugbw}} < 1000$ Hz.

E.3.13 REALIZED STEP SIZE (ΔV)

The change in V of a `tran` analysis output response, measured for defined large-signal rising, large-signal falling, small-signal rising, and small-signal falling transitions:

$$\Delta V = V_{\mathrm{final}} - V_{\mathrm{init}} \tag{62}$$

Extraction returns a null state if the settle time t_{settle} of the corresponding transient step is null, or if $|\Delta V|$ is less than two times a noise threshold of 10^{-5} V, ensuring the model does not attempt to evaluate metrics on numerical interpolation artifacts.

E.3.14 SLEW RATE (SR)

Represents the output stage's maximum large-signal voltage velocity in $\mathrm{V}/\mu\mathrm{s}$, measured for both large-signal rising (LR) and large-signal falling (LF) `tran` analysis steps to capture potential asymmetry in the output stage. This yields 2 distinct slew rate metrics in total.

We then isolate the slew interval via boundary extraction over a continuous PCHIP surrogate of the transient output $V(t)$. Given the realized transient step size ΔV we define the following 10% and 90% voltage thresholds:

$$\begin{aligned} V_{10} &= V_{\mathrm{init}} + 0.1\Delta V \\ V_{90} &= V_{\mathrm{init}} + 0.9\Delta V \end{aligned} \tag{63}$$

We then anchor the start time t_{10} to the final crossing of the 10% threshold and the end time t_{90} to the first crossing of the 90% threshold strictly following t_{10}:

$$\begin{aligned} t_{10} &= \sup \{ t \geq t_{\mathrm{step}} \mid V(t) \text{ crosses } V_{10} \} \\ t_{90} &= \inf \{ t > t_{10} \mid V(t) = V_{90} \} \end{aligned} \tag{64}$$

The slew rate is then evaluated as the derivative over this physically quarantined interval:

$$\mathrm{SR} = \frac{V_{90} - V_{10}}{t_{90} - t_{10}} \tag{65}$$

Extraction returns a null state if the realized step size ΔV or settle time t_{settle} of the corresponding transient step is also null.

E.3.15 SETTLE TIME (t_{settle})

Represents the time in ns for a `tran` analysis response to converge within a defined error tolerance band $\epsilon = \max(0.01 \cdot |\Delta V|, 10^{-5})$ of the final steady-state value V_{final}, where ΔV is the realized step size and 10^{-5} represents a noise floor for settling truncation errors. t_{settle} is measured for large-signal rising (LR), large-signal falling (LF), small-signal rising (SR), and small-signal falling (SF) steps, yielding 4 settle time metrics in total.

Search is constrained to the interval $t \in [t_{\text{split}}, t_{\text{final}}]$, where $t_{\text{final}} = 1.2 \cdot \texttt{stop_time}$ and t_{split} is anchored to the last time point crossing the midpoint threshold:

$$V_{50} = V_{\text{init}} + 0.5\Delta V \tag{66}$$

Defining $t_{\text{split}} = \sup\{t \mid V(t) = V_{50}\}$, we express the settling time relative to the delay time $t_{\text{delay}} = 0.2 \cdot \texttt{stop_time}$ by performing a reverse traversal of the PCHIP interpolated surrogate $V(t)$ over $[t_{\text{final}}, t_{\text{split}}]$:

$$t_{\text{settle}} = (\sup\{t \in [t_{\text{split}}, t_{\text{final}}] : |V(t) - V_{\text{final}}| > \epsilon\}) - t_{\text{delay}} \tag{67}$$

The extraction returns a null state if residual oscillations or drift in the final 5% of the post-delay time window exceed ϵ, the derivative of the final 5% of the post-delay time window is ever greater than $(0.01 \cdot |\Delta V|)$ / `stop_time`, or if $|\Delta V| < 10^{-5}$ V.

E.3.16 TRANSIENT DEVIATION PERCENTAGES ($\text{OS}_{\text{init}}, \text{US}_{\text{init}}, \text{OS}_{\text{act}}, \text{US}_{\text{act}}$)

Four deviation percentages characterize the stability of the `tran` analysis response between the initial voltage state V_{init} and terminal voltage state V_{final}. They are measured for large-signal rising (LR), large-signal falling (LF), small-signal rising (SR), and small-signal falling (SF) transitions, yielding $4 \times 4 = 16$ deviation metrics in total.

To decouple the initial passive injection time interval Φ_{inj} from the active recovery time interval Φ_{act}, we define a signal bisection anchored at V_{50} as defined in (66).

Passive Feedthrough: The initial offshoot OS_{init} and undershoot US_{init} percentages quantify parasitic coupling bounded strictly within Φ_{inj}, and are set to 0% if the deviation magnitude remains below a noise threshold of 10^{-5} V. This boundary is established to filter non-physical truncation errors and interpolation artifacts inherent to discrete time-stepping.

$$\begin{aligned} V_{\text{OS_init}} &= \max(0, \max(V_{\Phi_{\text{inj}}}) - V_{\text{init}}) \\ \text{OS}_{\text{init}} &= \left(\frac{V_{\text{OS_init}}}{|\Delta V|}\right) \cdot \mathbb{I}(V_{\text{OS_init}} > 10^{-5}) \times 100\% \end{aligned} \tag{68}$$

$$\begin{aligned} V_{\text{US_init}} &= \max(0, V_{\text{init}} - \min(V_{\Phi_{\text{inj}}})) \\ \text{US}_{\text{init}} &= \left(\frac{V_{\text{OS_init}}}{|\Delta V|}\right) \cdot \mathbb{I}(V_{\text{US_init}} > 10^{-5}) \times 100\% \end{aligned} \tag{69}$$

Active Ringing: The actual overshoot OS_{act} and undershoot US_{act} percentages evaluate the peak bound violations within the active loop's recovery regime Φ_{act}, where t_{cross} denotes the first intersection of the active phase with the settled value V_{final}. Values are set to 0% if the deviation magnitude remains below a noise threshold of 10^{-5} V.

$$\begin{aligned} V_{\text{OS_act}} &= \max(0, \max(V_{\Phi_{\text{act}}}) - V_{\text{final}}) \\ \text{OS}_{\text{act}} &= \left(\frac{V_{\text{OS_act}}}{|\Delta V|}\right) \cdot \mathbb{I}(V_{\text{OS_act}} > 10^{-5}) \times 100\% \end{aligned} \tag{70}$$

$$\begin{aligned} V_{\text{US_act}} &= \max(0, V_{\text{final}} - \min(V_{t > t_{\text{cross}}})) \\ \text{US}_{\text{act}} &= \left(\frac{V_{\text{US_act}}}{|\Delta V|}\right) \cdot \mathbb{I}(V_{\text{US_act}} > 10^{-5}) \times 100\% \end{aligned} \tag{71}$$

The extraction returns a null state for all deviation percentages of the corresponding transient step if the realized step size ΔV or settle time t_{settle} of the step is null.

E.3.17 TOTAL ON-CHIP AREA ESTIMATION (A_{total})

Aggregated from active transistor geometries and passive device components in μm^2, adjusted by a global layout overhead factor $\lambda_{\text{layout}} = 3.0$.

Active Device Area: For each transistor instance, the footprint is determined by the gate length L_{trans}, number of fins N_{fin}, contact poly pitch CPP, and metal pitch P_{metal}:

$$A_{\text{trans}} = (\max(L_{\text{trans}}, \text{CPP}) + \text{CPP}) \cdot (N_{\text{fin}} + 1) \cdot P_{\text{metal}} \tag{72}$$

The total active area is the sum of all M transistor instances:

$$A_{\text{active}} = \lambda_{\text{layout}} \cdot \sum_{i=1}^{M} A_{\text{trans},i} \tag{73}$$

Passive Device Area: Estimated using technology-dependent capacitance density for ρ_{cap} in F/m^2 and resistive coefficient κ_{res} in m^2/Ω:

$$A_{\text{passive}} = \sum_{j=1}^{C} \frac{C_j}{\rho_{\text{cap}}} + \sum_{k=1}^{R} (R_k \cdot \kappa_{\text{res}}) \tag{74}$$

Total Estimation: The combined area is normalized to μm^2:

$$A_{\text{total}} = (A_{\text{active}} + A_{\text{passive}}) \cdot 10^{12} \tag{75}$$

It is important to note that the predictive technology models (PTMs) utilized in this work primarily govern electrical simulation characteristics rather than physical layout constraints. Consequently, the geometric and parasitic density constants CPP, P_{metal}, ρ_{cap}, and κ_{res} are defined in Table 6 and derived from published literature characterizing equivalent foundry nodes. **Importantly, A_{total} is excluded as a metric from the forward modeling benchmark in this work.** It is strictly preserved within the ArrivalFM-OTA dataset to serve as a foundational proxy for future area-constrained design tasks.

E.3.18 PER-TRANSISTOR PELGROM MISMATCH ($\sigma_{V_{\text{th}}}$)

Defined as the statistical variance in threshold voltage between two identical, adjacent transistors, governed by the technology-specific Pelgrom coefficient $A_{V_{th}}$ in units of $\text{V} \cdot \mu\text{m}$. Because the predictive technology models (PTMs) used in this work do not intrinsically capture spatial yield variations, $A_{V_{\text{th}}}$ is derived from published literature characterizing the mismatch properties of equivalent physical foundry nodes.

For FinFET architectures, the effective channel width is derived from the fin height h_{fin} and fin thickness t_{fin}, scaled by the number of fins N_{fin}:

$$W_{\text{eff}} = N_{\text{fin}} \cdot (2 \cdot h_{\text{fin}} + t_{\text{fin}}) \tag{76}$$

The mismatch magnitude $\sigma_{V_{\text{th}}}$ for a transistor instance is then evaluated with the transistor gate length L_{trans} as follows:

$$\sigma_{V_{\text{th}}} = \frac{A_{V_{\text{th}}}}{\sqrt{L_{\text{trans}} \cdot W_{\text{eff}}}} \tag{77}$$

Representative Pelgrom and fin constants for the technology nodes utilized in this publication are cataloged in Table 6. **Furthermore, $\sigma_{V_{\text{th}}}$ is not incorporated into the forward modeling benchmark in this work.** It is instead preserved as a proxy to enable future investigations into variation-aware modeling and yield-constrained topology generation.

E.3.19 PER-TRANSISTOR TOTAL INTEGRATED NOISE ($V_{\text{noise_trans}}$)

Evaluated in units of V_{rms} by integrating the per-transistor power spectral density $S_{\text{trans}}(f)$ from `noise` analysis over the frequency interval $[1\ \text{Hz}, 10 \cdot f_{\text{ugbw}}]$, where f_{ugbw} is the measured unity-gain bandwidth.

Since $S_{\text{trans}}(f)$ is extracted in units of V^2/Hz, we directly employ piecewise power-law integration from (52), (53), and (54).

The result is then bounded to a minimum floor of $10^{-12}\ \text{V}_{\text{rms}}$. This lower bound acts as a numerical safeguard mapped directly to the simulator's absolute voltage tolerance (`vabstol` $= 10^{-12}$ V). Because the Newton-solver does not guarantee the accuracy of node voltages below this threshold, reporting $V_{\text{noise_trans}}$ beneath this limit fundamentally accumulates matrix solver artifacts rather than valid physical noise.

E.3.20 PER-TRANSISTOR FIN-STRENGTH RISK FACTOR (Ξ_{fin})

Characterizes susceptibility to stochastic process variations and discrete channel quantization against the increase in parasitic capacitance and physical footprint, quantifying the statistical dispersion of device current as a function of the total number of fins N_{fin} employed in a transistor instance:

$$\Xi_{\text{fin}} = \frac{1}{\sqrt{N_{\text{fin}}}} \tag{78}$$

It should be noted that Ξ_{fin} is not utilized in the forward modeling benchmark presented in this paper. Rather, it is provided within the dataset as foundation for future research into area and mismatch-constrained design tasks.

E.3.21 PER-TRANSISTOR STRESS SENSITIVITY FACTOR (Ψ_{stress})

Characterizes the proximity of a transistor instance to its inversion threshold. Given the gate-source voltage V_{gs} and threshold voltage V_{th} derived from `dcOp` analysis:

$$\Psi_{\text{stress}} = |V_{\text{gs}} - V_{\text{th}}| \tag{79}$$

Ψ_{stress} is not incorporated into the forward modeling benchmark in this work. It is instead preserved as a proxy metric to enable future investigations into stochastic and variation-aware circuit modeling.

E.3.22 PER-TRANSISTOR BANDWIDTH JITTER RISK FACTOR (Ω_{jitter})

Characterizes the sensitivity of a transistor's operating bandwidth to stochastic noise and supply voltage fluctuations. Defined as the product of the total integrated noise of the transistor $V_{\text{noise_trans}}$ and the device's effective time constant $\tau_{eff} \approx C_{gg}/|I_{ds}|$ derived from `dcOp` analysis:

$$\Omega_{\text{jitter}} = V_{\text{noise_trans}} \cdot \frac{C_{\text{gg}}}{|I_{\text{ds}}|} \tag{80}$$

Extraction returns a null value if I_{ds} falls below 10^{-15} A: exactly one order of magnitude above the simulator's absolute current tolerance (`iabstol` $= 10^{-16}$ A). In this regime, the device does not actively participate in dynamic signal transitions, rendering timing physically inapplicable. **Importantly, Ω_{jitter} is excluded as a metric from the forward modeling benchmark in this work.** It is preserved within the ArrivalFM-OTA dataset to serve as a foundational proxy for enabling future research into stochastic and variation-aware circuit modeling.

F Dataset Figures and Distributions

Evaluating surrogate fidelity on analog circuits requires understanding the underlying physical manifolds formed by continuous environmental sweeps and device sizing. Rather than following unconstrained, independent scalar distributions, OTA performance metrics are strictly coupled through nonlinear differential-algebraic equations that govern semiconductor physics. In this section, we analyze the 2,059,328 simulated OTA designs that comprise benchmark datapoints for **ArrivalFM-OTA**, examining the interplay between small-signal AC response, feedback stability margins, and large-signal transient dynamics across topologies and technology nodes. For exhaustive tabular distribution summaries of device-level and global performance metrics, consult Tables 21 and 22.

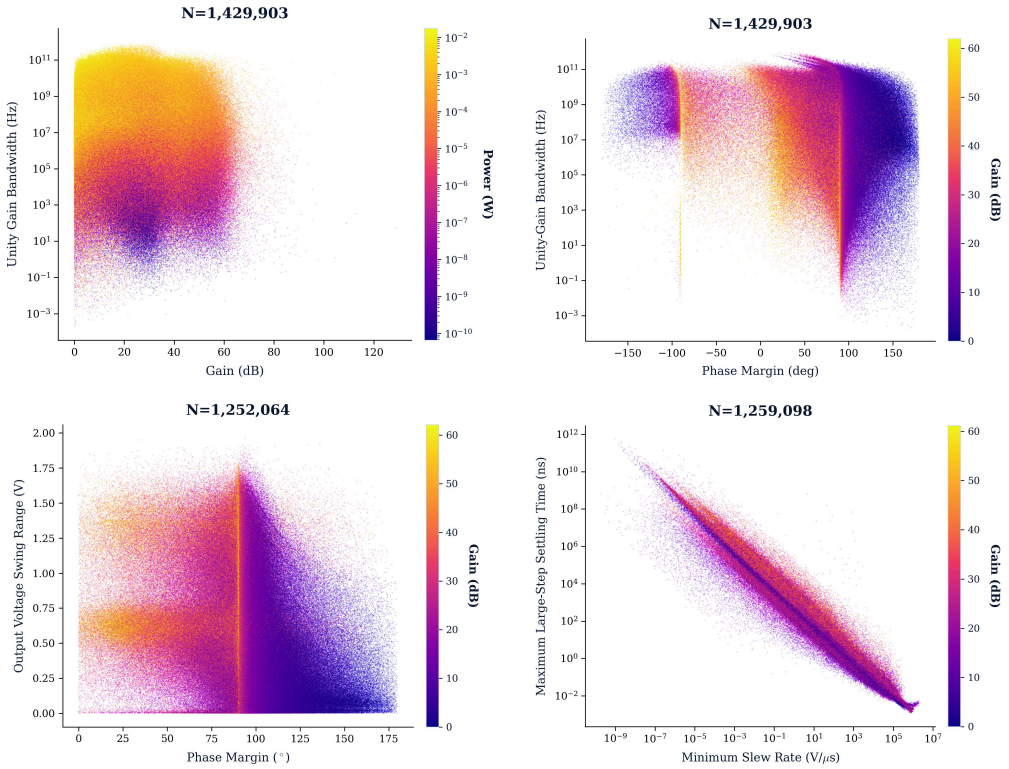

Figure 17: Global performance tradeoff manifolds across simulated OTA operating points in **ArrivalFM-OTA**. Designs are only plotted in a distribution if all 3 relevant performance metrics are non-NaN. **Top Left:** 1,429,903 OTA designs plotted by open-loop gain A_{ol} versus unity-gain bandwidth f_{ugbw} colored by static power dissipation P. **Top Right:** 1,429,903 designs plotted by phase margin PM versus unity-gain bandwidth f_{ugbw} colored by open-loop gain A_{ol}. **Bottom Left:** 1,252,064 designs plotted by output voltage swing range $V_{\text{out_range}}$ versus phase margin PM colored by open-loop gain A_{ol}. **Bottom Right:** 1,259,098 designs plotted by worst-case large-signal settling time t_{settle} versus worst-case slew rate SR colored by open-loop gain A_{ol}, revealing an invariant inverse power law spanning 14 orders of magnitude.

Figure 17 reveals the multi-dimensional structure of the dataset and confirms that our hybrid sampling covers both Pareto-optimal operating regimes and degenerate failure modes. In the top-left plot, the small-signal gain versus unity-gain bandwidth manifold exhibits the fundamental analog gain-bandwidth trade-off, gated by static power consumption. Low-power designs below $1\,\mu$W (dark purple) remain confined to sub-megahertz bandwidths due to transconductance starvation ($g_m \propto I_D$), whereas high-power configurations reaching 10 mW (yellow) push bandwidths past 10 GHz. The distribution spans from single-stage amplifiers clustered around 20–40 dB to multi-stage cascode topologies exceeding 100 dB.

The top-right and bottom-left plots map how stability margins constrain both high-frequency performance and dynamic range. A dense vertical pillar appears at $+90^\circ$ phase margin, marking the natural boundary for single-pole dominated, well-compensated designs. The region between 45° and 90° captures functional feedback amplifiers where high-gain configurations (yellow, 40–60 dB) achieve both wide bandwidth and stable closed-loop operation. Importantly, the distribution extends into negative phase margins (-180° to 0°), confirming that our sampling retains non-convergent, oscillatory, and conditionally stable configurations that models must learn to classify. In the bottom-left plot, the output swing range displays discrete horizontal bands centered near 0.7 V and 1.5 V, reflecting supply rail variations (V_{DD} sweeps) and topological differences between single-ended and fully differential stages. Symmetrical output stages achieve maximum swing (> 1.75 V) near the 90° phase margin boundary, where DC operating point biasing maintains all active devices in saturation without clipping.

The bottom-right plot demonstrates how small-signal frequency behavior directly governs large-signal transient execution across 14 decades of scale. The worst-case settling time $\max(t_{\text{settle,rise}}, t_{\text{settle,fall}})$ tracks worst-case slew rate $\min(\text{SR}_{\text{rise}}, \text{SR}_{\text{fall}})$ along a sharp inverse power-law envelope (SR $\in [10^{-9}, 10^{7}]\,\text{V}/\mu\text{s}$ against $t_{\text{settle}} \in [10^{-2}, 10^{12}]$ ns). The physical coupling across all four plots is immediately apparent in the upward dispersion above this envelope. While slew-dominated buffers hug the theoretical lower boundary ($t_{\text{settle}} \propto \Delta V/\text{SR}$), high-gain designs (yellow) scatter upward by several orders of magnitude. This latency penalty originates directly from the stability degradations seen in the top-right plot; circuits with marginal phase margins ($< 45^\circ$) or pole-zero doublets suffer from sustained small-signal ringing and slow tail settling long after initial slewing completes.



Figure 18: Cross-technology scaling and median device-level DC operating point conductance distributions across 5 technology nodes (7 nm to 20 nm) in Low-Standby Power (LSTP, purple) versus High Performance (HP, slate) flavors. **Top:** Open-loop gain (A_{ol}) density distributions per node. **Bottom Panels:** Median open-loop gain A_{ol} (dB), transistor transconductance g_m (μS), and output conductance g_{ds} (μS) across nodes, illustrating the physical origins of FinFET gain degradation and flavor divergence.

Figure 18 illustrates a fundamental inversion between digital and analog circuit scaling. In digital CMOS design, High Performance (HP) processes and advanced nodes are universally favored for their superior drive currents (I_{on}) and fast switching speeds. In analog OTA design, however, open-loop gain is strictly governed by intrinsic device gain, $A_v \approx g_m/g_{ds}$. While HP devices achieve hundreds of times higher raw transconductance than their LSTP counterparts ($597.8\,\mu$S versus $1.5\,\mu$S at 7 nm), aggressive threshold voltage reduction (V_{th}) and drain-induced barrier lowering cause output conductance g_{ds} to explode by over two orders of magnitude ($1.02\,\text{k}\mu$S versus $7.2\,\mu$S). As a result, LSTP consistently achieves superior open-loop gain over HP at every single technology node (30.4 dB versus 28.9 dB at 7 nm, widening to 14.0 dB versus 9.3 dB at 20 nm).

Furthermore, the technology sweep contradicts the classical planar heuristic that analog gain inevitably degrades with geometric scaling. As channel dimensions shrink from 20 nm down to 7 nm, median open-loop gain monotonically increases from 14.0 dB to 30.4 dB in LSTP, and from 9.3 dB to 28.9 dB in HP. While the 20 nm node employs a slightly tighter gate length sweep ($L_{\text{trans}} \in [11, 22]$ nm compared to $[11, 28]$ nm for the other nodes), this gain progression remains pronounced across 16 nm, 14 nm, 10 nm, and 7 nm where geometric bounds are identical. The lower two panels reveal the physical mechanism driving this trend: in advanced FinFET regimes, aggressive 3D multi-gate electrostatic control substantially tightens channel pinch-off and curtails channel-length modulation. This suppressive effect on g_{ds} outpaces the drop in transconductance, driving median g_{ds} down from $1.27\,\text{k}\mu$S at 20 nm to $7.2\,\mu$S at 7 nm in LSTP. Because these non-monotonic physical shifts cannot be captured by planar scaling rules or topology alone, neural surrogates must explicitly condition on internal transistor operating states (g_m, g_{ds}) to generalize across technology transitions.

Table 21: Comprehensive summary statistics of local device-level performance metrics across all simulated designs in **ArrivalFM-OTA**. For all exhaustive performance metric definitions, consult Table 16.

Metric	Count	Mean	Median	STD	IQR	Min	Max
Reg (Operating State)	24,083,538	1.955	3	1.233	2	0	3
I_{ds} (A)	24,083,538	4.91×10^{-4}	3.16×10^{-5}	1.14×10^{-3}	4.24×10^{-4}	0	0.0254
V_{ds} (V)	24,083,538	0.2649	0.2435	0.2238	0.3680	4.55×10^{-14}	1.035
V_{gs} (V)	24,083,538	0.3105	0.2889	0.1827	0.2430	2.59×10^{-15}	1.743
g_m (S)	24,083,538	4.22×10^{-3}	6.14×10^{-4}	7.37×10^{-3}	5.31×10^{-3}	0	0.0835
g_{ds} (S)	24,083,538	0.0207	1.97×10^{-4}	0.0728	2.27×10^{-3}	1.57×10^{-16}	1.625
V_{th} (V)	24,083,538	0.4160	0.3845	0.0925	0.1747	0.0541	0.5602
V_{dsat} (V)	24,083,538	0.0255	0.0238	8.85×10^{-3}	9.53×10^{-3}	7.97×10^{-3}	0.2158
C_{gg} (F)	24,083,538	6.50×10^{-15}	5.21×10^{-15}	5.12×10^{-15}	6.42×10^{-15}	1.68×10^{-17}	3.20×10^{-14}
C_{gs} (F)	24,083,538	3.59×10^{-15}	2.82×10^{-15}	2.87×10^{-15}	3.62×10^{-15}	8.42×10^{-18}	2.07×10^{-14}
C_{dd} (F)	24,083,538	2.82×10^{-15}	2.28×10^{-15}	2.25×10^{-15}	2.71×10^{-15}	8.42×10^{-18}	1.46×10^{-14}
C_{gd} (F)	24,083,538	2.91×10^{-15}	2.31×10^{-15}	2.39×10^{-15}	2.80×10^{-15}	8.42×10^{-18}	1.82×10^{-14}
C_{ss} (F)	24,083,538	3.17×10^{-15}	2.58×10^{-15}	2.46×10^{-15}	3.11×10^{-15}	8.42×10^{-18}	1.53×10^{-14}
V_{noise} (V_{rms})	13,926,043	2.04×10^{-5}	4.45×10^{-6}	8.65×10^{-5}	1.47×10^{-5}	1.00×10^{-12}	0.0656
$V_{\text{cm_ctrl}}$ (V)	1,477,264	4.07×10^{-3}	-5.87×10^{-4}	0.2675	0.3550	−0.9203	0.9294

Table 22: Comprehensive summary statistics of global performance metrics across all simulated designs in **ArrivalFM-OTA**. For all exhaustive performance metric definitions, consult Appendix E.3.

Metric	Count	Mean	Median	STD	IQR	Min	Max
A_{ol} (dB)	2,059,238	0.5120	14.075	45.388	34.870	−200.00	128.67
P (W)	2,059,238	1244.12	3.71×10^{-4}	8.40×10^{5}	2.16×10^{-3}	1.28×10^{-14}	9.76×10^{8}
f_{ugbw} (Hz)	1,429,903	3.72×10^{9}	2.52×10^{7}	1.77×10^{10}	4.54×10^{8}	2.13×10^{-4}	9.16×10^{11}
PM (°)	1,429,903	86.841	92.407	38.577	14.479	−179.78	179.97
f_{pc} (Hz)	1,899,952	9.18×10^{10}	9.52×10^{9}	1.84×10^{11}	7.65×10^{10}	1.02×10^{-6}	1.00×10^{12}
GM (dB)	1,899,952	56.303	49.904	40.251	50.671	−51.698	200.00
CMRR (dB)	1,381,479	112.84	137.66	85.751	171.12	−21.606	200.00
PSRR (dB)	1,381,479	109.96	137.49	88.161	178.08	−49.717	200.00
$V_{\mathrm{in_max}}$ (V)	1,252,064	0.3562	0.3314	0.2264	0.3599	−0.5852	1.035
$V_{\mathrm{in_min}}$ (V)	1,252,064	−0.3582	−0.3328	0.2248	0.3558	−1.035	0.6606
$V_{\mathrm{in_range}}$ (V)	1,252,064	0.7144	0.6293	0.4242	0.6046	1.00×10^{-4}	2.070
$V_{\mathrm{out_max}}$ (V)	1,252,064	0.4826	0.5277	0.2119	0.3249	-2.08×10^{-5}	1.013
$V_{\mathrm{out_min}}$ (V)	1,252,064	−0.1509	−0.0994	0.3356	0.5445	−0.9805	0.9119
$V_{\mathrm{out_range}}$ (V)	1,252,064	0.6335	0.5668	0.3915	0.5457	1.00×10^{-4}	1.961
V_{OS} (V)	1,252,270	-6.39×10^{-4}	-1.00×10^{-8}	0.0322	2.00×10^{-8}	−0.5958	0.6606
ΔV **(Large Rise,** V)	1,086,374	0.5774	0.6202	0.1821	0.2028	2.01×10^{-5}	1.432
SR **(Rise,** V/μs)	1,086,369	6527.51	34.513	4.00×10^{4}	559.03	1.96×10^{-10}	2.05×10^{6}
t_{settle} **(Large Rise,** ns)	1,086,374	1.13×10^{7}	26.099	1.03×10^{9}	719.57	1.20×10^{-3}	5.37×10^{11}
$\mathrm{OS}_{\mathrm{init}}$ **(Large Rise,** %)	1,086,374	1.336	0	158.61	0.0115	0	1.40×10^{5}
$\mathrm{US}_{\mathrm{init}}$ **(Large Rise,** %)	1,086,374	1.746	0	7.032	0.4592	0	2861.40
$\mathrm{OS}_{\mathrm{act}}$ **(Large Rise,** %)	1,086,374	2.450	0	181.56	0	0	5.94×10^{4}
$\mathrm{US}_{\mathrm{act}}$ **(Large Rise,** %)	1,086,374	0.0853	0	1.336	0	0	49.992
ΔV **(Large Fall,** V)	947,474	−0.5913	−0.6278	0.1697	0.1911	−1.432	-2.01×10^{-5}
SR **(Fall,** V/μs)	947,472	6752.04	31.042	4.24×10^{4}	479.86	1.96×10^{-10}	2.05×10^{6}
t_{settle} **(Large Fall,** ns)	947,474	1.25×10^{7}	30.891	1.09×10^{9}	882.42	1.20×10^{-3}	5.37×10^{11}
$\mathrm{OS}_{\mathrm{init}}$ **(Large Fall,** %)	947,474	0.9433	0	42.037	2.54×10^{-3}	0	3.50×10^{4}
$\mathrm{US}_{\mathrm{init}}$ **(Large Fall,** %)	947,474	1.820	0	15.447	0.4887	0	8530.10
$\mathrm{OS}_{\mathrm{act}}$ **(Large Fall,** %)	947,474	3.011	0	386.28	0	0	2.51×10^{5}
$\mathrm{US}_{\mathrm{act}}$ **(Large Fall,** %)	947,474	0.0830	0	1.242	0	0	49.945
ΔV **(Small Rise,** V)	1,244,440	0.0655	0.0683	0.0189	0.0154	2.00×10^{-5}	0.8423
t_{settle} **(Small Rise,** ns)	1,244,440	1.20×10^{7}	25.268	9.62×10^{8}	798.96	8.50×10^{-4}	4.20×10^{11}
$\mathrm{OS}_{\mathrm{init}}$ **(Small Rise,** %)	1,244,440	1.311	0	77.359	0	0	3.54×10^{4}
$\mathrm{US}_{\mathrm{init}}$ **(Small Rise,** %)	1,244,440	4.226	0	63.308	0.3401	0	3.14×10^{4}
$\mathrm{OS}_{\mathrm{act}}$ **(Small Rise,** %)	1,244,440	15.970	0	465.14	0	0	1.03×10^{5}
$\mathrm{US}_{\mathrm{act}}$ **(Small Rise,** %)	1,244,440	0.2778	0	1.796	0	0	49.948
ΔV **(Small Fall,** V)	1,248,284	−0.0656	−0.0684	0.0188	0.0153	−0.8423	-2.00×10^{-5}
t_{settle} **(Small Fall,** ns)	1,248,284	1.20×10^{7}	26.608	9.59×10^{8}	859.65	1.94×10^{-4}	4.20×10^{11}
$\mathrm{OS}_{\mathrm{init}}$ **(Small Fall,** %)	1,248,284	1.303	0	74.336	0	0	3.58×10^{4}
$\mathrm{US}_{\mathrm{init}}$ **(Small Fall,** %)	1,248,284	4.317	0	69.665	0.3186	0	4.01×10^{4}
$\mathrm{OS}_{\mathrm{act}}$ **(Small Fall,** %)	1,248,284	15.841	0	463.46	0	0	9.37×10^{4}
$\mathrm{US}_{\mathrm{act}}$ **(Small Fall,** %)	1,248,284	0.2641	0	1.740	0	0	49.960
$V_{\mathrm{noise_out}}$ ($\mathrm{V_{rms}}$)	1,315,936	8.82×10^{-5}	2.84×10^{-5}	2.86×10^{-4}	7.14×10^{-5}	1.07×10^{-6}	0.0934
$\mathrm{THD}_{\mathbf{dBc}}$ (dB)	1,272,202	−50.577	−48.138	26.400	45.930	−140.00	8.659

G BENCHMARK PROTOCOLS AND INSTRUCTIONS

The objective of the **ArrivalFM-OTA** benchmark is to evaluate machine learning models on multi-task forward physical performance prediction for OTAs in a single forward pass. Models evaluated on ArrivalFM-OTA are formulated to predict 68 performance metrics across global and component levels. The following benchmark documentation specifies dataset keys, physical definitions, and reference transformations for all classification and regression heads. Official benchmark splits are detailed in Table 2. **To ensure statistical significance and prevent initialization bias, official benchmark evaluation requires training across 10 independent random seeds from `seed42` through `seed51`, reporting mean and standard deviation across all runs.** Comprehensive data loading routines, baseline training scripts, and reproduction instructions are provided in the supplementary material and repository documentations. ArrivalFM-OTA comprises:

1. Global Classification Heads (10 Targets): Circuit level convergence and physical stability indicators. Each target is labeled as convergent (1) or non-convergent (0) as specified in Table 23.

2. Global Continuous Regression Heads (41 Targets): Global specifications spanning a wide variety of simulation types. Targets and their designated training transformations are detailed in Table 24.

3. Global Derived Regression Heads (2 Targets): Input and output voltage swing ranges are calculated directly from corresponding maximum and minimum regression heads:

$$\texttt{out_input_voltage_swing_range_v} = V_{\text{in,max}} - V_{\text{in,min}} \tag{81}$$

$$\texttt{out_output_voltage_swing_range_v} = V_{\text{out,max}} - V_{\text{out,min}} \tag{82}$$

4. Device-Level Classification Head (1 Target): Transistor operating region classification for each individual device across five discrete physical regimes as summarized in Table 17. The classification head `region_of_operation` predicts the physical operating state $\text{Reg}_m \in \{0, 1, 2, 3, 4\}$ for each transistor m.

5. Device-Level Continuous Regression Heads (14 Targets): DC operating points and noise metrics per transistor, alongside common-mode feedback control voltages, as listed in Table 25.

ArrivalFM-OTA is formulated as an open, representation-agnostic benchmark:

- **Model Architecture:** Any learning framework can be deployed, including graph neural networks, transformers, sequence models, neural ordinary differential equations, or multi-layer perceptrons.
- **Circuit Representation:** Circuits can be framed with any representation, including but not limited to our heterogeneous graph representation, homogeneous graphs, hypergraphs, tokenized netlist sequences, or tabular feature vectors.
- **Technology Node Encoding:** Technology models can be framed with any representation, including but not limited to our continuous 17-dimensional parameter vector implementation, one-hot indicators, or learned continuous embeddings.
- **Optimization and Inversion:** Training objectives, loss weighting schedules, and normalization schemes remain at the designer's discretion. Final evaluation requires mapping predictions back to physical and standardized logarithmic spaces using the benchmark inverse transforms detailed in Tables 24 and 25.

Table 23: Global classification targets and corresponding simulation validity criteria. For exhaustive definitions of the performance metrics defining validity classifiers, consult Appendix E.3.

Classifier Head	Physical Description	Convergence Condition
validity	Full simulation status validity	sim_status == 3
nominal_stab	Nominal loop stability	$f_{\text{ugbw}} \neq \text{NaN}$
threshold_stab	Phase crossover validity	$f_{\text{pc}} \neq \text{NaN}$
swing_stab	Output voltage swing validity	$V_{\text{out_range}} \neq \text{NaN}$
offset_stab	Input offset voltage validity	$V_{\text{OS}} \neq \text{NaN}$
transient_lr	Large-signal rise settling validity	$t_{\text{settle,lsr}} \neq \text{NaN}$
transient_lf	Large-signal fall settling validity	$t_{\text{settle,lsf}} \neq \text{NaN}$
transient_sr	Small-signal rise settling validity	$t_{\text{settle,ssr}} \neq \text{NaN}$
transient_sf	Small-signal fall settling validity	$t_{\text{settle,ssf}} \neq \text{NaN}$
periodic_stab	Total harmonic distortion validity	$\text{THD}_{\text{dBc}} \neq \text{NaN}$

Table 24: Global continuous regression targets and reference normalization methods. For exhaustive definitions of the performance metrics defining regression targets, consult Appendix E.3.

Regression Head	Physical Description	Units	Transform Method
out_gain_ol_dc_db	DC Open-Loop Voltage Gain	dB	Linear (Z-Score)
out_power_w	Static Power Dissipation	W	$\log_{10}$ (Z-Score)
out_ugbw_hz	Unity-Gain Bandwidth	Hz	$\log_{10}$ (Z-Score)
out_phase_margin_deg	Loop Phase Margin	deg	Linear (Z-Score)
out_phase_crossover_freq_hz	Phase Crossover Frequency	Hz	$\log_{10}$ (Z-Score)
out_gain_margin_db	Loop Gain Margin	dB	Linear (Z-Score)
out_cmrr_dc_db	Common-Mode Rejection Ratio	dB	Linear (Z-Score)
out_psrr_dc_db	Power Supply Rejection Ratio	dB	Linear (Z-Score)
out_input_voltage_swing_max_v	Maximum Input Swing Voltage	V	Linear (Z-Score)
out_input_voltage_swing_min_v	Minimum Input Swing Voltage	V	Linear (Z-Score)
out_output_voltage_swing_max_v	Maximum Output Swing Voltage	V	Linear (Z-Score)
out_output_voltage_swing_min_v	Minimum Output Swing Voltage	V	Linear (Z-Score)
out_vos_v	Input-Referred Offset Voltage	V	Symlog_{10} ($x_0 = 10^{-6}$, Z-Score)
out_slew_rate_rise_v_us	Slew Rate (Rise Large)	$\text{V}/\mu\text{s}$	$\log_{10}$ (Z-Score)
out_slew_rate_fall_v_us	Slew Rate (Fall Large)	$\text{V}/\mu\text{s}$	$\log_{10}$ (Z-Score)
out_realized_step_rise_large_sig_v	Large-Signal Step Height (Rise)	V	Linear (Z-Score)
out_realized_step_fall_large_sig_v	Large-Signal Step Height (Fall)	V	Linear (Z-Score)
out_realized_step_rise_small_sig_v	Small-Signal Step Height (Rise)	V	Linear (Z-Score)
out_realized_step_fall_small_sig_v	Small-Signal Step Height (Fall)	V	Linear (Z-Score)
out_settle_time_rise_large_sig_ns	Large-Signal Settle Time (Rise)	ns	$\log_{10}$ (Z-Score)
out_settle_time_fall_large_sig_ns	Large-Signal Settle Time (Fall)	ns	$\log_{10}$ (Z-Score)
out_settle_time_rise_small_sig_ns	Small-Signal Settle Time (Rise)	ns	$\log_{10}$ (Z-Score)
out_settle_time_fall_small_sig_ns	Small-Signal Settle Time (Fall)	ns	$\log_{10}$ (Z-Score)
out_offshoot_initial_rise_large_sig_pct	Initial Peak Overshoot (Rise Large)	%	$\log_{10}(1+x)$ (Z-Score)
out_undershoot_initial_rise_large_sig_pct	Initial Undershoot (Rise Large)	%	$\log_{10}(1+x)$ (Z-Score)
out_offshoot_actual_rise_large_sig_pct	Settled Overshoot (Rise Large)	%	$\log_{10}(1+x)$ (Z-Score)
out_undershoot_actual_rise_large_sig_pct	Settled Undershoot (Rise Large)	%	$\log_{10}(1+x)$ (Z-Score)
out_offshoot_initial_fall_large_sig_pct	Initial Peak Overshoot (Fall Large)	%	$\log_{10}(1+x)$ (Z-Score)
out_undershoot_initial_fall_large_sig_pct	Initial Undershoot (Fall Large)	%	$\log_{10}(1+x)$ (Z-Score)
out_offshoot_actual_fall_large_sig_pct	Settled Overshoot (Fall Large)	%	$\log_{10}(1+x)$ (Z-Score)
out_undershoot_actual_fall_large_sig_pct	Settled Undershoot (Fall Large)	%	$\log_{10}(1+x)$ (Z-Score)
out_offshoot_initial_rise_small_sig_pct	Initial Peak Overshoot (Rise Small)	%	$\log_{10}(1+x)$ (Z-Score)
out_undershoot_initial_rise_small_sig_pct	Initial Undershoot (Rise Small)	%	$\log_{10}(1+x)$ (Z-Score)
out_offshoot_actual_rise_small_sig_pct	Settled Overshoot (Rise Small)	%	$\log_{10}(1+x)$ (Z-Score)
out_undershoot_actual_rise_small_sig_pct	Settled Undershoot (Rise Small)	%	$\log_{10}(1+x)$ (Z-Score)
out_offshoot_initial_fall_small_sig_pct	Initial Peak Overshoot (Fall Small)	%	$\log_{10}(1+x)$ (Z-Score)
out_undershoot_initial_fall_small_sig_pct	Initial Undershoot (Fall Small)	%	$\log_{10}(1+x)$ (Z-Score)
out_offshoot_actual_fall_small_sig_pct	Settled Overshoot (Fall Small)	%	$\log_{10}(1+x)$ (Z-Score)
out_undershoot_actual_fall_small_sig_pct	Settled Undershoot (Fall Small)	%	$\log_{10}(1+x)$ (Z-Score)
out_integrated_noise_vrms	Integrated Input Noise	V_{rms}	$\log_{10}$ (Z-Score)
out_thd_db	Total Harmonic Distortion	dB	Linear (Z-Score)

Table 25: Device-level regression specifications and end transformation methods.

Regression Head	**Physical Description**	**Units**	**Transform Method**
op_ids	Drain Current	A	Symlog_{10} ($x_0 = 10^{-6}$)
op_vds	Drain-Source Voltage	V	Linear (Z-Score)
op_vgs	Gate-Source Voltage	V	Linear (Z-Score)
op_gm	Transconductance	S	Symlog_{10} ($x_0 = 10^{-6}$)
op_gds	Output Conductance	S	Symlog_{10} ($x_0 = 10^{-6}$)
op_vth	Threshold Voltage	V	Linear (Z-Score)
op_vdsat	Saturation Voltage	V	Linear (Z-Score)
op_cgg	Total Gate Capacitance	F	$\log_{10}$
op_cgs	Gate-Source Capacitance	F	$\log_{10}$
op_cdd	Total Drain Capacitance	F	$\log_{10}$
op_cgd	Gate-Drain Capacitance	F	$\log_{10}$
op_css	Total Source Capacitance	F	$\log_{10}$
op_noise_total_vrms	Total Integrated Noise	V_{rms}	$\log_{10}$
op_v_cm_ctrl	Common-Mode Control Voltage	V	Linear (Z-Score)

H Documentation of Architectural Baselines

H.1 Training and Hyperparameter Specifications

This section details the concrete circuit graph construction, tensor layouts, neural backbone implementations, and training hyperparameter protocols for all baseline architectures evaluated on the **ArrivalFM-OTA** benchmark. Figure 19 provides a structural schematic of our attributed bipartite graph formulation on a representative OTA graph schematic alongside example sub-vector layouts for component, net, pin edge, and global context tensors. Structural configurations across all 8 evaluated architectures (4 baseline backbones and 4 DAWN-modified variants) are formalized in Table 26, with universal training hyperparameters and loss utilizations documented in Tables 27 and 28.

Distributed Compute Infrastructure and Execution Protocol: All baseline models and DAWN variants are trained and evaluated on the MIT Office of Research Computing and Data (ORCD) Engage Compute Cluster. Distributed hardware acceleration is executed across $6\times$ NVIDIA GPUs using PyTorch `DistributedDataParallel`. Compute instances are allocated 16 physical CPU cores, 512 GB of host system RAM, and 8 parallel `DataLoader` workers to maximize I/O throughput. To ensure robust execution on preemptable cluster allocations, our training protocol incorporates native SLURM signal trapping for both `SIGUSR1` and `SIGTERM`. Thus, when a preemption notification is received, all complete model weights, AdamW optimizer moments, and Cosine Annealing scheduler states are serialized before cleanly requeuing. To guarantee statistical significance and eliminate initialization bias, every architecture is trained from scratch across 10 independent random seeds (`seed42` through `seed51`), with all benchmark results reported as the mean and sample standard deviation across all 10 runs.

Table 26: Architectural specifications across all 8 evaluated neural backbones. Evaluated baselines maintain strictly matched layer depths ($L = 4$) and hidden representations ($d = 512$) to isolate structural inductive bias, spanning topological ablation (MLP), standard unweighted graph convolution (GCN), relational pin self-attention (GT), and asymmetric recurrent pin projections (ParaGraph). DAWN-augmented variants incorporate a 128-dimensional nodal residual cascade that conditions circuit-level readouts on intermediate transistor DC operating points with minimal parameter overhead.

Model Architecture	Depth (L)	Width (d)	Edge Handling	Global $\mathbf{u}$	Parameters
MLP	4	512	Disabled ($\mathcal{E} = \emptyset$)	Readout Concatenation	2.16M
GCN	4	512	Unweighted	Readout Concatenation	2.17M
GT	4	512	9D Edge Injection	Readout Concatenation	2.29M
ParaGraph	$T = 4$	512	9 Relation $\mathbf{W}_r$	Readout Concatenation	2.15M
DAWN MLP	$4 + 2$	$128 \,/\, 512$	Disabled ($\mathcal{E} = \emptyset$)	Injected + Readout	2.31M
DAWN GCN	$4 + 2$	$128 \,/\, 512$	Unweighted	Injected + Readout	2.32M
DAWN GT	$4 + 2$	$128 \,/\, 512$	9D Edge Injection	Injected + Readout	2.44M
DAWN ParaGraph	$T = 4 + 2$	$128 \,/\, 512$	9 Relation $\mathbf{W}_r$	Injected + Readout	2.30M

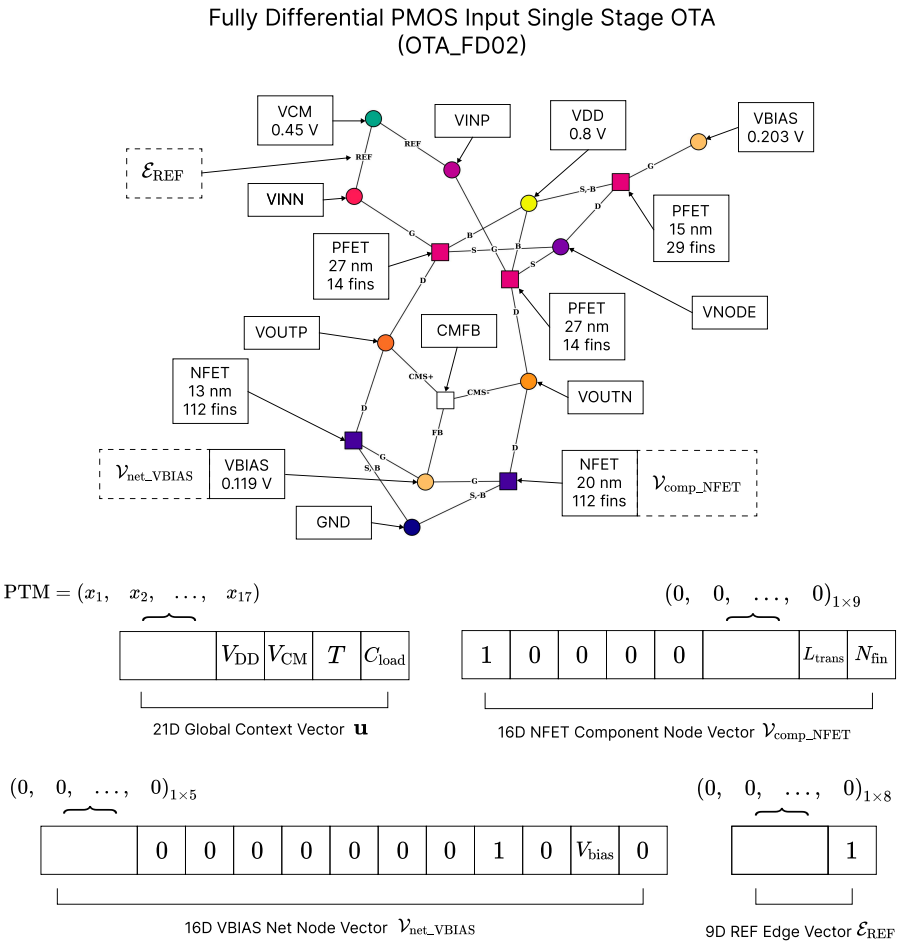

Figure 19: Attributed bipartite circuit graph formulation and tensor encoding layout on `OTA_FD02`, a fully differential single-stage amplifier. **Top:** The circuit topology is partitioned into disjoint component vertices represented by squares and electrical net vertices represented by circles. They are connected bidirectionally by directed pin edges $\mathcal{E}$ carrying physical terminal annotations. **Bottom:** Sub-vector tensor structure across the four constituent feature spaces: (1) **Global Context Vector** $\mathbf{u} \in \mathbb{R}^{21}$ concatenates 17 continuous FinFET PTM physical parameters with 4 operating conditions, namely supply voltage V_{DD}, input common-mode voltage V_{CM}, temperature T, and load capacitance C_{load}. (2) **Component Node Vector** $\mathcal{V}_{\text{comp}} \in \mathbb{R}^{16}$ combines a 5-dimensional component one-hot indicator, a 9-dimensional zero vector for net isolation, and 2 continuous sizing dimensions comprising transistor channel length L_{trans} and discrete fin count N_{fin}. (3) **Net Node Vector** $\mathcal{V}_{\text{net}} \in \mathbb{R}^{16}$ combines a 5-dimensional zero vector for device isolation, a 9-dimensional net type one-hot indicator, and 2 DC bias dimensions, exemplified by the fixed bias level V_{bias}. (4) **Edge Attribute Vector** $\mathcal{E} \in \mathbb{R}^{9}$ encodes terminal pin identity via a 9-dimensional one-hot indicator, exemplified by the reference terminal edge $\mathcal{E}_{\text{REF}}$.

Table 27: Universal training and optimization hyperparameters across all evaluated architectures on **ArrivalFM-OTA**.

Hyperparameter / System Configuration	**Value**
Optimizer	AdamW ($\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$)
Initial Learning Rate (η_0)	1.0×10^{-3}
Minimum Learning Rate ($\eta_{\min}$)	1.0×10^{-6}
Learning Rate Schedule	Cosine Annealing ($T_{\max} = 1000$, 0 warmup epochs)
Weight Decay (λ)	1.0×10^{-4}
Dropout Rate (p)	0.10
Gradient Norm Clipping	1.0 (maximum L2 norm)
Effective Batch Size	1024
Maximum Epochs	1000
Early Stopping Metric	Validation Combined Loss ($\mathcal{L}_{\text{total}}$)
Early Stopping Patience	50 epochs
Node & Edge Hidden Dimension (d)	512
Message Passing Layers (L)	4
DCOP Cascade Bottleneck Width (d_{DAWN})	128 (DAWN variants only)
Target Transformations	Per-head normalization as detailed in Tables 24 and 25
Loss Criteria	Detailed in Table 28

Table 28: Training loss criteria assigned per prediction head across our 8 evaluated neural baselines, adhering to the multi-task formulation in Section 4.3. For exhaustive definitions of mentioned performance metrics, consult Table 16 for device-level specifications and Appendix E.3 for global specifications.

Target Category	**Loss Function**	**Assigned Prediction Heads**
Simulation Validity	`BCEWithLogits`	All 10 global classification stability flags
Transistor Operating Region	`CrossEntropy`	5-class discrete state $\text{Reg}_m \in \{0, 1, 2, 3, 4\}$
Linear Continuous Specs	`MSE`	A_{ol}, PM, GM, CMRR, PSRR, V_{OS}, THD_{dBc}, $V_{\text{in_max}}$, $V_{\text{in_min}}$, $V_{\text{out_max}}$, $V_{\text{out_min}}$, all 4 transient step sizes ΔV, V_{ds}, V_{gs}, V_{th}, V_{dsat}, C_{gg}, C_{gs}, C_{dd}, C_{gd}, C_{ss}, and $V_{\text{cm_ctrl}}$
Heavy-Tailed Dynamic Specs	`Huber` ($\delta = 1.0$)	P, f_{ugbw}, f_{pc}, all 2 slew rates SR, all 4 settling times t_{settle}, all 4 initial offshoot percentages OS_{init}, all 4 actual offshoot percentages OS_{act}, all 4 initial undershoot percentages US_{init}, all 4 actual undershoot percentages US_{act}, $V_{\text{noise_out}}$, $V_{\text{noise_trans}}$, I_{ds}, g_m, and g_{ds}

H.2 Mathematical Formulations of Baseline Backbones

To ensure full reproducibility, we formalize the layer-wise update equations for all four baseline neural architectures over the attributed circuit graph $\mathcal{G} = (\mathcal{V}_{\text{comp}} \cup \mathcal{V}_{\text{net}}, \mathcal{E}, \mathbf{X}, \mathbf{E}, \mathbf{u})$. Across all backbones, initial node representations are projected via $\mathbf{h}_v^{(0)} = \text{GELU}(\mathbf{W}_v \mathbf{x}_v + \mathbf{b}_v)$ and global context via $\mathbf{h}_u = \text{GELU}(\mathbf{W}_u \mathbf{u} + \mathbf{b}_u)$, where $\mathbf{W}_v \in \mathbb{R}^{512 \times 16}$ and $\mathbf{W}_u \in \mathbb{R}^{512 \times 21}$.

H.2.1 MLP (Topological Ablation)

The MLP baseline serves as our topological ablation by disabling all graph edges ($\mathcal{E} = \emptyset$). Each node embedding $\mathbf{h}_v \in \mathbb{R}^{512}$ is updated independently across $L = 4$ feedforward blocks without spatial message passing:

$$\mathbf{h}_v^{(l+1)} = \text{Dropout}\left(\text{LayerNorm}\left(\text{GELU}\left(\mathbf{W}^{(l)}\mathbf{h}_v^{(l)} + \mathbf{b}^{(l)}\right)\right), p = 0.1\right), \quad l \in \{0, 1, 2, 3\} \tag{83}$$

where $\mathbf{W}^{(l)} \in \mathbb{R}^{512 \times 512}$. This baseline strictly isolates node-level parameter capacity from relational graph inductive biases.

H.2.2 GRAPH CONVOLUTIONAL NETWORK (GCN)

Our GCN baseline follows Kipf & Welling (2017), propagating isotropic normalized messages across neighboring device and net vertices. To preserve self-features, self-loops are added to the adjacency structure ($\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$). Layer-wise updates incorporate post-aggregation normalization:

$$\mathbf{h}_i^{(l+1)} = \text{Dropout}\left(\text{LayerNorm}\left(\text{GELU}\left(\sum_{j \in \mathcal{N}(i)\cup\{i\}} \frac{1}{\sqrt{\tilde{d}_i \tilde{d}_j}} \mathbf{W}^{(l)} \mathbf{h}_j^{(l)}\right)\right), p = 0.1\right) \tag{84}$$

where $\tilde{d}_i = 1 + |\mathcal{N}(i)|$ denotes the augmented node degree, and $\mathbf{W}^{(l)} \in \mathbb{R}^{512\times512}$. Directed pin edge attributes $\mathbf{e}_{ij}$ are omitted during propagation.

H.2.3 GRAPH TRANSFORMER (GT)

Our GT baseline deploys multi-head relational self-attention via `TransformerConv` (Shi et al., 2021). Raw 9-dimensional pin edge attributes $\mathbf{e}_{ji} \in \mathbb{R}^9$ are first projected to hidden width $d = 512$ via $\mathbf{h}_{ji}^e = \text{GELU}(\mathbf{W}_e \mathbf{e}_{ji} + \mathbf{b}_e)$. Across $K = 4$ attention heads, the attention coefficient from node j to node i incorporates the projected edge embedding:

$$\alpha_{ij}^k = \frac{\exp\left(\frac{1}{\sqrt{d}}\left(\mathbf{W}_Q^k \mathbf{h}_i^{(l)}\right)^\top \left(\mathbf{W}_K^k \mathbf{h}_j^{(l)} + \mathbf{W}_E^k \mathbf{h}_{ji}^e\right)\right)}{\sum_{u\in\mathcal{N}(i)} \exp\left(\frac{1}{\sqrt{d}}\left(\mathbf{W}_Q^k \mathbf{h}_i^{(l)}\right)^\top \left(\mathbf{W}_K^k \mathbf{h}_u^{(l)} + \mathbf{W}_E^k \mathbf{h}_{ui}^e\right)\right)} \tag{85}$$

where $\mathbf{W}_Q^k, \mathbf{W}_K^k, \mathbf{W}_E^k \in \mathbb{R}^{512\times512}$. Following `concat=False` in `TransformerConv`, multi-head messages are averaged across all K heads, combined with self-gating $\beta_i \in [0, 1]$, and passed through layer normalization:

$$\mathbf{h}_{\text{conv},i}^{(l)} = \beta_i \mathbf{W}_1 \mathbf{h}_i^{(l)} + (1 - \beta_i)\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^k \mathbf{W}_V^k \mathbf{h}_j^{(l)}\right) \tag{86}$$

$$\mathbf{h}_i^{(l+1)} = \text{Dropout}\left(\text{LayerNorm}\left(\text{GELU}\left(\mathbf{h}_{\text{conv},i}^{(l)}\right)\right), p = 0.1\right) \tag{87}$$

where $\mathbf{W}_1, \mathbf{W}_V^k \in \mathbb{R}^{512\times512}$.

H.2.4 PARAGRAPH (RELATIONAL GATED GRAPH NEURAL NETWORK)

Ren et al. (2020) introduced the foundational paradigm of modeling analog circuit schematics as multi-relational graphs with dedicated terminal-role transformations. While their original implementation combined relational attention with untied GraphSAGE layers on small hidden widths ($d \leq 64$), directly scaling untied multi-relational layers to our foundational hidden dimension ($d = 512$) would induce severe parameter explosion (> 9.5M parameters across 9 relations), breaking parity with our other baselines. Furthermore, relational attention is already isolated by our Graph Transformer (GT) baseline. Therefore, we implement this multi-relational paradigm as an R-GGNN (Relational Gated Graph Neural Network) utilizing a shared GRU cell. Recurrent weight tying across $T = 4$ propagation steps physically emulates iterative numerical relaxation sweeps while enforcing strict parameter parity (2.15M parameters) with our other $L = 4$ baselines:

$$\mathbf{m}_i^{(t)} = \mathbf{W}_{\text{root}} \mathbf{h}_i^{(t-1)} + \sum_{r=1}^{9}\sum_{j\in\mathcal{N}_r(i)} \frac{1}{|\mathcal{N}_r(i)|} \mathbf{W}_r \mathbf{h}_j^{(t-1)} \tag{88}$$

$$\mathbf{h}_i^{(t)} = \text{LayerNorm}\left(\text{GRUCell}\left(\mathbf{m}_i^{(t)}, \mathbf{h}_i^{(t-1)}\right)\right). \tag{89}$$

H.2.5 DAWN Residual Cascade and Readout Pooling

To instantiate the four DAWN-augmented variants (**DAWN MLP, DAWN GCN, DAWN GT, DAWN ParaGraph**), each backbone's final node embeddings are conditioned on operating corners via $\tilde{\mathbf{h}}_v = \mathbf{h}_v^{(L)} + \mathbf{h}_u$. The additive physical residual is computed through a 128-dimensional bottleneck cascade:

$$\mathbf{z}_v = \text{Dropout}\left(\text{GELU}\left(\mathbf{W}_1\tilde{\mathbf{h}}_v + \mathbf{b}_1\right)\right), \quad \boldsymbol{\Delta}_v = \mathbf{W}_{\text{up}}(\mathbf{W}_2\mathbf{z}_v + \mathbf{b}_2) + \mathbf{b}_{\text{up}} \tag{90}$$

where $\mathbf{W}_1 \in \mathbb{R}^{128\times512}$, $\mathbf{W}_2 \in \mathbb{R}^{128\times128}$, and $\mathbf{W}_{\text{up}} \in \mathbb{R}^{512\times128}$ with $\mathbf{W}_{\text{up}} = \mathbf{0}, \mathbf{b}_{\text{up}} = \mathbf{0}$ at initialization. Across all models, global circuit-level embeddings $\mathbf{h}_{\text{circuit}} \in \mathbb{R}^{512}$ are computed by pooling node states alongside global context $\mathbf{h}_u$:

$$\mathbf{h}_{\text{circuit}} = \text{MLP}_{\text{global}}\left(\left[\frac{1}{|\mathcal{V}|}\sum_{v\in\mathcal{V}}\mathbf{h}_v^{\text{phys}} \,\middle\|\, \max_{v\in\mathcal{V}}\mathbf{h}_v^{\text{phys}} \,\middle\|\, \mathbf{h}_u\right]\right) \tag{91}$$

where $\mathbf{h}_v^{\text{phys}} = \mathbf{h}_v^{(L)} + \boldsymbol{\Delta}_v$ for DAWN models, and $\mathbf{h}_v^{\text{phys}} = \mathbf{h}_v^{(L)}$ for standard baselines. Simultaneously, component-level operating states are predicted from $\mathbf{h}_v^{\text{dcop}} = \tilde{\mathbf{h}}_v + \boldsymbol{\Delta}_v$ (or $\tilde{\mathbf{h}}_v$ for standard baselines).