Article | Received 31 March 2026; Revised 10 June 2026; Accepted 3 July 2026; Published 30 July 2026
https://doi.org/10.55092/interdiscipline20260002

Physics-guided multi-objective Bayesian optimization for process-aware CMOS LNA design

Jithish Jayarajan[1,*], Shuoyu Ji[2], Bharatha Kumar Thangarasu[3], Nagarajan Mahalingam[3], Baoji Miao[4] and Kiat Seng Yeo[1,3]

[1] Engineering Product Development, Singapore University of Technology and Design, 487372, Singapore
[2] The International Joint Institute of Tianjin University, Fuzhou, Tianjin University, Tianjin 300072, China
[3] School of Microelectronics, Tianjin University, Tianjin 300072, China
[4] School of Materials Science and Engineering, Henan University of Technology, Zhengzhou 450001, China

* Correspondence author; E-mail: jithish_jayarajan@sutd.edu.sg.

Highlights:

- Physics-guided initialization improves valid RF simulations from 47.5% to 95.7%.
- Validity-gated Bayesian optimization for mixed discrete–continuous process design kit (PDK)-aware LNA design.
- 62% power reduction and +10.7 dB IIP3 gain @ 2.4 GHz over expert manual LNA design (via P_{dc}–NF–IIP3 Pareto tradeoff; minimum-power design NF).
- Discrete PDK inductor libraries embedded via continuous GP-compatible encoding.
- Dominated hypervolume expands 105% beyond Latin hypercube sampling alone.

Abstract: Automated analog radio-frequency (RF) circuit design remains challenging due to nonlinear device behavior, sparse feasible regions, expensive simulation-based evaluations, and discrete constraints imposed by process design kits (PDKs). This work presents a validity-gated multi-objective Bayesian optimization (MOBO) framework for Complementary Metal-Oxide-Semiconductor (CMOS) low-noise amplifier (LNA) design over a mixed discrete-continuous search space defined by foundry PDK constraints. The approach employs a physics-informed parameter initialization combined with Latin hypercube sampling (LHS) to construct an initial dataset that ensures circuit-level feasibility while maintaining device operation within bounded regions. Discrete PDK elements, including quantized inductor selections and constrained device dimensions, are directly embedded in the optimization search space. A Gaussian process (GP) surrogate model, combined with a validity-gated acquisition strategy, then guides the search by filtering out infeasible evaluations and improving simulation efficiency throughout the optimization process. The framework is validated on a 2.4 GHz current reuse cascode LNA designed in Cadence Virtuoso and simulated using Spectre in a Taiwan Semiconductor Manufacturing Company (TSMC) 40 nm CMOS process. Under a fixed budget of 190 simulations (120 LHS + 70 MOBO), the method achieves a 62% reduction in power consumption and a 10.7 dB

improvement in the input third-order intercept point (IIP3) @ 2.4 GHz relative to a handcrafted expert baseline, reflecting a P_{dc}–noise figure (NF)–IIP3 Pareto trade-off. The minimum-power Pareto design exhibits a modest NF increase (2.19 dB *vs* 1.98 dB for the manual baseline), while NF-optimized Pareto solutions achieve NF as low as 1.84 dB. The dominated hypervolume expands by 105% over LHS initialization alone, and valid simulation yield improves from 47.5% to 95.7% under guided optimization. These results demonstrate that physics-informed initialization combined with validity-gated Bayesian optimization enables efficient exploration of constrained, mixed-variable CMOS RF design spaces, achieving expert-level performance without manual heuristic tuning.

Keywords: multi-objective Bayesian optimization; Gaussian process surrogate; CMOS low-noise amplifier; RF circuit synthesis; mixed discrete–continuous optimization; physics-informed initialization

1. Introduction

Automated analog and radio frequency (RF) circuit design remains fundamentally more challenging than digital design due to nonlinear device behavior, tightly coupled performance metrics, and strict feasibility constraints [1,2]. In Complementary Metal-Oxide-Semiconductor (CMOS) low-noise amplifier (LNA) design, objectives such as power consumption, noise figure (NF), gain, and linearity (input third-order intercept point, IIP3) form mutually conflicting trade-off surfaces. Unlike purely continuous formulations, practical RF circuit optimization must also satisfy discrete constraints imposed by process design kits (PDKs) [3], such as quantized inductor libraries and layout grid constraints on device geometries [4]. These constraints produce highly nonconvex, mixed discrete–continuous design spaces with sparse feasible regions [5] and frequent simulation failures under unguided exploration [6,7].

Traditional LNA design methodologies rely on analytical approximations, g_m/I_D heuristics, impedance matching theory, and iterative simulation refinement within design tools [8,9]. While effective under experienced designers, these approaches are time-intensive and do not guarantee systematic exploration of global multi-objective trade-offs. Moreover, random or uniform sampling strategies often yield high invalid-simulation rates due to violations of transistor operating regions, matching constraints, or biasing conditions.

Bayesian optimization (BO) has emerged as an effective framework for sample-efficient optimization under expensive black-box evaluations [10]. However, direct application of BO to RF circuit synthesis presents several challenges, including sparse feasible regions arising from stringent RF stability and bias constraints, mixed discrete–continuous decision variables under PDK constraints, high invalid-simulation rates under unguided exploration, and conflicting multi-objective tradeoffs [11,12]. Our preliminary experiments on LNA design using uniform LHS initialization resulted in an invalid simulation rate of 52.5%, primarily caused by violations of transistor operating regions, impedance matching constraints, and bias feasibility conditions. In sparse RF design spaces, such invalid evaluations can degrade surrogate model quality and reduce sample efficiency.

These observations motivate the need for physics-guided initialization, which embeds physical relationships between circuit elements into the sampling process to ensure valid device operation while maintaining sufficient diversity for surrogate learning. Towards this objective, this work proposes a physics-guided multi-objective Bayesian optimization (MOBO) framework for PDK-constrained LNA design. The key novelty of this work is a unified framework that, unlike existing approaches, jointly

integrates physics-guided initialization, strict discrete PDK library enforcement, and validity-gated surrogate model training for RF circuit design. In contrast to conventional manual LNA design cycles that may span several days of iterative simulation-tune cycles, the proposed MOBO framework achieves competitive solutions within 190 simulations, corresponding to approximately 16–19 hours of raw simulation time in our setup.

Figure 1 summarizes the proposed three-phase framework. Phase I generates PDK-realizable warm-start designs via physics-informed sampling, combining LHS for continuous parameters with discrete inductor sampling, followed by resonance-guided derivation and library snapping. Phase II encodes designs into a multidimensional space, evaluates them in Spectre, and applies a validity gate, in which only physically valid designs are retained for surrogate training. Phase III fits Gaussian process (GP) surrogate models and performs constrained MOBO, snapping candidates to discrete libraries and iteratively refining them to expand the feasible Pareto front within a fixed simulation budget.

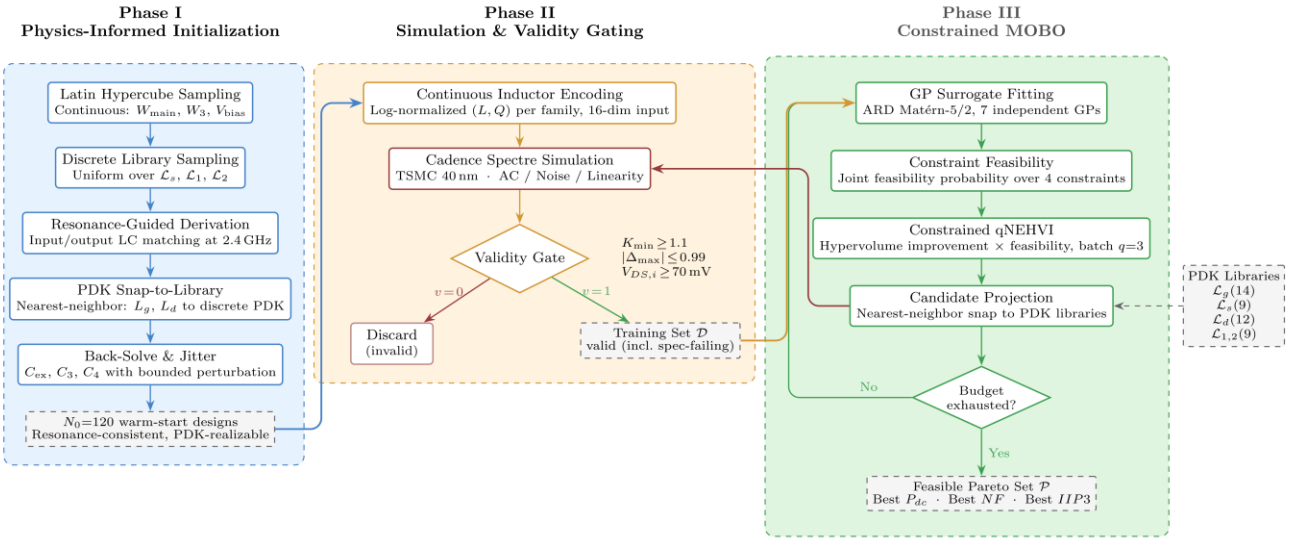

Figure 1. Overview of the proposed physics-guided MOBO framework. Phase I generates resonance-consistent, PDK-realizable warm-start designs via physics-guided sampling. Phase II encodes and evaluates designs using Spectre with validity gating. Phase III performs GP-based constrained MOBO.

The main contributions of this work are summarized as follows:

- A physics-guided initialization strategy that reduces early-stage infeasibility in sparse RF design spaces.
- A continuous embedding of discrete PDK inductor libraries that enables GP-based MOBO while ensuring physically realizable component selection.
- A validity-gated constrained multi-objective optimization framework that improves surrogate model quality and sample efficiency.
- Experimental validation on a 2.4 GHz current reuse CMOS LNA implemented in the TSMC 40 nm PDK under fixed simulation budgets.

Experimental results demonstrate substantial improvements in power efficiency, linearity, and valid simulation yield relative to uniform LHS initialization and a handcrafted expert baseline. The remainder of this paper is organized as follows. Section 2 reviews related Radio-Frequency Integrated Circuit (RFIC)

synthesis approaches. Section 3 introduces the target LNA topology. Section 4 presents the formal problem formulation. Section 5 describes the proposed physics-guided initialization, Section 6 presents the proposed constrained discrete-aware MOBO framework. Section 7 details the experimental setup. Section 8 presents experimental results. Section 9 highlights the limitations of the proposed approach, and Section 10 concludes the paper.

2. Related works

Stringent RFIC performance targets increasingly expose the limits of manual, iteration-heavy design flows [13]. LNA optimization requires a careful balance among mutually conflicting objectives such as wideband gain, NF, linearity (e.g., IIP3), and power consumption. To reduce reliance on designer intuition and costly post-layout electromagnetic (EM) iterations, Artificial Intelligence/Machine Learning (AI/ML)-based methods are widely adopted for automated parameter sizing and circuit synthesis [14]. Recent surveys highlight a shift from closed-form, equation-driven optimization toward data-driven modeling of highly nonlinear RF metrics in nanometer-scale processes [15]. The optimization approaches can be broadly categorized as:

2.1. Neural network-based surrogate modeling for performance prediction

NN-based surrogate models map device-level sizing parameters to RF performance and enable rapid evaluation after training. Cascade architectures of shallow networks have been used to progressively constrain the prediction space, enabling LNA and Voltage-Controlled Oscillator (VCO) sizing while capturing interdependent components [16]. Deep learning and multi-level perceptron (MLP) models have also been applied to predict S-parameters and NF for differential cascode LNAs across temperature and frequency variations [17]. Graph convolutional neural networks (GCN) can natively represent topology and facilitate transfer of circuit knowledge [18], and automatic differentiation can accelerate matching-network synthesis [19]. Despite these advances, NN surrogates remain data-intensive as acquiring large, high-fidelity SPICE/EM datasets for wideband RF blocks is often prohibitively expensive, limiting scalability in layout-aware flows [20,21].

2.2. RL-based automated sizing and synthesis

To overcome the limitations of static pre-simulated datasets, Reinforcement Learning (RL) has emerged as a dynamic alternative in which an agent learns continuous sizing policies through direct interaction with electronic design automation (EDA) simulators. Deep reinforcement learning (DRL) algorithms, such as proximal policy optimization (PPO), have demonstrated expert-level performance in the design of wideband distributed architectures, while multi-actor RL frameworks have further improved sampling efficiency and convergence robustness for analog optimization [22,23]. In [24], RL is combined with swarm intelligence to develop a hierarchical co-design framework that efficiently synthesizes complex impedance matching networks. Recent open-source frameworks further extend RL to heterogeneous graph neural networks, enabling optimization of mixed-signal circuits from the schematic stage through post-layout constraints [25]. The integration of neural simulators with RL agents has further shortened RF front-end design cycles [26]. However, RL often suffers from sample inefficiency and the curse of

dimensionality in continuous action spaces [27], making convergence difficult for tightly constrained, multi-objective LNA designs.

2.3. Evolutionary algorithms: global search and limitations

Before the rise of deep learning, evolutionary algorithms (EAs) such as particle swarm optimization (PSO) and genetic algorithms (GA) were standard tools for global exploration in analog sizing. They have been applied to optimize gain–power tradeoffs in RF receivers [28,29]. Despite robust global search behavior and implementation simplicity, EAs typically require thousands of simulations to converge. In advanced nodes, where layout-dependent effects (LDE) and parasitic coupling demand heavy verification, this simulation burden becomes a major bottleneck [30]. Foundational surveys of GA for system engineering [31] document their strength in global, population-based search but underscore the large number of function evaluations required for convergence, which is a critical drawback when each evaluation involves an expensive multi-hour circuit simulation. In contrast, BO replaces population-based random variation with a probabilistic surrogate model. This model estimates both the expected objective value and the uncertainty at each candidate point. An acquisition function then guides evaluations toward regions that are either promising or uncertain. As a result, BO can often converge within tens to hundreds of simulations rather than thousands.

2.4. Bayesian Optimization for sample-efficient sizing

BO has gained traction in RFIC design because of its strong sample efficiency for black-box objectives under tight simulation budgets [32]. Gaussian process (GP) surrogates combined with acquisition functions balance exploration and exploitation, reducing the number of SPICE evaluations. Recent advances include mixed-variable BO for discrete/continuous components [33] and multitask neural-network-enhanced GP models (MTNN-GP) that exploit cross-metric correlations to accelerate convergence [34,35].

2.5. Comparison and positioning of this work

Table 1 summarizes the presence or absence of critical capabilities in a binary feature format, highlighting distinguishing characteristics of the proposed approach relative to prior work. Table 2 provides a descriptive qualitative comparison of representative automated analog/RF circuit synthesis methods across key dimensions, including optimization strategy, initialization, constraint handling, and automation level. Together, these two tables position this work relative to classic baselines and motivate the need for a tailored MOBO framework for constrained RFIC design. The key novelty of the proposed approach lies in the joint integration of resonance-guided initialization, strict enforcement of discrete PDK libraries, and validity-gated surrogate training within a unified optimization pipeline. This addresses key limitations of existing methods, which typically treat these aspects independently. By explicitly modeling the mixed discrete–continuous design space and incorporating resonance-informed initialization, the proposed framework mitigates challenges arising from sparse feasibility and discontinuous objective landscapes due to discrete PDK inductor selection.

Table 1. Capability-based binary comparison of automated analog/RF circuit synthesis methods.

Feature	This Work	Beaulieu *et al.* [16]	Wang *et al.* [18]	Sun *et al.* [22]	Shreeharsha *et al.* [28]	Huang *et al.* [34]	Budak *et al.* [36]	Lyu *et al.* [37]	Li *et al.* [38]	Somayaji *et al.* [39]	Budak *et al.* [40]
Physics-informed initialization	✓	×	×	×	×	×	×	×	×	✓	×
Hard discrete PDK enforcement	✓	×	×	×	×	×	×	×	×	×	×
Validity-gated acquisition	✓	×	×	×	×	×	×	×	×	×	×
Multi-objective (Pareto-based)	✓	×	×	×	×	×	×	✓	×	×	×
SPICE-level simulation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fully automated flow	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓
GP surrogate model	✓	×	×	×	×	✓	×	✓	×	×	×
No expert tuning required	✓	×	✓	×	×	×	×	✓	✓	×	✓
Mixed discrete–continuous domain	✓	×	×	×	×	×	×	×	×	×	×
Commercial PDK compatible	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 2. Qualitative comparison of automated analog/RF circuit synthesis methods.

Aspect	This Work	Beaulieu *et al.* [16]	Wang *et al.* [18]	Sun *et al.* [22]	Shreeharsha *et al.* [28]	Huang *et al.* [34]	Budak *et al.* [36]	Lyu *et al.* [37]	Li *et al.* [38]	Somayaji *et al.* [39]	Budak *et al.* [40]
Optimization Framework	MOBO (GP)	C-SNN	GCN-RL	DRL (MDP)	EB-PSO	MTNN-GP BO	DNN-Opt	MOBO (GP)	ANN+GA	RL	ANN+DE
Initialization Strategy	Physics-guided	Random/ LHS	Random	Random exploration	Random	Random	Augmented samples	LHS	LHS	Probabilistic	LHS
Discrete PDK Handling	Hard enforcement	Rounding	Continuous	Stepwise tuning	Rounding	Continuous	Continuous	Continuous	Continuous	Continuous	Continuous
Feasibility Enforcement	Validity-gated	Constraint handling	Scalar FoM	Exponential reward	Velocity penalty	PoF-weighted	Aggregation	Dominance	Weighted cost	FoM	Ranking
Surrogate Model	GP	C-SNN +GA	GCN+DDPG	None	None	MTNN-GP	DNN	GP	ANN	DDPG	ANN
Simulator Integration	Spectre	Spectre	Spectre/ HSPICE	ADS	Spectre	SPICE-level	SPICE-level	SPICE-level	SPICE-level	HSPICE	Spectre
Automation Level	Full	Semi	Full	Full	Full	Full	Full	Full	Full	Full	Full
Expert Knowledge Req.	Moderate	Moderate	Low	High	Low	Moderate	Moderate	Low	Low	High	Low
Circuit Topology	CS/CG LNA	Cascode LNA, VCO, CCGC Mixer	2-Stage TIA, 2-Stage V-Amp, 3-Stage TIA, LDO	NDPA	Op-Amp	XFMR, 3-Stage Cascade Diff. LNA, PA	FC OTA, SA Comp.	3-Stage PA	2-Stage RR Op-Amp, 5th-Order Active-RC CBPF	2-Stage Diff. Amp., FC Amp., Hyst. Comp.	FC OTA, ISA, SLC, VCO
Technology Node	TSMC 40 nm	180/ 130 nm	180 nm	GaN HEMT	45 nm	65/40 nm	180 nm	55 nm	180/ 130 nm	90 nm	180/40 nm
Multi-Objective Aware	Yes	No	No	No	No	No	No	Yes	No	No	No

Abbreviations: MOBO–multi-objective Bayesian optimization; GP–Gaussian process; C-SNN–cascade shallow neural network; GCN–graph convolutional neural network; RL–reinforcement learning; DRL–deep reinforcement learning; EB-PSO–error-bound particle swarm optimization; MTNN–multitask neural network; DNN–deep neural network; ANN–artificial neural network; GA–genetic algorithm; DE–differential evolution; DDPG–deep deterministic policy gradient; LHS–Latin hypercube sampling; PoF–probability of feasibility; CS/CG–common-source/common-gate; LNA–low noise amplifier; VCO–voltage-controlled oscillator; CCGC–cross-coupled Gilbert cell; TIA–transimpedance amplifier; LDO–low-dropout regulator; NDPA–non-uniform distributed power amplifier; Op-Amp–operational amplifier; OTA–operational transconductance amplifier; FC–folded-cascode; XFMR–transformer; RR–rail-to-rail; CBPF–complex band-pass filter; ISA–inverter-stacking amplifier; SLC–strong-arm latch comparator.

3. Target LNA topology

This section describes the 2.4 GHz LNA [41] in TSMC 40 nm CMOS used as the optimization target (Figure 2). This baseline is adapted from the 2.4 GHz high-gain LNA topology reported by Radic *et al.* [41]. The original circuit was reimplemented by the authors in TSMC 40 nm CMOS using Cadence Virtuoso, as the technology node in [41] differs from the 40 nm PDK used in this study. Component values were then manually tuned by an experienced RF designer within the same Cadence Virtuoso–Spectre environment used for MOBO evaluation, targeting the same specification constraints listed in Table 3 ($S_{21} \geq 19$ dB, $S_{11} \leq -12$ dB, $S_{22} \leq -12$ dB, OP1dB ≥ -5 dBm, stability margins). The resulting manual design achieves $P_{dc} = 14.1$ mW, NF = 1.98 dB, and IIP3 = −12.0 dBm at 2.4 GHz, and represents a hand-crafted baseline for this topology and technology node.

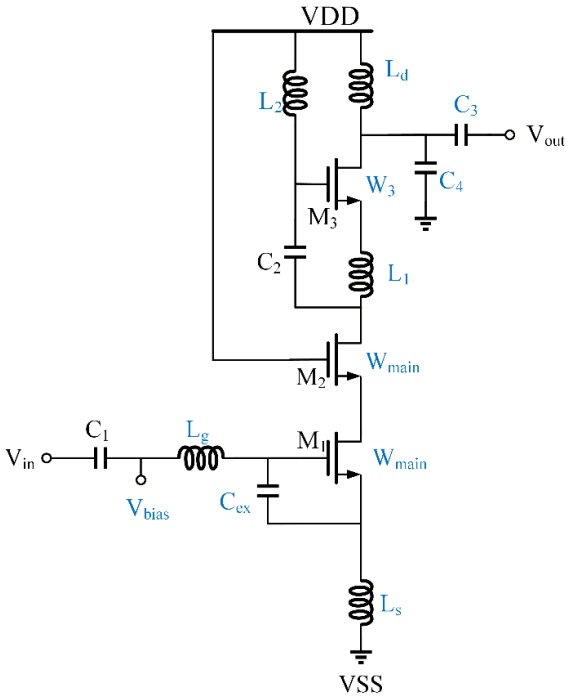

Figure 2. Circuit topology of the 2.4 GHz current reuse cascode CMOS LNA with inductive source degeneration and inter-stage matching network from modified circuit topology based on [41].

Table 3. Design constraints used during optimization.

Constraint	Metric	Limit
Gain	S_{21} @ 2.4 GHz	≥ 19 dB
Input match	S_{11} @ 2.4 GHz	≤ –12 dB
Output match	S_{22} @ 2.4 GHz	≤ –12 dB
Compression	OP1dB	≥ –5 dBm
Stability	K_{min}	≥ 1.1
Stability	$\lvert\Delta_{max}\rvert$	≤ 0.99
Saturation (M1)	$V_{DS,margin}$	≥ 70 mV
Saturation (M3)	$V_{DS,margin}$	≥ 70 mV

The proposed baseline retains the current-reuse cascode concept of Radic *et al.* [41] but differs in technology node, matching, and reported metrics, as summarized below. Reference [41] is realized in a 0.35 µm SiGe BiCMOS process (MOS devices) at 3.3 V, matches its input by simple inductive source degeneration, and tunes the output to 100 Ω for a differential mixer. The present baseline is reimplemented

in a 40 nm bulk-CMOS PDK at 1.1 V, adds an auxiliary gate–source capacitor for simultaneous noise and input matching under the reduced voltage headroom, and matches the output to 50 Ω. It attains a lower NF (1.98 *vs* 2.7 dB) and lower power (14.1 *vs* 16.5 mW) at comparable gain.

The design employs an inductively degenerated cascode with a CS transistor (M_1) and CG transistor (M_2). To balance NF_{min} and S_{11}, a PCSNIM-based input network [42] is used, consisting of C_1, L_g, and L_s, with an auxiliary capacitor C_{ex} between gate and source of M_1 to enable flexible simultaneous noise and input matching under low-power operation. A third transistor (M_3) is stacked to reuse bias current. The inter-stage network (L_1, L_2, C_2) enables efficient signal transfer and gain enhancement through resonance tuning. Output matching to $50\,\Omega$ is achieved using L_d, C_4, and C_3. Despite its high performance, the resulting high-order interactions make stable parameter sizing challenging for conventional EDA methods.

4. Problem formulation and design space

4.1. LNA topology and design variables

Quantized inductor selections and bounded device geometries from the PDK create discontinuous and nonconvex regions in the LNA's objective landscape. To ensure fabrication feasibility and stable surrogate learning across discontinuous regions, the optimization framework must explicitly account for discrete PDK constraints in the search space. Let the design vector be

$$\boldsymbol{x} = (\boldsymbol{x_c}, \boldsymbol{x_d}) \in \boldsymbol{X} \tag{1}$$

where $x_c \in R^{d_c}$ denotes a d_c-dimensional real-valued vector of continuous circuit parameters subject to bounded constraints, and $x_d \in D$ denotes discrete inductor selections from EM-characterized libraries. The continuous variables are

$$\boldsymbol{x_c} = \{\boldsymbol{W_{main}}, \boldsymbol{W_3}, \boldsymbol{V_{bias}}, \boldsymbol{C_{ex}}, \boldsymbol{C_3}, \boldsymbol{C_4}\} \tag{2}$$

and the discrete variables correspond to inductors selected from role-specific libraries:

$$\boldsymbol{x_d} = \{\boldsymbol{L_g}, \boldsymbol{L_s}, \boldsymbol{L_d}, \boldsymbol{L_1}, \boldsymbol{L_2}\} \qquad \boldsymbol{L_t} \in \boldsymbol{\mathcal{L}_t}, \tag{3}$$

where $t \in g, s, d, 1,2$ indexes the inductor role and $\mathcal{L}_t$ denotes the discrete library set for that role. Each library entry is characterized at the operating frequency $f_0 = 2.4$ GHz by

$$\boldsymbol{\ell_{t,j}} = \left(\boldsymbol{L_{t,j}}, \boldsymbol{Q_{t,j}}, \boldsymbol{SRF_{t,j}}\right), \qquad \boldsymbol{j} = \boldsymbol{1}, \ldots, |\boldsymbol{L_t}|, \tag{4}$$

Where, $\ell_{t,j} \in \mathbb{R}^3$ denotes the log-descriptor vector (continuous embedding) for the j^{th} entry in the discrete PDK inductor library for role t, where $t \in \{L_g, L_d, L_s, L_1, L_2\}$ indexes the functional role of inductor in the circuit. Here, $L_{t,j}$ is the inductance, $Q_{t,j}$ is the quality factor, and $SRF_{t,j}$ is the self-resonant frequency.

4.2. Mixed discrete–continuous design space

The discrete inductor domain is defined as the Cartesian product

$$\boldsymbol{D} = \boldsymbol{L_g} \times \boldsymbol{L_s} \times \boldsymbol{L_d} \times \boldsymbol{L_1} \times \boldsymbol{L_2} \tag{5}$$

where each L_t denotes the finite EM-characterized library associated with inductor role t. The continuous design domain is given by

$$B = \prod_{k=1}^{d_c} [a_k, b_k] \subset R^{d_c}, \tag{6}$$

which forms a d_c-dimensional hyper-rectangular region defined by physically feasible bounds on each continuous parameter. Here, a_k and b_k denote the lower and upper bounds of the k^{th} continuous design parameter, respectively, as listed in Table 4.

The admissible mixed design domain is therefore

$$\boldsymbol{X} = \{(\boldsymbol{x_c}, \boldsymbol{x_d}) \mid \boldsymbol{x_c} \in \boldsymbol{B},\ \boldsymbol{x_d} \in \boldsymbol{D}\} = \boldsymbol{B} \times \boldsymbol{D}. \tag{7}$$

Here, x_c denotes the vector of continuous design variables, including transistor sizing and bias parameters, while x_d denotes the vector of discrete PDK inductor selections. By incorporating discrete PDK elements together with continuous transistor and bias parameters, the design space remains fabrication-realistic and avoids errors or performance loss from post-optimization discretization. The transistor channel length is fixed at 40 nm for all devices in this work. The use of a discrete PDK inductor library is an implementation choice in the proposed workflow, adopted to ensure that all candidate designs are directly realizable and do not require post-optimization rounding and is not a universal requirement of RFIC design. This choice is mainly motivated by the layout- and electromagnetic-dependent nature of on-chip inductors, for which the required inductance and associated parasitics cannot be accurately determined from simple closed-form equations alone.

Table 4. Tunable design variables.

Parameter	Symbol	Type	Physical Role
Gate inductor	L_g	Discrete	Input resonance
Source degen. inductor	L_s	Discrete	NF–linearity trade-off
Drain inductor	L_d	Discrete	Output resonance
Inter-stage inductors	L_1, L_2	Discrete	Resonance tuning
Main transistor width	W_{main}	Continuous	Gain and NF control
Cascode width	W_3	Continuous	Linearity control
Bias voltage	V_{bias}	Continuous	Operating region
Coupling capacitance	C_{ex}	Continuous	Input tuning
Output capacitances	C_3, C_4	Continuous	Output impedance match

4.3. Multi-objective formulation

Let $x \in X$ denote the circuit design parameter vector, and define the vector-valued objective function

$$\boldsymbol{f}: \boldsymbol{X} \rightarrow \boldsymbol{R^3} \tag{8}$$

As

$$\boldsymbol{f(x)} = \begin{bmatrix} \boldsymbol{f_1(x)} \\ \boldsymbol{f_2(x)} \\ \boldsymbol{f_3(x)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{P_{dc}(x)} \\ \boldsymbol{NF(x)} \\ \boldsymbol{-IIP3(x)} \end{bmatrix}, \tag{9}$$

where P_{dc} denotes direct current (DC) power consumption, NF denotes NF @ 2.4 GHz, and $IIP3$ denotes the input third-order intercept point @ 2.4 GHz. Since $IIP3$ is to be maximized, it is negated. Table 5 summarizes the optimization objectives. To preserve surrogate modeling stability and improve sample efficiency, the number of optimization objectives is limited to three. Additional RF

requirements, including gain, matching, compression point, and stability margins, are enforced as constraints rather than additional objectives.

This multi-objective formulation is a modeling choice with two motivations. First, from a surrogate modelling perspective, restricting the Pareto objectives to three (P_{dc}, NF, IIP3) preserves GP surrogate quality and keeps q-noisy expected hypervolume improvement (qNEHVI) computationally tractable. The hypervolume improvement acquisition scales exponentially with the number of objectives, and GP inference quality degrades in high-dimensional objective spaces within the modest 190-evaluation budget used here. Second, from an RF design practice perspective, gain, impedance matching, compression point, and stability margins are fundamentally binary acceptance criteria rather than continuous trade-off dimensions. A design that fails $S_{21} \geq 19$ dB or $S_{11} \leq -12$ dB is not deployable regardless of its power or linearity.

Table 5. Optimization objectives.

Objective	Metric	Direction
Power consumption	P_{dc} (mW)	Minimize
Noise performance	NF @ 2.4 GHz (dB)	Minimize
Linearity	$IIP3$ @ 2.4 GHz (dBm)	Maximize

4.4. Constraint set and validity gate

The proposed framework distinguishes between two types of constraints: validity constraints, which ensure physical realizability and determine eligibility for surrogate training, and performance constraints, which define electrical specifications and are enforced through the acquisition function.

4.4.1. Performance constraints

Performance constraints evaluated at the operating frequency $f_0 = 2.4$ GHz are defined as

$$\begin{aligned} \boldsymbol{c_{S21}(x)} &= \boldsymbol{S_{21}^{min} - S_{21}(x)}, \\ \boldsymbol{c_{S11}(x)} &= \boldsymbol{S_{11}(x) - S_{11}^{max}}, \\ \boldsymbol{c_{S22}(x)} &= \boldsymbol{S_{22}(x) - S_{22}^{max}}, \\ \boldsymbol{c_{OP1}(x)} &= \boldsymbol{OP1dB^{min} - OP1dB(x)}. \end{aligned} \tag{10}$$

Designs that satisfy validity conditions but violate performance constraints remain part of the surrogate training dataset, enabling the constraint surrogates to learn both feasible and infeasible performance regions.

4.4.2. Validity constraints

Validity constraints ensure physically meaningful circuit operation. Wideband stability constraints evaluated over a frequency sweep are

$$\begin{aligned} \boldsymbol{c_K(x)} &= \boldsymbol{K_{req} - K_{min}(x)}, \\ \boldsymbol{c_\Delta(x)} &= \boldsymbol{\Delta_{max}(x) - \Delta_{req}}, \end{aligned} \tag{11}$$

where:

$$\boldsymbol{K_{min}(x) = \min_f K(x;f)}, \qquad \boldsymbol{\Delta_{max}(x) = \max_f |\Delta(x;f)|}. \tag{12}$$

The Rollett stability factor is

$$K(x;f) = \frac{1-|S_{11}|^2-|S_{22}|^2+|\Delta|^2}{2|S_{12}S_{21}|}, \tag{13}$$

with determinant

$$\Delta = S_{11}S_{22} - S_{12}S_{21}. \tag{14}$$

Device saturation validity is enforced as

$$c_{V_{DS},i}(x) = V_{req} - V_{DS,margin,i}(x), \tag{15}$$

where $V_{req} = 70\,\text{mV}$ and $i \in \{M1, M3\}$ indexes the main and cascode transistors.

4.4.3. Validity gate

The validity operator is defined as

$$v(x) = I\left(c_K(x) \le 0 \ \wedge \ c_\Delta(x) \le 0 \ \wedge \ \bigwedge_i c_{V_{DS},i}(x) \le 0\right), \tag{16}$$

where $I(\cdot)$ denotes the indicator function. Only samples satisfying *v(x) = 1* are incorporated into surrogate model training. Samples violating validity constraints correspond to unstable or improperly biased circuit configurations whose simulator outputs are physically unreliable and are therefore excluded from surrogate training. The numerical limits for all constraints are summarized in Table 3.

A simulation is classified as invalid if the Spectre simulator fails to converge or returns a non-physical result (e.g., negative power or divergent S-parameters), or one or more validity constraints (stability, device saturation) are violated. The valid simulation yield (or valid simulation rate) is defined as the fraction of evaluated design points for which the validity operator returns *v(x) = 1*. Using the validity operator:

$$\text{Valid yield} = \frac{1}{N}\sum_{i=1}^{N} v(x_i), \tag{17}$$

where $v(x_i) = 1$, if simulation i is valid and $v(x_i) = 0$ otherwise.

4.5. Feasible region and pareto optimality

The feasible region is defined as the set of designs satisfying both validity and performance constraints:

$$F = \{x \in X \mid v(x) = 1, \ c_k(x) \le 0 \ \ \forall k\}. \tag{18}$$

A feasible design $x^\star \in F$ is Pareto optimal if no other feasible design dominates it:

$$\nexists x \in F \text{ such that } f(x) \prec f(x^\star), \tag{19}$$

where Pareto dominance $\prec$ is defined as

$$f(x) \prec f(x^\star) \Leftrightarrow \quad \forall\, i:\ f_i(x) \le f_i(x^\star), \quad \exists\, j:\ f_j(x) < f_j(x^\star). \tag{20}$$

The Pareto set is

$$P = \{x^\star \in F \mid x^\star \text{ is Pareto optimal}\}. \tag{21}$$

Given a reference point r, the dominated hypervolume of P is

$$HV(P; r) = \lambda\left(\bigcup_{u \in P} [u, r]\right), \tag{22}$$

where $\lambda(\cdot)$ denotes Lebesgue measure and $[u, r]$ is the axis-aligned hyperrectangle bounded by objective vector u and reference point r.

4.6. Complete constrained multi-objective problem

The complete constrained multi-objective problem solved in this work can be expressed as

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & x_d \in D, \\ & v(x) = 1, \\ & c_k(x) \le 0, \quad \forall k. \end{aligned} \tag{23}$$

where:

- $x_d \in D$ enforces discrete component selections from the PDK library.
- $v(x) = 1$ is a validity gate ensuring only physically reliable simulations are used for surrogate modeling.
- $c_k(x) \le 0$ are performance and feasibility constraints for all indices k.

This unified formulation incorporates discrete constraints, physics-based validity, and multi-objective trade-offs in a single constrained framework, guiding the optimizer to find feasible, physically realizable solutions.

The constraint thresholds in Table 3 were selected based on reference design specifications and practical RFIC design margins. The gain specification ($S_{21} \ge 19$ dB), input match ($S_{11} \le -12$ dB), and output match ($S_{22} \le -12$ dB) are consistent with standard 2.4 GHz ISM-band LNA performance requirements. The linearity floor ($OP1dB \ge -5$ dBm) reflects a minimum acceptable compression-point target for low-power receiver front ends in this band. The stability margins ($K_{min} \ge 1.1$ and $|\Delta_{max}| \le 0.99$) extend the theoretical unconditional-stability thresholds ($K > 1$ and $|\Delta| < 1$) by 10% and 1%, respectively, to provide a guard-band against near-boundary instability across process, voltage, and temperature (PVT) variations. The saturation margin ($V_{DS,margin} \ge 70$ mV) ensures both transistors remain well into the saturation region under nominal bias, providing robustness against threshold-voltage and supply variation in 40 nm CMOS.

5. Physics-guided initialization

Instead of uninformed random sampling, which produces many unstable or non-resonant designs in high-dimensional mixed spaces, we use physics-guided initialization. This method generates near-resonant, PDK-feasible designs that maintain circuit consistency and enough diversity for surrogate learning. The

warm-start dataset combines space-filling sampling of continuous transistor variables, uniform sampling of independent discrete inductors, and physics-guided derivation of dependent inductors and capacitors via resonance relations. This ensures physically meaningful initial designs while strictly enforcing realizability within discrete component libraries. The term "physics-guided" is used to describe the proposed initialization procedure. The warm-start strategy is based on first-order inductor–capacitor (LC) resonance relations, linear capacitance-density approximations, reference-design scaling, and multiplicative jitter. These components incorporate circuit-level prior knowledge into the sampling process but remain heuristic approximations rather than strict physical constraints derived from a complete analytical model.

5.1. Continuous variable sampling via Latin Hypercube Sampling

The primary continuous transistor variables are sampled using LHS to ensure uniform coverage of the feasible design space. Let the continuous design vector be

$$x_c = [W_{main},\ W_3,\ V_{bias}]^T. \tag{24}$$

Samples are drawn in normalized space:

$$u_i \sim LHS([0,1]^3), \qquad i = 1, \dots, N_0, \tag{25}$$

and mapped to physical bounds:

$$x_{c,i}^{(k)} = a_k + u_i^{(k)}(b_k - a_k), \qquad k = 1,2,3, \tag{26}$$

where, $[a_k, b_k]$ denote physically valid parameter ranges. Here, a_k and b_k denote the lower and upper physical bounds (from Table 3) of the k^{th} continuous parameter.

5.2. Independent discrete inductor sampling

Certain inductors are treated as independent design variables and are sampled uniformly from their respective discrete libraries:

$$\begin{aligned} L_s &\sim Uniform(L_s) \\ L_1 &\sim Uniform(L_1) \\ L_2 &\sim Uniform(L_2) \end{aligned} \tag{27}$$

The inductors (L_s, L_1, L_2) are independently sampled, as they are not uniquely constrained by resonance, enabling structural diversity and unbiased exploration of the design space. In contrast, L_g and L_d are derived from input and output resonance conditions to ensure physically consistent impedance matching. Specifically, L_g is computed from (W_{main}, L_s) and L_d from W_3, after which both are snapped to the nearest realizable values in the discrete PDK libraries.

The gate and drain inductors, $L_g$and L_d, are primarily associated with the input and output matching networks at 2.4 GHz, respectively. For these elements, first-order LC resonance relations provide physically meaningful and analytically tractable estimates, making resonance-guided initialization appropriate.

By contrast, L_s participates in input matching through source-degeneration feedback, but also strongly influences the NF-linearity trade-off and is therefore not determined solely by a single resonance condition. Similarly, $L_1$and L_2 affect interstage resonance and gain transfer, but operate in a multi-node impedance environment that depends on transistor operating point, cascode loading, and

interstage parasitics. Reducing these effects to a single closed-form resonance expression would require additional simplifying assumptions and may introduce modelling error. Therefore, L_s, L_1, and L_2 are sampled independently from the PDK inductor library. This allows the GP surrogate to learn effective source-degeneration and interstage tuning conditions directly from Spectre simulations, rather than imposing potentially inaccurate analytical constraints on the initialization.

5.3. Input resonance-guided derivation and snapping (L_g, C_{ex})

The intrinsic device input capacitance is modeled using a calibrated capacitance density:

$$C_{in,dev}(W_{main}) = \kappa_{in}\, W_{main} \tag{28}$$

where, W_{main} is the width of the main amplifying transistor (in μm), and κ_{in} (in fF/μm) is the technology-calibrated input capacitance density at the minimum channel length L = 40 nm extracted from reference device characterization.

While the gate capacitance generally depends on both device width W and channel length L ($C_{\text{gate}} \approx C_{\text{ox}}WL$), all transistors in this design use the fixed minimum channel length $L = 40$ nm throughout the optimization. Therefore, with L held constant, κ_{in} is treated as an effective input capacitance density per unit width, expressed in fF/μm, and extracted from reference device characterization at $L = 40$ nm.

Let the target angular resonance frequency be

$$\omega_0 = 2\pi f_0, \tag{29}$$

where, $f_0 = 2.4$ GHz is the RF center frequency at which input and output matching are enforced. To compute a resonance-consistent gate inductance, a reference coupling capacitance $C_{ex,ref}$ is used to define a proxy total input capacitance:

$$C_{in,tot,ref} = \kappa_{in}\, W_{main} + C_{ex,ref}, \tag{30}$$

where, $C_{ex,ref}$ is a nominal reference value for the external coupling capacitance, set to the midpoint of its allowed range, *i.e.*, $C_{ex,ref} = \left(C_{ex,min} + C_{ex,max}\right)/2$. Here $C_{ex,min}$ and $C_{ex,max}$ denote the lower and upper physical bounds of the coupling capacitance (Table 6). The corresponding ideal gate inductance is

$$L^{\star}_{g,H} = \frac{1}{\omega_0^2\, C_{in,tot,ref}} - L_s, \tag{31}$$

where, L_s is the source degeneration inductance, sampled uniformly from its discrete library L_s prior to this step. For implementation, a small floor is applied prior to discrete snapping:

$$L^{\star}_g = max\left(10^9\, L^{\star}_{g,H},\ 0.1\right), \tag{32}$$

where, the floor of 0.1 nH prevents non-physical negative or zero values that can arise when L_s is large relative to the total required inductance. Since only discrete inductors are available, the ideal value is snapped to the nearest realizable element in the inductor library. The value of 0.1 nH is used only as a conservative numerical guard. It was chosen to be smaller than the minimum available entry in the L_s discrete library, which is 0.2 nH as listed in Table 6. This ensures that, if the resonance-based estimate yields a near-zero or slightly negative $L^{\star}_g$, the subsequent library-snapping step remains well defined and maps the value to the nearest valid PDK inductor entry, namely 0.2 nH.

$$L_g = \underset{L \in L_g}{argmin} \left| L - L^{\star}_g \right|. \tag{33}$$

Given the snapped inductors (L_g, L_s), the resonance-consistent coupling capacitance is back-solved:

$$C_{ex}^{\star} = \frac{1}{\omega_0^2 (L_g + L_s)} - \kappa_{in} W_{main}. \tag{34}$$

To enforce practical bounds while introducing controlled diversity, clipping and multiplicative jitter are applied sequentially, where the clipping operator enforces hard bounds:

$$clip(x, a, b) = min(max(x, a), b). \tag{35}$$

$$\tilde{C}_{ex} = clip\left(C_{ex}^{\star}, C_{ex,min}, C_{ex,max}\right), \tag{36}$$

$$C_{ex} = clip\left(\eta_{ex} \tilde{C}_{ex}, C_{ex,min}, C_{ex,max}\right), \tag{37}$$

Where:

$$\eta_{ex} \sim U\left[\eta_{ex,min}, \eta_{ex,max}\right] \tag{38}$$

is a multiplicative jitter factor drawn uniformly from the interval $[\eta_{ex,min}, \eta_{ex,max}]$. The jitter introduces controlled diversity around the resonance-consistent capacitance value, while the outer clipping ensures the final value remains within the physical bounds $[C_{ex,min}, C_{ex,max}]$. All bound and jitter values are listed in Table 6.

Table 6. Design space, reference parameters, and constraints.

Parameter	Value/Range	Unit
Technology & Frequency		
Input cap. density, κ_{in}	0.50	fF/μm
Output cap. density, $\kappa_{out,p}$	5.47	fF/μm
Reference Design Anchors		
Cascode width, $W_{3,\mathrm{ref}}$	180	μm
Drain inductance, $L_{d,ref}$	4.46	nH
Coupling cap., $C_{ex,ref}$	0.357	pF
Series output cap., $C_{3,ref}$	0.80	pF
Shunt output cap., $C_{4,ref}$	0.36	pF
Parasitic output cap., $C_{par,ref}$	0.626	pF
Continuous Parameters (d_c = 6)		
W_{main}	[101, 400]	μm
W_3	[101, 400]	μm
V_{bias}	[0.45, 0.95]	V
C_{ex}	[0.1, 1.2]	pF
C_3	[0.2, 2.5]	pF
C_4	[0.1, 2.5]	pF
Multiplicative Jitter		
η_{flex}	U([0.90, 1.10])	
η_{cap}	U([0.90, 1.10])	
η_3	U([0.85, 1.15])	
Discrete Inductor Libraries		
L_g (gate)	14 entries, 1.5–10.2	nH
L_s (source)	9 entries, 0.2–1.5	nH
L_d (drain)	12 entries, 1.5–8.0	nH
L_1 (interstage)	9 entries, 2.1–6.5	nH
L_2 (interstage)	9 entries, 1.5–6.1	nH

5.4. Output resonance-guided derivation and snapping (L_d, C_3, C_4)

The output matching network is derived from the cascode common-gate device width W_3 (in μm). A proxy output capacitance model, calibrated from reference device characterization, captures the width-dependent parasitic loading at the drain node:

$$C_{out,proxy}(W_3) = \kappa_{out,proxy}\, W_3, \tag{39}$$

where, $\kappa_{out,proxy}$ (in fF/μm) is the technology-calibrated output capacitance density used for initial inductance sizing.

The corresponding ideal drain inductance is

$$L_d^{\star} = \frac{1}{\omega_0^2\, C_{out,proxy}(W_3)}, \tag{40}$$

where, $\omega_0 = 2\pi f_0$ is the target angular resonance frequency at $f_0 = 2.4$ GHz, as defined in the input resonance derivation.

This ideal value is snapped to the nearest realizable discrete inductor in the drain inductor library L_d:

$$L_d = \underset{L \in L_d}{argmin} |L - L_d^{\star}|. \tag{41}$$

Once L_d is fixed, the explicit output capacitors are derived using a separate parasitic capacitance model calibrated from a reference design:

$$C_{par,out}(W_3) = \kappa_{out}\, W_3, \qquad \kappa_{out} = \frac{C_{par,out,ref}}{W_{3,ref}}, \tag{42}$$

where, $C_{par,out,ref}$ is the parasitic output capacitance measured at the reference device width $W_{3,ref}$, and κ_{out} (in fF/μm) is the resulting linear capacitance density. Note that $\kappa_{out,proxy}$ and κ_{out} serve distinct roles: the former is used for initial L_d sizing, while the latter models the parasitic capacitance subtracted when back-solving for C_4. The resonance-consistent total shunt capacitance corresponding to the snapped L_d is

$$C_{tot,out}^{\star} = \frac{1}{\omega_0^2\, L_d}. \tag{43}$$

The required explicit shunt capacitance is therefore

$$C_4^{\star} = C_{tot,out}^{\star} - C_{par,out}(W_3). \tag{44}$$

Clipping and multiplicative jitter are applied sequentially to enforce practical bounds while introducing controlled diversity:

$$\tilde{C}_4 = clip\left(C_4^{\star},\ C_{4,min},\ C_{4,max}\right), \tag{45}$$

$$C_4 = clip\left(\eta_4\, \tilde{C}_4,\ C_{4,min},\ C_{4,max}\right), \tag{46}$$

Where,

$$\eta_4 \sim U\left[\eta_{4,min},\ \eta_{4,max}\right] \tag{47}$$

is a multiplicative jitter factor drawn uniformly from $\left[\eta_{4,min}, \eta_{4,max}\right]$ to introduce controlled variability around the resonance-derived shunt capacitance. The bounds $C_{4,min}$ and $C_{4,max}$ denote the physical limits

of the shunt capacitance (Table 6). The series coupling capacitor C_3 is assigned using a reference-scaled relation that preserves proportional trends with device width and output shunt capacitance:

$$C_3^{\star} = C_{3,ref}\sqrt{\frac{W_3}{W_{3,ref}}}\sqrt{\frac{max(C_4,\,\epsilon)}{C_{4,ref}}}, \tag{48}$$

where $C_{3,ref}$ is the series coupling capacitance of the reference design, $W_{3,ref}$ is the reference cascode device width, $C_{4,ref}$ is the reference shunt capacitance, and $\epsilon = 10^{-6}$ pF is a small numerical safeguard to ensure positivity under the square root. Finally, multiplicative jitter and a single clipping step are applied:

$$C_3 = clip\left(\eta_3\, C_3^{\star},\ C_{3,min},\ C_{3,max}\right), \tag{49}$$

Where,

$$\eta_3 \sim U\left[\eta_{3,min},\ \eta_{3,max}\right] \tag{50}$$

is a multiplicative jitter factor drawn uniformly from $\left[\eta_{3,min},\ \eta_{3,max}\right]$ to introduce controlled variability in the scaled coupling capacitance. The bounds $C_{3,min}$ and $C_{3,max}$ denote the physical limits of the series coupling capacitance. All reference values, jitter ranges, and capacitance bounds are listed in Table 6. As illustrated in Figure 3, continuous device parameters are sampled via LHS, while discrete inductors are selected from quantized libraries, and dependent reactive components are derived through resonance relations with bounded jitter and snapping.

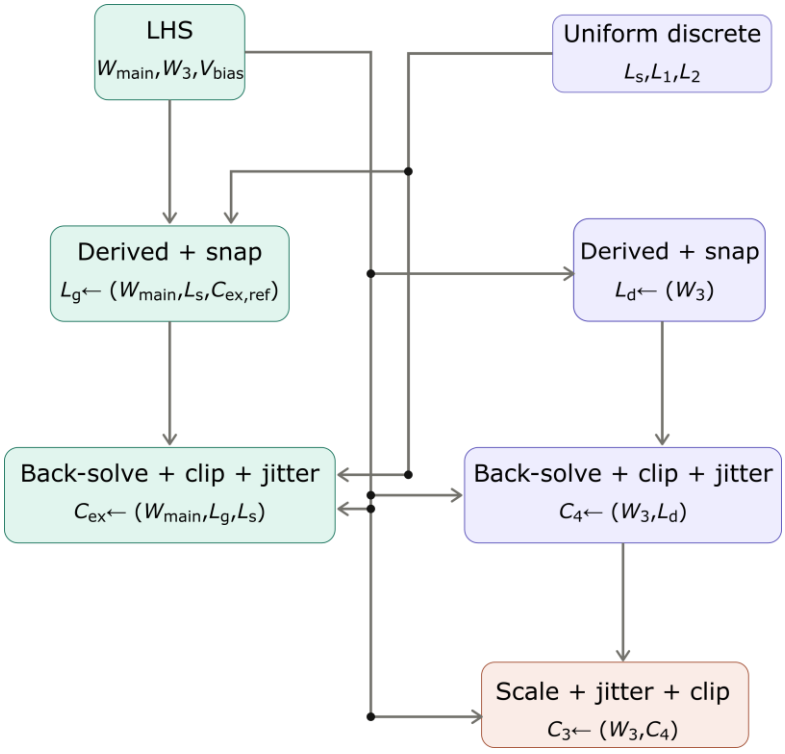

Figure 3. Dependency structure of the physics-guided warm-start sampling.

The jitter ranges were selected according to the sensitivity of each capacitance to resonance detuning. The narrower ±10% ranges for η_{flex} and η_{cap} are used for capacitances that directly affect the input and output resonance conditions, where larger perturbations may shift the resonance away from the 2.4 GHz target band. The wider ±15% range for η_3 is used because the series output coupling capacitance C_3 mainly affects output matching and is less sensitive to resonance-frequency detuning.

These ranges were also checked empirically and found to maintain an initialization yield above 30% while still providing sufficient sample diversity for early surrogate-model training.

6. Constrained MOBO framework

6.1. Discrete-inductor encoding strategy

Since BO process operates in a continuous parameter space, each discrete inductor must be represented by a continuous embedding that maps to a valid library component.

6.1.1. Continuous embedding of discrete inductors

For each library entry $\ell_{t,j}$, a log-descriptor vector is formed:

$$z_{t,j} = \begin{bmatrix} logL_{t,j} \\ logQ_{t,j} \end{bmatrix}. \tag{51}$$

The logarithmic transformation compresses the dynamic range of inductance and quality factor values, yielding a more uniform embedding space. Self-resonant frequency (SRF) data is available in the library but is excluded from the embedding, since at the 2.4 GHz, all inductors have SRF values well above the band of interest, so SRF provides negligible discriminative power while adding optimization dimensions.

Min–max normalization is applied per inductor family to map all entries to the unit hypercube:

$$e_{t,j}^{(k)} = \frac{z_{t,j}^{(k)} - \min_j z_{t,j}^{(k)}}{\max_j z_{t,j}^{(k)} - \min_j z_{t,j}^{(k)}}, \qquad k \in \{L, Q\}. \tag{52}$$

Thus, each inductor family t contributes a two-dimensional continuous encoding:

$$\hat{e}_t = \left(e_{t,L}^{enc},\ e_{t,Q}^{enc}\right) \in [0,1]^2. \tag{53}$$

With five inductor families, the total encoding dimension is $5 \times 2 = 10$, yielding a full optimization vector of dimension $d_c + 10 = 16$ that is entirely continuous.

6.1.2. Discrete realization via nearest-neighbor projection

After the optimizer proposes a continuous encoding $\hat{e}_t$ for each inductor family, the nearest library entry in embedding space is selected:

$$j_t^\star = arg \min_{j \in \{1,\ldots,|L_t|\}} \| \hat{e}_t - e_{t,j} \|_2, \qquad L_t \leftarrow L_{t,j_t^\star}. \tag{54}$$

This defines a projection operator

$$\Pi_{L_t}: [0,1]^2 \to L_t \tag{55}$$

that maps any point in the continuous encoding space to a valid, fabrication-realizable inductor from the EM-characterized library.

6.2. GP modeling

Each scalar output metric $y(x)$, whether an objective (e.g., P_{dc}, NF) or a constraint function (e.g., c_{S21}, c_K) is modeled independently by a GP.

6.2.1. Observation model

The observed output is modeled as a latent function corrupted by Gaussian noise:

$$y(x) = f(x) + \epsilon, \qquad \epsilon \sim N(0, \sigma_n^2), \tag{56}$$

with the GP prior

$$f(x) \sim GP\big(\mu(x),\ k(x, x')\big). \tag{57}$$

6.2.2. Kernel function

An Automatic Relevance Determination (ARD) Matérn-5/2 kernel is employed:

$$k(x, x') = \sigma_f^2 \left(1 + \sqrt{5}\, r + \frac{5}{3}\, r^2\right) exp\big(-\sqrt{5}\, r\big), \tag{58}$$

where the weighted distance is

$$r = \sqrt{\sum_{d=1}^{D} \frac{(x_d - x_d{}')^2}{\ell_d^2}} \tag{59}$$

and $\{\ell_d\}_{d=1}^{D}$ are per-dimension ARD lengthscales. The ARD mechanism enables the GP to automatically down-weight irrelevant or weakly influential design variables, which is critical in this mixed encoding space where inductor embeddings and circuit parameters exhibit different sensitivities.

6.2.3. Posterior prediction

Given a training dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, let $K \in R^{N \times N}$ with $K_{ij} = k\big(x_i, x_j\big)$, and $k_* = [k(x, x_1), \ldots, k(x, x_N)]^\top$. The posterior predictive distribution at a new point x is

$$\mu_N(x) = k_*^\top (K + \sigma_n^2 I)^{-1} y, \tag{10}$$

$$\sigma_N^2(x) = k(x, x) - k_*^\top (K + \sigma_n^2 I)^{-1} k_*. \tag{11}$$

6.2.4. Hyperparameter estimation

The GP hyperparameters $\theta = \{\sigma_f^2, \sigma_n^2, \ell_1, \ldots, \ell_D\}$ are estimated by maximizing the log marginal likelihood:

$$\begin{aligned} logp(y \mid X, \theta) = \quad & -1/2\, y^\top (K_\theta + \sigma_n^2 I)^{-1} y \\ & -1/2\, log|K_\theta + \sigma_n^2 I| - N/2\, log(2\pi). \end{aligned} \tag{62}$$

Surrogates are refitted after each new observation is added to the training set.

6.3. Constraint modeling via probability of feasibility

For each constraint $c_k(x) \leq 0$ modeled by its own GP posterior $c_k(x) \sim N\left(\mu_{c_k}(x), \sigma_{c_k}^2(x)\right)$, the probability of satisfying constraint k at a candidate point x is

$$p_k(x) = P(c_k(x) \le 0) = \Phi\left(\frac{-\mu_{c_k}(x)}{\sigma_{c_k}(x)}\right), \tag{63}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Assuming approximate independence across constraints, the joint probability of feasibility is

$$p_{feas}(x) = \prod_{k=1}^{K} p_k\,(x). \tag{64}$$

6.4. Constrained noisy expected hypervolume improvement

The acquisition function combines the noisy expected hypervolume improvement (qNEHVI) with the probability of feasibility. Let P_T be the feasible Pareto set at iteration T. The hypervolume improvement for a candidate x is

$$HVI(x) = max\big(0,\ HV(P_T \cup \{f(x)\};\, r) - HV(P_T;\, r)\big). \tag{65}$$

The constrained multi-objective acquisition function is

$$\alpha(x) = E[HVI(x)] \cdot p_{feas}(x), \tag{66}$$

where the expectation is taken with respect to the GP posterior. The noisy variant (qNEHVI) accounts for observation noise in the existing Pareto set and supports joint optimization over a batch of q candidates, enabling parallel evaluation. In the implementation, the batch size is set to $q = 3$.

Although the circuit simulations are deterministic, small numerical variations can arise from solver tolerances, convergence thresholds, and nonlinear analysis behavior across different operating regions. In addition, a small learned or regularizing noise term improves the numerical conditioning of the GP model, especially in the mixed continuous-discrete design space. Thus, the noisy formulation is used to avoid treating all observed objective values as exact and to prevent overconfident Pareto-front updates under surrogate uncertainty.

6.5. Validity-gated dataset update

After each simulation, the validity indicator $v(x)$ is evaluated, and designs that fail the check are discarded from surrogate training. The objective and constraint models are updated only with valid observations:

$$D \leftarrow D \cup \big\{\big(x,\, f(x),\, c(x)\big) \;\big|\; v(x) = 1\big\} \tag{67}$$

Thus, D contains only valid samples and their corresponding objective $f(x)$ and constraint $c(x)$ values.

The valid designs that violate performance specifications (e.g., $S_{21} < 19$ dB) are retained in D. The constraint surrogates require both spec-passing and spec-failing observations to learn the feasibility boundary. Invalid designs which violate stability, passivity, or bias-point requirements produce non-physical simulator outputs and are therefore excluded from all GP training sets. Algorithm 1 summarizes the complete optimization flow.

Algorithm 1. Physics-guided multi-objective Bayesian optimization framework.

Input: Continuous bounds $\{[a_k, b_k]\}$, discrete libraries $\{\mathcal{L}_t\}$, constraint thresholds, initial budget N_0, MOBO evaluation budget N, batch size q
Output: Feasible Pareto set $\mathcal{P}$ and selected best designs
Initialize training dataset $\mathcal{D} \leftarrow \emptyset$
— **Physics-Informed Warm Start** —
for $i = 1, \ldots, N_0$ **do**
  Sample $\mathbf{x}_{c,i}$ via LHS over $[\mathbf{a}_c, \mathbf{b}_c]$
  Select L_1, L_2, L_s randomly from libraries
  Compute L_g^* from input resonance; snap $L_g \leftarrow \Pi_{\mathcal{L}_c}(L_g^*)$
  Back-solve C_{ex} from resonance; apply clip/jitter
  Compute $L_d^*(W_3)$; snap $L_d \leftarrow \Pi_{\mathcal{L}_c}(L_d^*)$
  Derive C_3, C_4 from output resonance; apply clip/jitter
  Simulate and record $y(\mathbf{x}_i)$
  Evaluate validity gate $v(\mathbf{x}_i)$
  if $v(\mathbf{x}_i) = 1$ **then**
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\mathbf{x}_i,\ \mathbf{f}(\mathbf{x}_i),\ \mathbf{c}(\mathbf{x}_i))\}$
  else
    Discard from surrogate training (mark failed)
  end if
end for
— **Surrogate Initialization** —
Fit GP surrogates for objectives and constraints on $\mathcal{D}$ via marginal likelihood
—BO **Loop** —
for $n = 1, \ldots, \lceil N/q \rceil$ **do**
  Optimize $\alpha(\mathbf{x}) = \mathbb{E}[\text{HVI}(\mathbf{x})] \cdot p_{\text{feas}}(\mathbf{x})$ to obtain batch $\{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_q\}$
  for each candidate $\hat{\mathbf{x}}_i$ **in** batch **do**
    Snap inductors: $L_t \leftarrow \Pi_{\mathcal{L}_c}(\hat{e}_t)$ for each family t
    Simulate realized design; obtain objectives $\mathbf{f}(\hat{\mathbf{x}}_i)$ and constraints $\mathbf{c}(\hat{\mathbf{x}}_i)$
    Evaluate validity gate $v(\hat{\mathbf{x}}_i)$
    if $v(\hat{\mathbf{x}}_i) = 1$ **then**
      $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\hat{\mathbf{x}}_i,\ \mathbf{f}(\hat{\mathbf{x}}_i),\ \mathbf{c}(\hat{\mathbf{x}}_i))\}$ {includes spec-failing designs}
    else
      Discard from all surrogate training (mark failed)
    end if
  end for
  Refit GP surrogates on updated $\mathcal{D}$
end for
return Feasible Pareto set $\mathcal{P}$ and best designs (min P_{dc}, min NF, max $IIP3$)

7. Experimental setup

The proposed framework was validated on a 2.4 GHz current reuse cascode CMOS LNA implemented in the TSMC 40 nm PDK. The circuit was designed using Cadence Virtuoso and simulated using the Spectre simulator. RF performance metrics, including gain, NF, impedance matching, linearity, and power consumption, were extracted from Spectre simulations for use in the optimization process. The MOBO framework was implemented using the Ax platform [43]. Optimization was performed on a cloud-based CPU environment (Google Colab) with approximately 12.7 GB RAM and completed within standard session limits, demonstrating that the proposed framework does not require specialized high-performance computing resources. Here, 'standard session limits' refers to the maximum uninterrupted runtime provided by a Google Colab Pro subscription (up to 24 hours per session). The full 190-simulation workflow completed within a single overnight session without requiring any session restart or manual intervention. All circuit evaluations were performed using Spectre under TT process corner at 27 ℃ with $V_{DD} = 1.1$ V. A total of 190 simulations were conducted (120 LHS initialization + 70 MOBO evaluations).

The 120 LHS + 70 MOBO budget split was selected as a practical balance between initial coverage and simulation cost. The 120-point LHS budget corresponds to approximately 11 samples per variable for the 11-variable mixed continuous-discrete search space, which lies within the commonly used 10–15 samples-per-variable range for BO initialization. This choice is also supported by Figure 4 (top panel), where the hypervolume increases rapidly during the first 100 LHS evaluations and shows only modest improvement from evaluations 101–120. For the MOBO phase, 70 evaluations were allocated under the practical wall-clock constraint of Spectre simulation. The batch size was set to $q = 3$ as a compromise between runtime and candidate diversity: sequential BO $(q = 1)$would require more frequent GP refitting and acquisition optimization, whereas larger batches may produce correlated candidates in the sparse feasible region. Figure 4 presents an evaluation budget analysis based on the full 190-evaluation simulation log. The upper plot shows the dominated hypervolume, normalized to the final value at evaluation 190, as a function of evaluation count. The hypervolume increases rapidly during the first ~100 LHS evaluations and then shows only limited improvement from evaluations 101–120, indicating that the LHS phase had largely saturated. After the transition to MOBO at evaluation 120, the hypervolume increases again, showing the benefit of guided refinement. The lower plot reports the MOBO hypervolume gain over 10-evaluation windows. Approximately 80% of the total MOBO gain is obtained within the first 30 MOBO evaluations, followed by diminishing returns. These results support the selected budget split: 120 LHS evaluations provide sufficient initial coverage, while 70 MOBO evaluations recover most of the available Pareto-front improvement within the practical simulation budget.

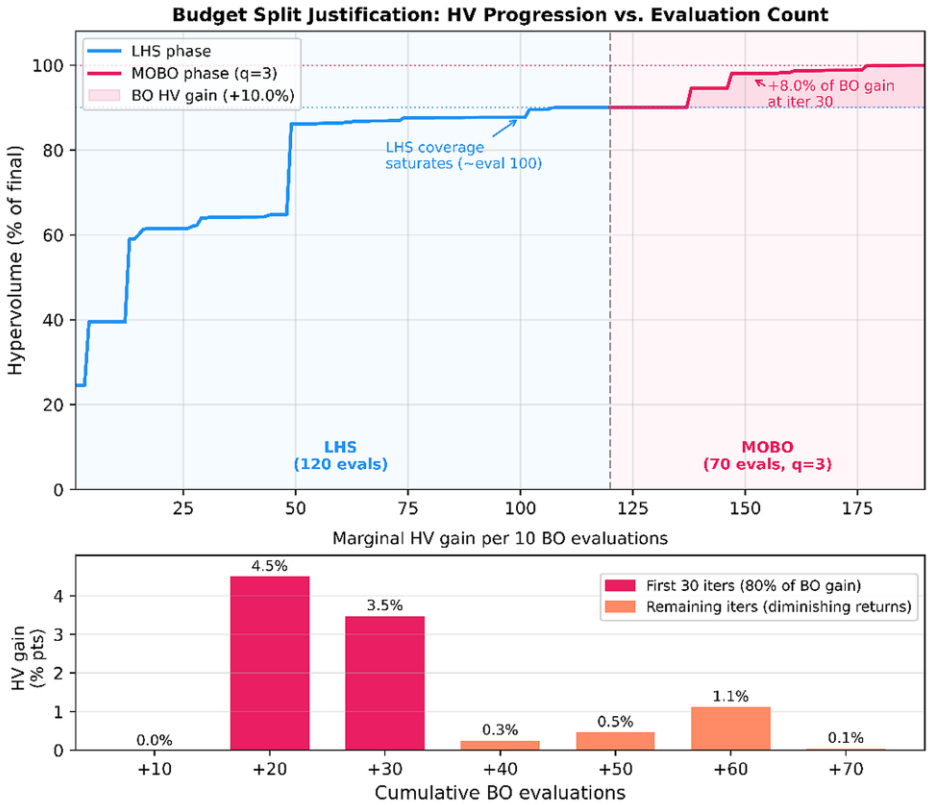

Figure 4. Budget ablation from the 190-evaluation simulation log. Upper: Dominated hypervolume (% of final) *vs.* cumulative evaluation count; the LHS phase saturates after ~100 evaluations, and the MOBO phase (right of dashed line) provides a clear HV improvement (shaded region). Lower: Marginal HV gain per 10-iteration MOBO window; ~80% of the total MOBO gain is captured within the first 30 iterations (dark red bars), confirming diminishing returns thereafter.

7.1. Metric extraction

Simulation-based metric extraction was performed using a structured Cadence Ocean workflow, with each script corresponding to a distinct circuit operating regime as follows.

- DC and small-signal characterization was performed using a DC operating-point analysis followed by S-parameter and noise simulations from 0.1 to 10 GHz over 500 frequency points. This script extracted S_{11}, S_{21}, S_{22}, NF, and DC power consumption P_{dc}. Crucially, it also evaluated the absolute unconditional stability of the circuit by extracting the worst-case Rollett stability factor (global minimum K) and the stability determinant (global maximum $|\Delta|$) across the entire 0.1 to 10 GHz spectrum, thereby rigorously preventing out-of-band low-frequency oscillations during algorithmic structural exploration.
- Weakly nonlinear performance was evaluated using a combined PSS/PAC analysis for IIP3 extraction. A large-signal excitation tone at 2.4 GHz (−50 dBm) drove the PSS analysis to establish the periodic operating point, while a small-signal PAC tone at 2.405 GHz was used to extract the fundamental component and the third-order intermodulation (IM3) sideband. The relatively low drive level kept the circuit in the weakly nonlinear region, where the input-referred third-order intercept point (IIP3) could be reliably obtained from the simulated fundamental and IM3 responses.
- Single-tone Harmonic Balance (HB) power sweep at 2.4 GHz. Input power was swept from −50 dBm to 10 dBm to capture both the small-signal gain plateau and the full transition into gain compression, enabling reliable OP1dB extraction.

The total end-to-end runtime was approximately 16–19 hours. This includes about 2 hours for the 120 LHS simulations, with an average runtime of ∼3 min/evaluation; about 3 hours for the 70 MOBO Spectre evaluations, averaging ∼6 min/evaluation due to full nonlinear analyses for feasible candidates; about 8 hours for GP fitting and qNEHVI acquisition optimization across BO iterations, corresponding to ∼15–20 min/iteration; and about 4–6 hours for one-time setup and integration tasks, including the Spectre—Python interface, Colab environment preparation, and dataset preprocessing.

8. Experimental results

8.1. Design space characteristics

Figure 5 shows sampled designs in the gain (S_{21}) *versus* input return loss (S_{11}) plane. The constraints ($S_{21} \geq 19$ dB, $S_{11} \leq -12$ dB) define a narrow feasible region within a nonlinear, sparse design space. Uniform LHS initialization yields limited feasible samples, with only 57 of 120 trials valid (47.5% valid rate). In contrast, the surrogate-guided acquisition focuses evaluations near the feasible boundary and Pareto-relevant region, achieving a higher validity rate (67/70, 95.7%). Pareto-optimal designs are marked with stars, illustrating the trade-off frontier within the feasible region.

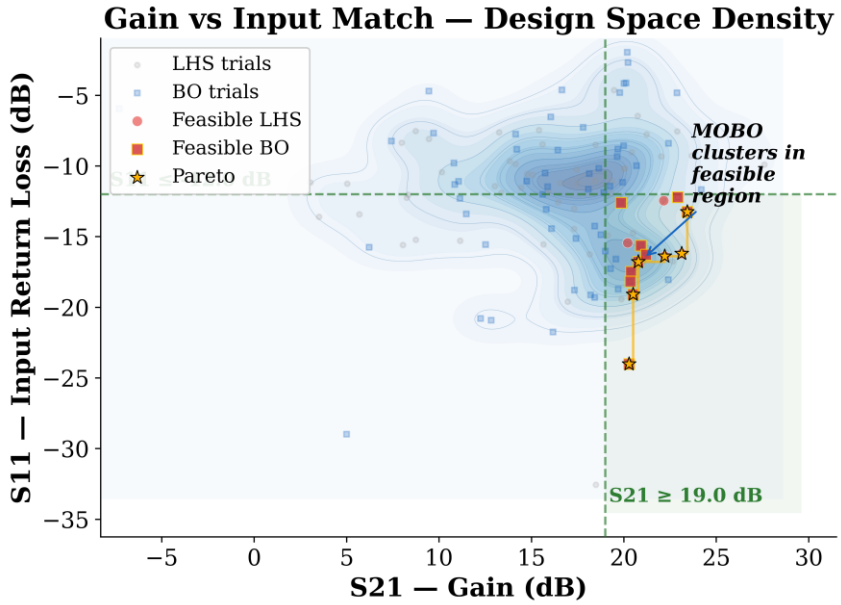

Figure 5. Design-space density visualization in the gain (S_{21}) *versus* input return loss (S_{11}) plane.

8.2. Feasible design discovery rate

Figure 6 compares valid simulation yield during LHS initialization and the subsequent MOBO phase. The blue curve shows rolling yield (window = 20), and the orange curve shows cumulative yield, with the dashed line separating the 120 LHS samples from BO-guided evaluations.

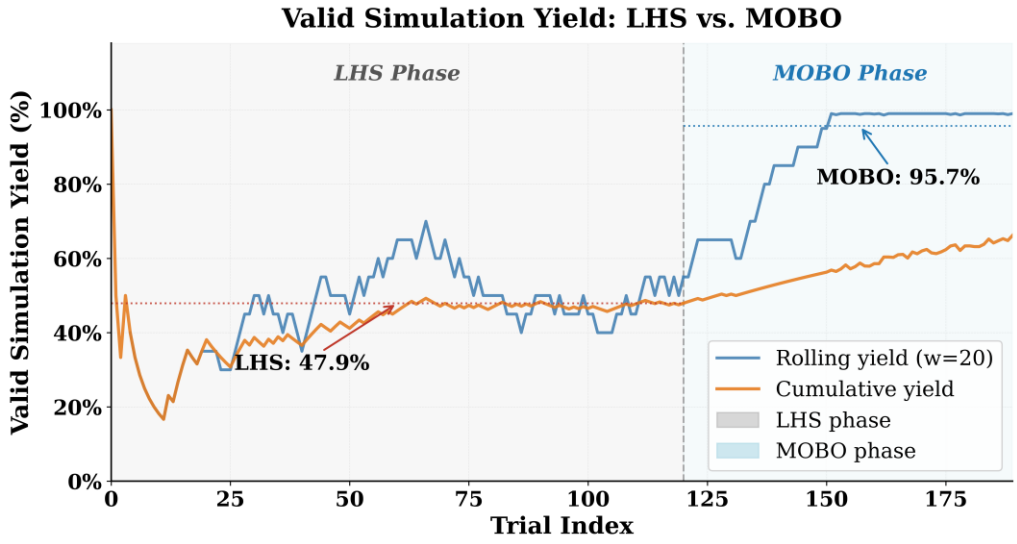

Figure 6. Cumulative valid simulation yield *vs* trial index.

The cumulative yield at trial $n$is defined as

$$Y_{\text{cum}}(n) = (1/n)\sum_{i=1}^{n} v(x_i), \text{ where } v(x_i) = 1 \tag{68}$$

if the i^{th} simulation is valid and $v(x_i) = 0$ otherwise. It represents the running fraction of simulations up to trial $n$that return physically consistent and convergent results. Unlike the rolling yield, which captures local behavior over a sliding window, the cumulative yield provides a stable, noise-resistant measure of overall simulation efficiency across the full run.

During LHS, the rolling yield fluctuates between 30% and 60%, with a cumulative yield of ~47.5%, indicating many invalid samples. In the MOBO phase, the rolling yield quickly approaches 100%, and the cumulative yield rises to 95.7%, demonstrating significantly improved simulation efficiency through validity-gated constrained BO.

During the LHS initialization phase, corresponding to trials 1–120, the 47.5% valid yield reflects the sparsity of the feasible design region. Although the initialization is physics-guided, it cannot ensure that every independently sampled combination of continuous parameters and discrete PDK inductors will satisfy the resonance, matching, and stability requirements.

During the MOBO phase, corresponding to trials 121–190, the remaining ~ 4.3% invalid designs arise mainly from candidates proposed near constraint boundaries, such as the wideband stability limit $K \approx 1.1$or the V_{DS}saturation-margin floor of 70 mV. After discrete inductor snapping and full Spectre evaluation, some of these boundary candidates are classified as invalid.

8.3. Convergence and sample efficiency

Figure 7 shows convergence of the best feasible power P_{dc} *versus* simulation count, under stability, bias, and performance constraints. During LHS, uniform sampling finds an initial best of 7.3 mW, surpassing the manual baseline (14.1 mW) after only 26 simulations. In the MOBO phase, the solution improves further to 5.4 mW, a 62% reduction from the manual design and 1.9 mW lower than the best LHS result. This improvement highlights the effectiveness of surrogate-guided exploration in identifying lower-power feasible regions.

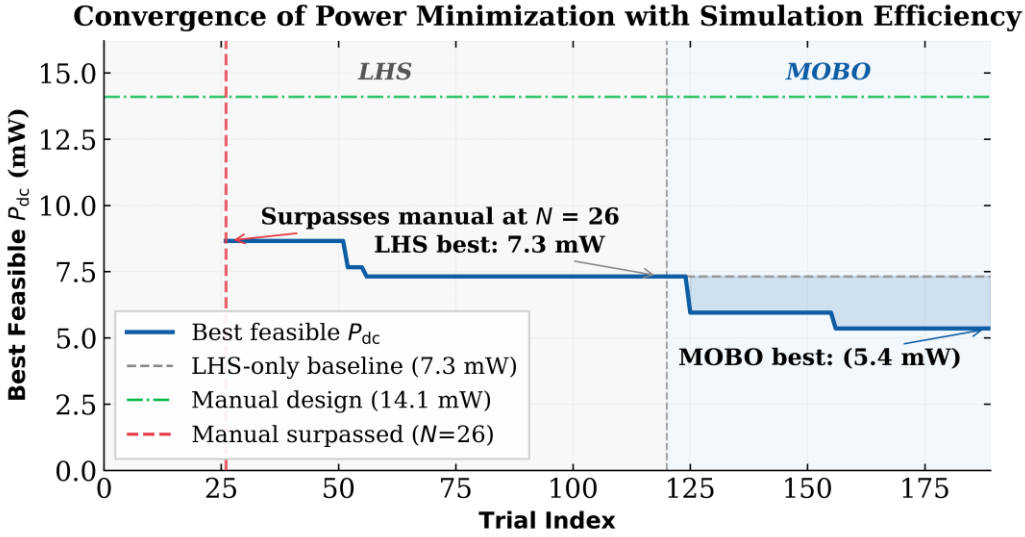

Figure 7. Convergence of best feasible power consumption P_{dc} *versus* number of simulations.

8.4. Multi-objective pareto expansion.

Figure 8 shows the dominated hypervolume (HV) progression over the full 190-evaluation optimization run. The HV is computed in the minimization objective space $(P_{\mathrm{dc}}, NF, -IIP3)$ and measures the objective-space volume dominated by the current Pareto front with respect to a fixed reference point. A larger HV therefore indicates a Pareto front that is both farther from the reference point and more diverse across the objective space, reflecting joint improvement in power, NF, and linearity.

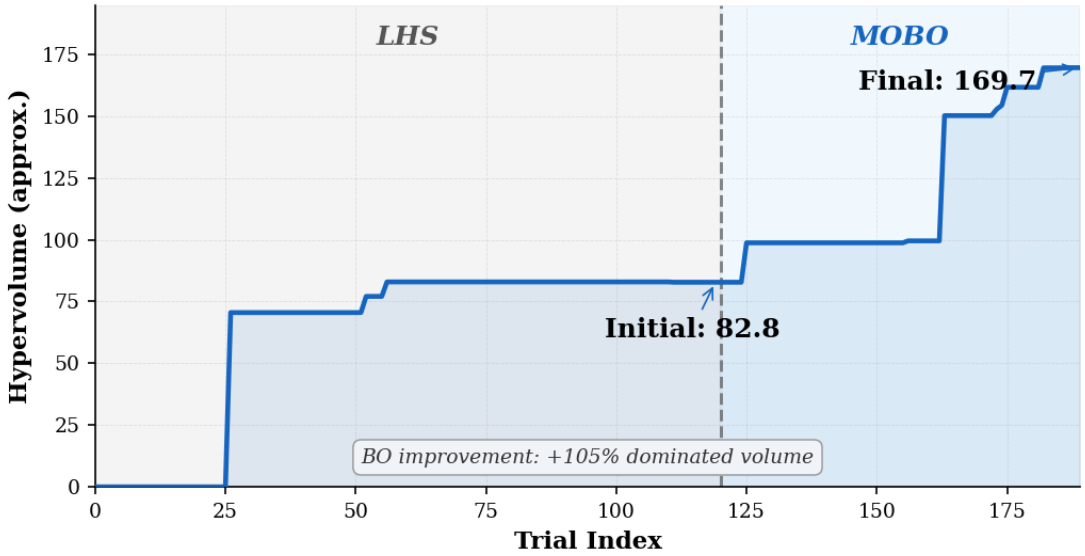

Figure 8. Hypervolume progression for the multi-objective problem $(P_{dc}, NF, -IIP3)$.

The reference point was set to $(P_{\mathrm{dc}}, NF, -IIP3) = (30{,}10{,}20)$, corresponding to $P_{\mathrm{dc}} = 30\,\mathrm{mW}$, $NF = 10\,\mathrm{dB}$, and $IIP3 = -20\,\mathrm{dBm}$. These values represent conservative worst-case bounds outside the feasible design range observed in this study.

During the LHS phase, the HV increases rapidly to 82.8 as initial feasible trade-offs are identified and then begins to saturate as unguided space-filling exploration yields diminishing improvements. After switching to MOBO, the HV further increases to 169.7, corresponding to a 105% improvement over the LHS-only value. This indicates that MOBO more than doubles the dominated objective-space volume by identifying designs with improved trade-offs among lower P_{dc}, lower NF, and higher $IIP3$, compared with LHS alone under the same evaluation budget. Figure 4 (bottom panel) further shows that approximately 80% of the MOBO hypervolume gain is obtained within the first 30 MOBO evaluations, with diminishing returns, thereafter, supporting the chosen budget as a practical stopping point.

8.5. Comparison with manual design

Figure 9 compares the Pareto frontiers obtained after LHS initialization and MOBO refinement, together with the manual expert design. The shaded region in each projection indicates the additional dominated objective space identified by surrogate-guided optimization beyond unguided initialization.

In Figure 9a, the P_{dc}-NF projection shows that MOBO reduces power consumption by up to 62%, from 14.1 mW to 5.4 mW. This point corresponds to the minimum-power MOBO design and reflects a deliberate power-noise trade-off: its NF is 2.19 dB, slightly higher than the manual baseline of 1.98 dB. However, the MOBO Pareto set also contains designs with NF as low as 1.84 dB, showing that NF-competitive solutions are attainable at intermediate power levels.

In Figure 9b, the P_{dc} -IIP3 projection shows a substantial linearity improvement, with IIP3 increasing from $-12.0\,\mathrm{dBm}$ to $-1.3\,\mathrm{dBm}$, corresponding to a $+10.7\,\mathrm{dB}$ improvement while also reducing power. Figure 9c further shows the NF-IIP3 trade-off, where MOBO identifies higher-linearity designs, some with moderately increased NF and others with both improved linearity and lower NF than the manual baseline. Overall, the MOBO Pareto frontier expands beyond the LHS frontier across all projections, confirming improved exploration of the P_{dc}-NF-IIP3 tradeoff space.

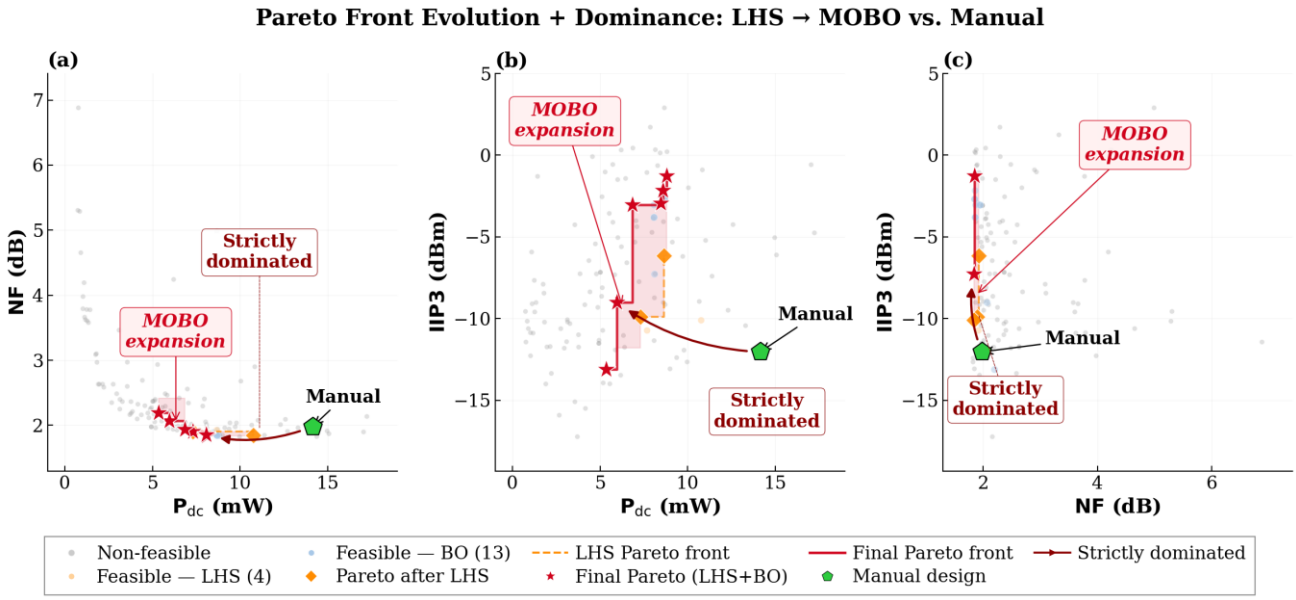

Figure 9. Multi-objective pareto expansion.

Figure 10 provides a three-dimensional visualization of the Pareto surface in the objective space (P_{dc} [mW], NF [dB], IIP3 [dBm]). Spec-feasible designs satisfying $S_{21} \geq 19$ dB, $S_{11} \leq -12$ dB, and $S_{22} \leq -12$ dB are shown in blue (LHS phase) and light red (MOBO phase). The 11 Pareto-optimal designs are highlighted with larger markers, and the convex-hull approximation of the Pareto front is rendered as a translucent pink surface. The gold star marks the manually designed reference, which lies inside the Pareto surface, confirming that MOBO improves the overall Pareto trade-off frontier relative to the manual design.

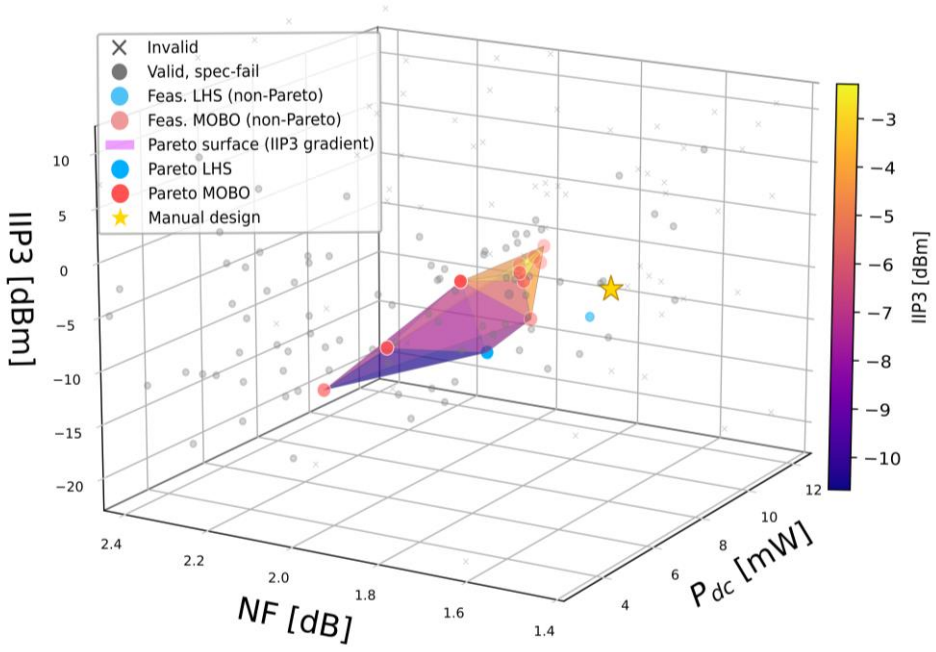

Figure 10. Three-dimensional Pareto surface of the objective space (P_{dc}, NF, IIP3). All 190 simulated designs are shown: invalid designs (gray crosses), valid but spec-failing designs (medium gray), and spec-feasible designs (blue = LHS, light red = MOBO). Pareto-optimal points are highlighted with larger markers (dark blue = LHS, dark red = MOBO). The pink surface is the approximation of the Pareto front. The gold star marks the manual reference design (P_{dc}, NF, IIP3).

Table 7 shows that the selected MOBO Pareto solution achieves the lowest power consumption among the phases compared, reducing P_{dc} from 14.1 mW in the manual baseline to 5.4 mW, corresponding to a 62% reduction. It also improves IIP3 from −12.0 dBm to −1.3 dBm, while incurring a modest NF increase from 1.98 dB to 2.19 dB. This result highlights the power–linearity–noise tradeoff captured by the proposed MOBO framework, rather than representing the independent best value of each metric across the full Pareto front.

Table 8 compares the manual reference design with the best representative feasible designs obtained after the LHS and MOBO phases. Each column corresponds to a single design from the respective phase, so the reported P_{dc}, IIP3, and NF values are jointly achieved by that design. Compared with the manual baseline, the best MOBO design reduces P_{dc} by 62% and improves IIP3 by 10.7 dB, with only a 0.21 dB NF penalty.

Table 7. Performance comparison with open-source approaches.

Framework	# evals	Valid yield	# Feasible	Best feas. P_{dc} (mW)	Best feas. NF (dB)	Best feas. IIP3 (dBm)
Manual (expert baseline)	—	—	1 (manual)	14.10	1.98	−12.00
NSGA-II	190	42.6%	0	—	—	—
GP-EI	190	49.5%	30 (+29)	7.01 (+50.3%)	1.74 (+12.1%)	−6.75 (+5.25 dB)
NGOpt	190	36.8%	1	11.03 (+21.8%)	1.73 (+12.6%)	−13.05 (−1.05 dB)
This work	**190**	**95.7%**	**17 (+16)**	**5.40 (−61.7%)**	**1.84 (−7.1%)**	**−1.30 (+10.70 dB)**

Note: The reported "Best feasible" values are computed independently for each metric among all feasible designs produced by each framework. Therefore, the best P_{dc}, NF, and IIP3 values in each row may correspond to different feasible designs, and should not be interpreted as a single design simultaneously achieving all three values. # denotes a count: # evals is the number of simulations performed, and # Feasible is the number of designs satisfying all constraints.

Table 8. Performance comparison across phases.

Metric	Manual	Best LHS	Best MOBO	Improvement
P_{dc} (mW)	14.1	8.6	5.4	+62%
IIP3 @ 2.4 GHz (dBm)	−12.0	−8.5	−1.3	+10.7 dB
NF @ 2.4 GHz (dB)	1.98	2.15	2.19	−0.21 dB

Note: Each column reports one representative feasible design from the corresponding phase; hence, the listed P_{dc}, IIP3, and NF values are jointly achieved by the same design.

8.6. Comparison with open-source optimization frameworks

To benchmark the proposed framework against widely available black-box optimizers, we conducted a controlled comparison with three representative approaches: Non-dominated Sorting Genetic Algorithm II (NSGA-II) [44], Gaussian-process Expected Improvement (GP-EI) [45], and the Nevergrad Optimizer (NGOpt) auto-portfolio [46,47]. All methods were evaluated using the same mixed continuous–discrete design space, objectives $(-P_{dc}, -NF, +IIP3)$, validity gates, Table 3 constraints, Spectre backend, 190-simulation evaluation budget, and random seed.

The three open-source baselines were selected to cover representative algorithmic families commonly used in black-box circuit optimization. NSGA-II is an evolutionary multi-objective optimization method that maintains a population of candidates and ranks them using Pareto dominance and crowding distance. The scikit-optimize baseline implements GP-EI with random Tchebycheff scalarization [48], where a Gaussian-process (GP) surrogate is fitted to a randomly weighted Tchebycheff aggregation of the three objectives and the next design is selected by maximizing single-objective EI. NGOpt is an adaptive gradient-free auto-portfolio that selects among optimizers such as DE, Covariance Matrix Adaptation Evolution Strategy (CMA-ES), PSO, and $(1+1)$-Evolution Strategy (ES) depending on the dimensionality and remaining budget. Together, these baselines span evolutionary multi-objective search, scalarized surrogate-based BO, and adaptive gradient-free portfolio optimization.

As summarized in Table 7, the proposed method achieves a valid simulation yield of 95.7%, approximately twice that of the strongest open-source baseline, scikit-optimize, at 49.5%. It also identifies 17 fully feasible designs, whereas NGOpt finds one at most. Among strictly feasible solutions, the proposed framework attains the lowest P_{dc} of 5.40 mW and the best IIP3 of −1.30 dBm. Although the open-source baselines obtain slightly lower NF values of 1.73–1.74 dB, these solutions incur worse P_{dc} and IIP3, indicating a scalarization-induced trade-off rather than a Pareto improvement.

As summarized in Table 8, the proposed MOBO framework substantially improves the LNA performance over the manual baseline, reducing P_{dc} from 14.1 mW to 5.4 mW, corresponding to a 62% power reduction, while improving IIP3 by 10.7 dB from −12.0 dBm to −1.3 dBm. This improvement is achieved with only a small NF penalty of 0.21 dB.

These results show that, under limited Spectre budgets, domain-aware constrained MOBO provides a clear advantage over general-purpose optimizers for mixed-integer multi-objective RFIC design. The improvement is mainly due to resonance-aware initialization, joint objective–constraint GP modeling, and feasibility-weighted qNEHVI-based hypervolume search.

8.7. Generalization to a second design case (different topology and technology node)

To assess whether the proposed framework generalizes beyond the single example studied above, we applied the identical pipeline to a second, structurally different LNA: a 2.4 GHz resistive shunt-feedback LNA implemented in a 55 nm SOI CMOS process and operating at VDD = 1.2 V. This case differs from the primary cascode current-reuse design in circuit topology, technology node, and nominal operating point, and therefore exercises the method under conditions for which it was not tuned.

The validation circuit, shown in Figure 11, is adapted from the 2.4 GHz shunt-feedback CMOS LNA of Aditi and Bansal [49] but was independently redesigned and simulated in this work in the XMC 55-nm RF-SOI process to establish a new baseline. It adopts a two-stage cascaded configuration: the first stage is a cascode amplifier (M1–M2) with a Π-type input-matching network (L_1, C_1, C_2) and a resistive shunt-feedback path (R_1), and the second stage is a common-source amplifier (M3) independently biased through resistor R2. Both stages are loaded with tuned LC resonant tanks (L_3–C_3 and L_4–C_4) that set the operating band.

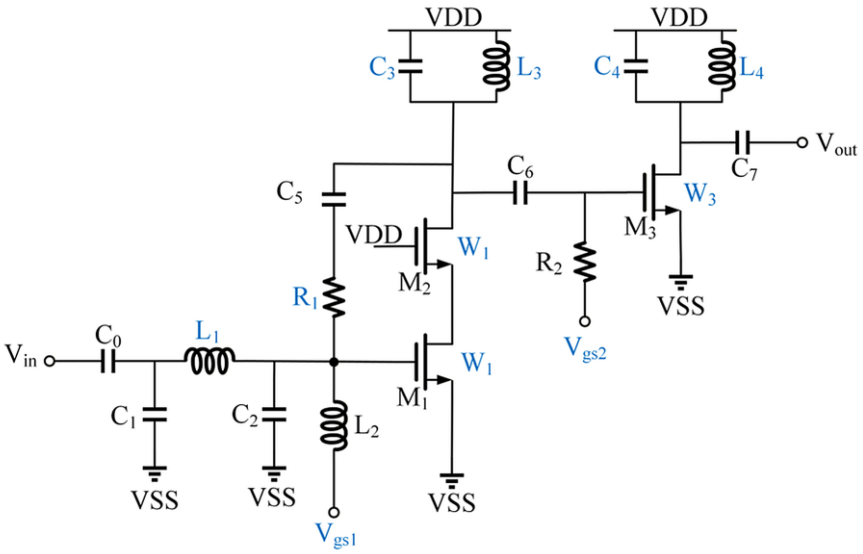

Figure 11. Circuit topology of the 2.4 GHz two-stage resistive shunt-feedback LNA (XMC 55-nm RF-SOI), the second validation case, from modified circuit topology based on [49].

The discrete inductor library was re-characterized for the SOI process inductor model, and the resonance-guided initialization was re-derived for the new matching network (a Π-type input match, resistive shunt feedback, and an LC output) instead of the inductively degenerated input of the primary design. The shunt-feedback resistor R1 was introduced as an additional continuous variable, yielding a 13-dimensional mixed space (i.e., R_1, V_{gs1}, W_1, V_{gs2}, W_3, C_3, C_4 as the seven continuous variables, and L_1/L_3/L_4 each encoded as normalized log L and log Q for the six discrete dimensions). The objectives (minimize P_{dc}, minimize NF, maximize IIP3), the validity gate, and the feasibility-weighted qNEHVI acquisition were left unchanged; the performance constraints were set to $S_{21} \geq 18$ dB, $S_{11} \leq -10$ dB, and $S_{22} \leq -10$ dB for this topology. The design was optimized under a 150-simulation Spectre budget (120 physics-guided LHS samples followed by 30 MOBO evaluations in batches of three).

As summarized in Table 9, the framework again achieved a high valid-simulation yield of 89.3% and discovered 15 spec-feasible designs forming an 11-point Pareto front. Relative to the expert manual baseline (P_{dc} = 18.56 mW, NF = 3.385 dB, IIP3 = −0.64 dBm), the optimizer reduced power to 13.92 mW (−25.0%), lowered NF to 3.03 dB (−0.36 dB), and improved IIP3 to +1.95 dBm (+2.59 dB). Obtained on a different topology, technology node, and supply voltage with no change to the optimization algorithm, these results indicate that the proposed physics-guided, validity-gated MOBO framework transfers across design cases, directly addressing the concern that a single example may not by itself establish generality.

Table 9. Performance comparison for the second design case (XMC 55 nm SOI shunt-feedback LNA).

Metric	Manual	Best feasible (MOBO)	Improvement
P_{dc} (mW)	18.56	13.92	−25.0%
NF @ 2.4 GHz (dB)	3.385	3.026	−0.36 dB
IIP3 @ 2.4 GHz (dBm)	−0.64	+1.95	+2.59 dB

Note: Best-feasible values are computed independently for each objective among all 15 feasible designs (as in Table 7) and need not correspond to a single design. The manual baseline additionally satisfies S11 = −12.5 dB, S21 = 21 dB, and S22 = −16.1 dB.

Table 10 further reports a single representative Pareto design, illustrating a jointly achievable operating point rather than the per-objective bests of Table 9. Relative to the manual baseline, this design lowers P_{dc} by 9.3% and improves IIP3 by 2.59 dB while satisfying all matching and stability constraints ($S11 \leq -10$ dB, $S21 \geq 18$ dB, $S22 \leq -10$ dB), at the cost of a 0.65 dB NF increase. This mirrors the power–linearity gain with a modest noise penalty observed in the primary design case, confirming that the same trade-off behaviour is reproduced on a different topology and technology node.

Table 10. Best representative design versus manual baseline (second design case).

Metric	Manual	Best MOBO	Change
P_{dc} (mW)	18.56	16.84	−9.3%
NF @ 2.4 GHz (dB)	3.385	4.03	+0.65 dB
IIP3 @ 2.4 GHz (dBm)	−0.64	+1.95	+2.59 dB
S_{11} @ 2.4 GHz (dB)	−12.5	−10.49	≤ -10 (met)
S_{21} @ 2.4 GHz (dB)	21.0	19.91	≥ 18 (met)
S_{22} @ 2.4 GHz (dB)	−16.1	−15.21	≤ -10 (met)
OP1dB (dBm)	6.13	6.79	+0.66 dB

Note: The MOBO column reports one representative feasible design; hence the listed P_{dc}, NF, IIP3, and S-parameter values are jointly achieved by the same design.

Figure 12 visualizes these trade-offs by projecting the feasible designs onto the noise-figure axis. Panel (a) shows the characteristic power–noise knee, in which lowering NF below roughly 3.1 dB requires a sharp increase in power, whereas panel (b) shows the noise–linearity trade-off. In both projections the manual baseline (green star) lies on the dominated side of the discovered Pareto frontier: designs exist with simultaneously lower NF and lower power in (a), and with simultaneously lower NF and higher IIP3 in (b), visually corroborating the pairwise improvements reported in Tables 9 and 10.

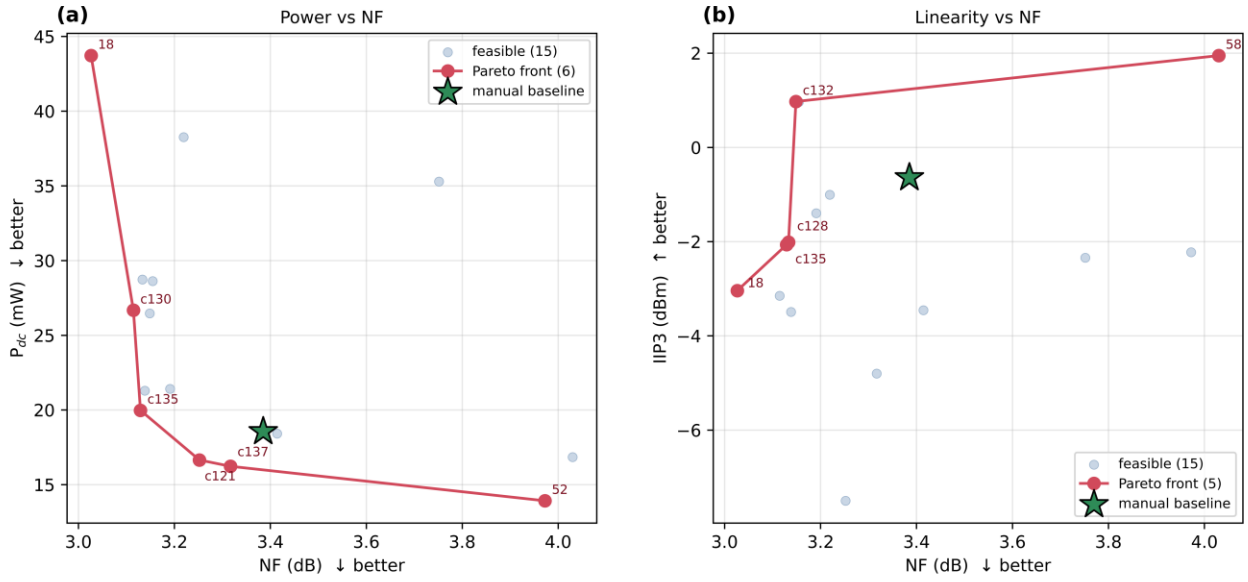

Figure 12. Pareto fronts for the second design case (XMC 55 nm SOI shunt-feedback LNA), projected against noise figure with a shared NF axis: **(a)** power *versus* NF and **(b)** IIP3 *vs* NF. Light markers denote the 15 spec-feasible designs, the red curve is the two-dimensional Pareto frontier (annotated by design ID), and the green star marks the manual baseline. In both projections the manual design is dominated by the discovered frontier.

9. Limitations

Despite the improvements demonstrated, several limitations should be acknowledged. The present work is scoped as a proof-of-concept study, with the primary contribution being methodological, demonstrating that physics-guided initialization, discrete PDK enforcement, and validity-gated MOBO can be jointly integrated into a unified optimization pipeline. The framework is demonstrated on two schematic-level LNA topologies; a 2.4 GHz cascode current-reuse design in TSMC 40 nm and a 2.4 GHz resistive shunt-feedback design in 55 nm SOI (Section 8.7), both under nominal TT, 27 ℃ conditions and without post-layout extraction or PVT analysis. Broader validation across additional circuits, technology nodes, layout-extracted designs, and multi-corner operating conditions is therefore an important direction for future work.

A full PVT sweep would substantially increase the simulation budget, while post-layout extraction may introduce layout-dependent parasitic effects that shift inductance-sensitive metrics such as NF and S_{11}. In addition, the current physics-guided initialization relies on first-order resonance and capacitance models, which may require recalibration for topologies where higher-order parasitic coupling or nonlinear effects dominate feasibility.

The validity-gating concept is extensible to nonlinear RF blocks such as PAs, mixers, and VCOs; however, the validity criteria must be redesigned for each circuit class. For example, nonlinear circuits may require gates based on HB convergence, steady-state oscillation, or large-signal stability rather than the small-signal stability and bias-margin criteria used in this LNA example. Defining such simulation-extractable validity criteria is a key direction for future work.

Finally, real-world imperfections such as process variation, compact-model mismatch, and layout-dependent parasitics may shift circuit performance relative to nominal schematic simulations. However, the proposed MOBO framework can naturally support uncertainty-aware design by identifying diverse Pareto-optimal solutions and, in future work, by incorporating multi-corner constraints directly into the optimization loop.

10. Conclusion

This work presented a physics-guided dataset generation and validity-gated MOBO framework for CMOS LNA design under discrete PDK constraints. The key novelty lies in the joint integration of resonance-guided initialization, strict discrete PDK library enforcement, and validity-gated surrogate training within a unified optimization pipeline, addressing limitations of existing approaches that treat these aspects independently. By explicitly modeling the mixed discrete–continuous design space and embedding resonance-informed initialization, the proposed approach addresses two central challenges in RF design automation: sparse feasibility and discontinuous objective landscapes induced by fabrication libraries. Validation on a 2.4 GHz current reuse cascode CMOS LNA implemented in the TSMC 40 nm PDK and evaluated using Cadence Spectre demonstrates a 62% reduction in power consumption and a +10.7 dB improvement in IIP3 @ 2.4 GHz relative to a handcrafted expert baseline, with a modest NF increase in the minimum-power design (2.19 dB *vs* 1.98 dB for the manual baseline), consistent with the inherent P_{dc}–NF–IIP3 Pareto tradeoff. NF-optimized Pareto solutions achieve NF as low as 1.84 dB. The dominated hypervolume expands by 105% compared to LHS initialization alone, and the valid simulation yield increases from 47.5% to 95.7% under guided optimization. These results indicate that

physics-aware BO can systematically extend the attainable Pareto frontier under realistic fabrication constraints, providing a practical pathway toward automated analog/RF synthesis in industrial PDK environments. Future work will extend the framework to diverse RF topologies and process nodes, incorporate layout-aware parasitic modeling with multi-corner evaluation, and investigate transfer learning to reduce sample complexity by lowering the number of required simulations.

Data availability statement

The data generated and analyzed during this study, including the simulation results used to produce the figures and tables, are available from the corresponding author upon reasonable request. The circuit design files and technology-specific models are subject to confidentiality and licensing restrictions and therefore cannot be made publicly available.

Declaration of generative AI and AI-assisted technologies

The authors used Prism (OpenAI) and Grammarly (Grammarly, Inc.) solely for language editing to improve clarity and grammar. The authors are fully responsible for the technical content, experimental design, analysis, and conclusions of this work.

Authors' contribution

Conceptualization, K.S.Y., J.J. and S.J.; methodology, J.J. and S.J.; software, J.J and S.J; validation, J.J., S.J. and K.S.Y.; formal analysis, K.S.Y., J.J. and S.J.; resources, K.S.Y.; data curation, J.J.; writing—original draft preparation J.J., S.J., N.M., B.K.T., B.M., and K.S.Y.; writing—review and editing, J.J., S.J., N.M., B.K.T., B.M., and K.S.Y.; visualization, J.J.; supervision, K.S.Y.; project administration, K.S.Y. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

Kiat Seng Yeo holds the position of Editor-in-Chief for *Interdiscipline* and has not peer reviewed or made any editorial decisions for this paper. The authors declare no conflicts of interest.

References

[1] Rayas-Sánchez JE, Zhang QJ, Rautio JC, Nikolova NK, Boria VE, *et al.*, Microwave modeling and design optimization: the legacy of john bandler. *IEEE Trans. Microw. Theory Tech.* 2024, 73(1):87–101.

[2] He B, Zhang S, Wang Y, Gao T, Yang F, *et al*. A batched bayesian optimization approach for analog circuit synthesis via multi-fidelity modeling. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst*. 2022, 42(2):347–359.

[3] Jung S, Choi Y, Kim J. Variability-aware, discrete optimization for analog circuits. In *Proceedings of the 49th Annual Design Automation Conference*, San Francisco, USA, 2012, June 3–7, 2012, pp. 536–541.

[4] Chen C, Wang H, Song X, Liang F, Wu K, *et al*. High-dimensional bayesian optimization for analog integrated circuit sizing based on dropout and gm/ID methodology. *Comput.-Aided Des. Integr. Circuits Syst*. 2022, 41(1):4808–4820.

[5] Wei PH, Murmann B. Analog and mixed-signal layout automation using digital place-and-route tools. *IEEE Trans. Very Large Scale Integr. Syst.* 2021, 29(11):1838–1849.

[6] Yang KE, Tsai CY, Shen HH, Chiang CF, Tsai FM, *et al.* Trust-region method with deep reinforcement learning in analog design space exploration. In *Proceedings of the 2021 58th ACM/IEEE Design Automation Conference (DAC)*, San Francisco, USA, December 5–9, 2021, pp. 1225–1230.

[7] Cao W, Gao J, Ma T, Ma R, Benosman M, *et al*. RoSE-opt: robust and efficient analog circuit parameter optimization with knowledge-infused reinforcement learning. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 2025, 44(2):627–640.

[8] Tarkhan M, Sawan M. A novel current density based design approach of low-noise amplifiers. *IEEE Access* 2022, 10:42309–42320.

[9] Lin H, Gao L, Li H, Xu J, Zhang X. A 23.6–46.5 GHz LNA with 3 dB NF and 24 dB gain tuning range in 28-nm cmos technology. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2024, 71(1):29–39.

[10] Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N. Taking the human out of the loop: a review of bayesian optimization. *Proc. IEEE* 2016, 104(1):148–175.

[11] Touloupas K, Sotiriadis PP. LoCoMOBO: a local constrained multiobjective bayesian optimization for analog circuit sizing. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 2022, 41(9):2780–2793.

[12] Touloupas K, Sotiriadis PP. Mixed-variable bayesian optimization for analog circuit sizing using variational autoencoders. In *Proceedings of the 18th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, Villasimius, Italy, June 12–15, 2022, pp. 1–4.

[13] Rutenbar RA, Gielen GGE, Roychowdhury J. Hierarchical modeling, optimization, and synthesis for system-level analog and RF designs. *Proc. IEEE* 2007, 95(3):640–669.

[14] Mina R, Jabbour C, Sakr GE. A review of machine learning techniques in analog integrated circuit design automation. *Electronics* 2022, 11(3):435.

[15] Guven I, Parlak M, Lederer D, De Vleeschouwer C. AI-driven integrated circuit design: a survey of techniques, challenges, and opportunities. *IEEE Access* 2025, 13:167364–167389.

[16] Beaulieu PO, Dumesnil E, Nabki F, Boukadoum M. Analog RF circuit sizing by a cascade of shallow neural networks. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 2023, 42(12):4391–4401.

[17] Subburaman B, Thangaraj V, Balu V, Pandyan UM, Kulkarni J. Artificial neural network modeling of a cmos differential low-noise amplifier using the bayesian regularization algorithm. *Sensors* 2023, 23(21):8790.

[18] Wang H, Wang K, Yang J, Shen L, Sun N, *et al.* GCN-RL circuit designer: transferable transistor sizing with graph neural networks and reinforcement learning. In *Proceedings of the 2020 57th ACM/IEEE Design Automation Conference (DAC)*, Online, July 20–24, 2020, pp. 1–6.

[19] Shila KA. Computationally efficient design of an LNA input matching network using automatic differentiation. *IEEE J. Microwaves* 2025, 5:972–982.

[20] Koziel S, Pietrenko-Dabrowska A. Rapid surrogate-aided multicriterial optimization of compact microwave passives employing machine learning and ANNs. *IEEE Trans. Microwave Theory Tech*. 2024, 72(8):4475–4488.

[21] Yu Y, Zhang Z, Cheng Q, Liu B, Wang Y, *et al.* State-of-the-art: AI-assisted surrogate modeling and optimization for microwave filters. *IEEE Trans. Microwave Theory Tech.* 2022, 70(1):4635–4651.

[22] Sun Y, Benosman M, Ma R. GaN distributed RF power amplifer automation design with deep reinforcement learning. In *Proceedings of the 2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, Incheon, Republic of Korea, June 13–15, 2022, pp. 54–57.

[23] Choi Y, Park S, Choi M, Lee K, Kang S. MA-Opt: reinforcement learning-based analog circuit optimization using multi-actors. *IEEE Trans. Circuits Syst. I Regul. Pap*. 2024, 71(5):2045–2056.

[24] Wang S, Liu Y, Zhang J, Lan M, Jin K, *et al.* Improved reinforcement learning-based design of impedance matching networks for RFIC applications. In *Proceedings of the 2025 IEEE MTT-S International Wireless Symposium (IWS)*, Xi'an, China, May 19–22, 2025, pp. 1–3.

[25] Li Z, Carusone AC. An open-source ams circuit optimization framework based on reinforcement learning—from specifications to layouts. *IEEE Access* 2024, 12:150032–150045.

[26] Oh Y, Kim D, Lee YH, Hwang B. CRONuS: circuit rapid optimization with neural simulator. In *Proceedings of the 2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Valencia, Spain, March 25–27, 2024, pp. 1–6.

[27] Oh S, Kim HJ. Sample-efficient reinforcement learning for analog circuit optimization with intrinsic reward. *J. Semicond. Technol. Sci*. 2025, 25(5):469–475.

[28] Shreeharsha KG, Siddharth RK, Vasantha MH, Kumar YBN. An error bound particle swarm optimization for analog circuit sizing. *IEEE Access* 2024, 12:50126–50136.

[29] Wu S, Li X, Tan T, Huang Z. Optimized design of low-noise amplifier based on behavior-level parameter modeling. In *Proceedings of the 2023 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, Qingdao, China, May 14–17, 2023, pp. 1–3.

[30] Gao X, Zhang H, Ye S, Liu M, Pan DZ, *et al.* Post-layout simulation driven analog circuit sizing. *Sci. China Inf. Sci*. 2024, 67(4):142401.

[31] Caponetto R, Fortuna L, Graziani S, Xibilia MG. Genetic algorithms and applications in system engineering: a survey. *Trans. Inst. Meas. Control*. 1993, 15(3):143–156.

[32] Lyu W, Xue P, Yang F, Yan C, Hong Z, *et al.* An efficient bayesian optimization approach for automated optimization of analog circuits. *IEEE Trans. Circuits Syst. I Regul. Pap*. 2018, 65(6):1954–1967.

[33] Touloupas K, Chouridis N, Sotiriadis PP. Local bayesian optimization for analog circuit sizing. In *Proceedins of the 2021 58th ACM/IEEE Design Automation Conference (DAC)*, San Francisco, USA, December 5–9, 2021, pp. 1237–1242.

[34] Huang J, Tao C, Yang F, Yan C, Zhou D, *et al.* Bayesian optimization approach for rf circuit synthesis via multitask neural network enhanced gaussian process. *IEEE Trans. Microwave Theory Tech.* 2022, 70(11):4787–4795.

[35] Garbuglia F, Spina D, Deschrijver D, Couckuyt I, Dhaene T. Bayesian optimization for microwave devices using deep gp spectral surrogate models. *IEEE Trans. Microwave Theory Tech.* 2023, 71(16):2311–2318.

[36] Budak AF, Bhansali P, Liu B, Sun N, Pan DZ, *et al*. DNN-Opt: an RL inspired optimization for analog circuit sizing using deep neural networks. *arXiv* 2021, arXiv:2110.00211.

[37] Lyu W, Yang F, Yan C, Zhou D, Zeng X. Multi-objective Bayesian optimization for analog/RF circuit synthesis. In *Proceedings of the 55th Annual Design Automation Conference*, San Francisco, USA, June 24–28, 2018, pp. 1–6.

[38] Li Y, Wang Y, Li Y, Zhou R, Lin Z. An artificial neural network assisted optimization system for analog design space exploration. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 2020, 39(10):2640–2653.

[39] Somayaji NSK, Hu H, Li P. Prioritized reinforcement learning for analog circuit optimization with design knowledge. In *Proceedings of the 58th ACM/IEEE Design Automation Conference (DAC)*, San Francisco, USA, December 5–9, 2021, pp. 1231–1236.

[40] Budak AF, Gandara M, Shi W, Pan DZ, Sun N, *et al*. An efficient analog circuit sizing method based on machine learning assisted global optimization. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 2022, 41(5):1209–1221.

[41] Radic J, Djugova A, Videnovic-Misic M. A 2.4 GHz high-gain low noise amplifier. In *Proceedings of the International Symposium on Signals, Circuits and Systems*, Iași, Romania, July 9–10, 2009, pp. 1–4.

[42] Nguyen TK, Kim CH, Ihm GJ, Yang MS, Lee SG. CMOS low noise amplifier design optimization technique. *IEEE Trans. Microwave Theory Tech.* 2004, 52(5):1433–1442

[43] Olson M, Santorella E, Tiao LC, Cakmak S, Garrard M, *et al.* Ax: a platform for adaptive experimentation. In *Proceedings of the AutoML 2025 ABCD Track*, New York, USA, September. 8–11, 2025.

[44] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 2002, 6(2):182–197.

[45] Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* 1998, 13(4):455–492.

[46] Rapin J, Teytaud O. Nevergrad-a gradient-free optimization platform. 2018. Available: https://github.com/facebookresearch/nevergrad (accessed 25 January 2026).

[47] Meunier L, Rakotoarison H, Wong PK, Rozière B, Rapin J, *et al.* Black-box optimization revisited: Improving algorithm selection wizards through massive benchmarking. *IEEE Trans. Evol. Comput.* 2022, 26(3):490–500.

[48] Knowles J. ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Trans. Evol. Comput.* 2006, 10(1):50–66.

[49] Aditi, Bansal M. High linearity and low noise shunt resistive feedback CMOS LNA in 2.4 GHz ISM band. In *Proceedings of the 2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE)*, Noida, India, October 26–27, 2017, pp. 95–99.