Received: 28 May 2023 | Revised: 19 August 2023 | Accepted: 13 September 2023

DOI: 10.1002/cta.3825

WILEY

ORIGINAL PAPER

Multistage Op Amp design space exploration by g_m/I_D sampling and symbolic design equations

Qixu Xie | **Guoyong Shi**

Department of Micro/Nano Electronics, School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

Correspondence
Guoyong Shi, Department of Micro/Nano Electronics, School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China.
Email: shiguoyong@sjtu.edu.cn

Funding information
Natural Science Foundation of China (NSFC) under grant number 61974087, Hisilicon Semiconductor Co. Ltd., Shenzhen, China, under contract (2022–2024) TC20211101667 and TC20230629049

Summary

Multiple-stage complementary metal-oxide-semiconductor (CMOS) operational amplifiers (Op Amps) have been extensively studied in the literature; often, each new design reports several performance advancements over the existing ones. However, the design space boundaries of those new proposals were rarely explored. Predicting the design space boundary of a new circuit topology is usually challenging due to the dependency on many aspects of technical consideration regarding the process technology, sizing priority adopted in practice, and balancing of performance metrics on power, area, noise immunity, and so on. Knowing the design space boundary could be beneficial to a fair appraisal of new designs. In this paper, we propose a gm/ID-based method for the design space exploration of multistage Op Amps. By introducing a sampling technique while taking advantage of the device grouping property with gm/ID, it is possible to explore the circuit performance boundary by combining approximate symbolic modeling and gm/ID table lookup. Due to the introduction of a fast performance evaluation method, a large ensemble of circuit sizing samples can be evaluated in parallel, by which an efficient data mining procedure could be incorporated to deduce the circuit samples that can achieve extremal performance at Pareto boundaries. The fidelity of this proposal has been validated via simulation program with integrated circuit emphasis (SPICE) . The sampling method is also compared with a multi-objective genetic algorithm to show the superiority in capturing Pareto boundaries.

KEYWORDS

analog integrated circuit (IC), gm/ID method, Latin hypercube sampling (LHS), operational amplifier (Op Amp), symbolic design equation

1 | INTRODUCTION

Operational amplifiers (Op Amps) are ubiquitous in modern electronic systems; their design details could play a crucial role in determining the overall analog system performance. However, Op Amp design today still lacks well-developed automation tools. Any kind of circuit topology innovation would take a great deal of design experience and profound appreciation of design trade-offs. Since Op Amps are a representative class of integrated analog circuit modules, design

automation of this class of circuits often becomes the first testbed of other general automation ideas in the analog domain.

Op Amp design is traditionally a "pencil and paper" work field, highly relying on the designer's expertise, experience, and hand analytics. Designers must be well-trained and skillful in the following: (1) topological manipulation of circuit (simplifying, macromodeling, and quick performance anticipation), (2) deriving design equations (both simple and complex ones), (3) drafting sizing strategies, and (4) conducting SPICE simulation to validate the design intuitions. Many authors have published their own reference design procedures for the class of multistage Op Amps.[1–3] However, those procedures have not been accepted as universal design methods so far. Whenever limitations of prior design tactics need to be overcome, it is often the deep analytical facts that are most relevant for innovation, making the pursuit of superior bandwidth and drivability possible.[4,5] As far as circuit performance is concerned, the design space boundary of an existing design is usually most convincing when a new design is invented. However, it is also most difficult to obtain that. Consequently, most of the time why a new circuit design turns out to be superior is not well-justified in reality. When working on design space, precise simulation is actually not the most important. What we need is design space exploration methods that could run fast but maintain accuracy in capturing the performance boundaries with good fidelity. This type of design aid is pursued in this paper.

In the early years, quite a portion of research effort was directed toward the so-called *knowledge-based* approach to analog design automation, resulting in research tools such as IDAC,[6] OASYS,[7] and OPASYN.[8] These automation tools have partly incorporated *design equations* and *circuit topology* synthesis. However, they finally have not made it to practice due to many reasons, mainly because of (1) lack of auto-generation tools for essential design equations and lack of consensus on how to utilize them, (2) lack of co-design methods interfacing approximate design equations and SPICE-precision model data that are neutral to technology advancement, and (3) lack of flexibility adapting to the evolving circuit topologies, among others.[9]

In the recent years, we have witnessed a growing interest in application of the machine learning methods to analog design automation, such as applying artificial neural networks,[10] genetic algorithms (GA),[11,12] Bayesian optimization (BO),[13] and other hybrid methods for analog design tasks.[14] Apart from these, heuristic optimization methods such as particle swarm optimization (PSO),[15] MOEA/D,[16] NSGA-II,[17] or other GA variants[18,19] targeting at multi-objective optimization have been attempted in numerous papers. Some of the efforts were attempted at seeking the multi-objective design space boundaries in the sense of Pareto front. However, most of the prior data-driven methods require SPICE-precision simulation-in-the-loop computations in order to achieve the goal. As an apparent shortcoming, these methods do not exploit the topological circuit properties that can be made use of to guide the design space search. The motivation of exploring the design space boundary is necessary in the analog domain, but the way computation is conquered can be explored in a much deeper way, including incorporating behavioral approaches instead of fully simulation-based.

Facing the ever-existing difficulty in the analog automation field, we focus this work on the special class of multiple-stage Op Amps, whose inherent topological structure can be incorporated. Several such salient features are worth mentioning.

- Firstly, we have noticed that the g_m/I_D design methodology[20] is well-suited to multistage Op Amp circuits because in most cases, it is adequate to derive from the g_m/I_D design parameters the DC operating conditions and the frequency compensation conditions.[21]
- Secondly, the g_m/I_D design parameters can be well-characterized at the device level by a small number of sweeping simulations, collected once for a target technology then used for many types of circuits and different operating conditions. This basic fact circumvents the necessity of using simplistic device model equations in transistor sizing because they are becoming less trustful facing advanced technology nodes.
- Thirdly, as a basic observation, the g_m/I_D quantity falls in a very narrow numerical interval regardless the transistor type (NMOS or PMOS) and the technology node. Typically, considering an interval of values between 0 and 30 V^{-1} would be sufficient for covering all kinds of device operating regions. Since the g_m/I_D quantities can be used to determine the aspect ratios W/L (i.e., sizing), it virtually transfers the sizing design space to the exploration of the g_m/I_D parameter space, whose dimension seems to be much lower, hence computationally feasible by sampling.
- Fourthly, in order to enhance the efficiency in design space exploration, we seek a behavioral approach to performance assessment without fully resorting to SPICE simulations. This strategy is enabled by advanced symbolic computation techniques developed for symbolic design equations generation.[22] Although forsaking SPICE-precision simulation, the accuracy of performance estimation can be proven to be adequate for design space exploration. This

is also in accordance to what designers often do in their daily work; namely, *pencil–paper*-based design seeks intuition rather than accuracy.

Following the above discussions, two key innovative points of this paper are emphasized: (1) Sampling-based g_m/I_D data mining and (2) adoption of circuit performance evaluation based on behavioral symbolic equations. We shall use examples to validate that accuracy of circuit performance metrics predicted by behavioral models is adequate in comparison with SPICE simulation by experimenting on different circuit topologies and several technology nodes. Furthermore, effectiveness in design space exploration by the proposed approach will be demonstrated by graphical visualization means with focused inspection on the high-frequency performance trade-off that is most critical for multistage amplifiers. The proposed methodology is in principle automatable, but the full implementation of automation is not the chief goal of this work.

A preliminary research based on the above idea was presented in the conference paper,[23] where only a limited case study and data analytics were disclosed. This journal version has enclosed the full fundamentals and an extended justification on the multitude of aspects regarding this new methodology. Furthermore, we also make a comparison with a multi-objective GA algorithm (NSGA-II) in searching the performance Pareto boundaries, showing the superiority of the sampling-based data mining method.

This paper is organized as follows. In Section 2, we explain the motivations and advantages of sampling g_m/I_D for the design of multistage CMOS Op Amps, then introduce the principle of behavioral performance evaluation. The detailed sampling strategy and data mining methods are elaborated in Section 3, which leads to the tool named *Sampling-based Op Amp Design Aid* (SODA). In Section 4, experimental studies are present in order to validate the effectiveness by applying the proposed SODA tool to the sizing space exploration of several multistage Op Amps, with a comparison made to the NSGA-II method in searching the Pareto fronts. Finally, Section 5 concludes the paper.

2 | RATIONALE OF g_m/I_D SAMPLING

The g_m/I_D design method is now popular in the analog design community.[20,21,24–27] Continuous research in the past three decades has basically justified that this methodology is not only suitable to low power design by adopting the moderate inversion operating region but also applicable in the strong inversion design. Professor Sansen predicted in his paper in 2015[24] that the g_m/I_D design methodology was expected to sustain beyond the FinFET technology. As long as the basic small-signal analysis remains valid for advanced technology node, design equations that are commonly used in multistage Op Amps design can safely work with the g_m/I_D quantities in design reasoning for GBW (gain-bandwidth product), phase margin, DC gain, and power, and so on. g_m/I_D data are flexible in that they can work with performance deduction in both the DC domain (traditional) and the AC domain by the so called C/I_D extension.[28] When working with multistage Op Amps, the C/I_D method can be replaced by frequency-domain design equations.

2.1 | Limited range of admissible g_m/I_D values

For the sizing purpose, the following list of g_m/I_D tables is sufficient.

- Transistor intrinsic gain A_i versus g_m/I_D,
- V_{GS} versus g_m/I_D,
- $C_{g(s|d)}/I_D$ versus g_m/I_D,
- g_{ds}/I_D versus $(g_m/I_D, V_{ds})$ (three-dimensional),
- $I_D/(W/L)$ versus g_m/I_D, etc.

Depending on the special design needs, it is possible to extend the above list to other types of g_m/I_D tables. Figure 1 shows two sample tables of V_{GS} and the intrinsic gain with respect to the x-axis g_m/I_D for the CMOS Predictable Technology Model (PTM) 45 nm process.[29] A notable fact seen from these plots is that the g_m/I_D axis has a short admissible range, say, $[2.5, 20]$ V^{-1} in this case. Other technology libraries bear the similar property.

It henceforth leads to a simple observation: The performance of an analog circuit is basically determined by a combination of transistors with their g_m/I_D values restricted in such a small range. We further point out that, although the

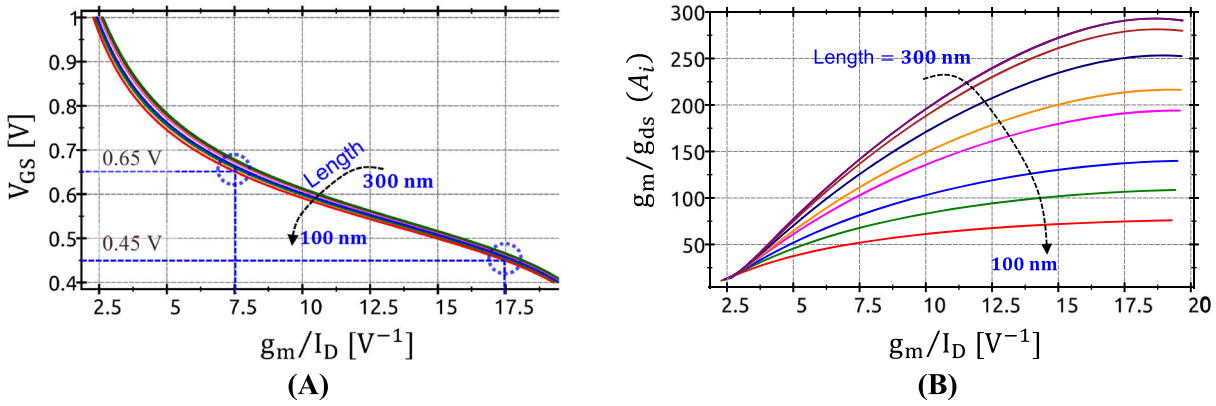

FIGURE 1 g_m/I_D curves for NMOS transistor in Predictable Technology Model (PTM) 45 nm CMOS technology library. (A) V_{GS} versus g_m/I_D and (B) intrinsic gain versus g_m/I_D.

number of transistors in a typical analog circuit could be large, implying a nontrivial g_m/I_D search space for circuit optimization, a closer look at certain categories of circuits, such as multiple-stage Op Amps, may reveal that many devices would share common g_m/I_D values, a basic property known as g_m/I_D sharing. Furthermore, other forms of sharing mechanisms exist, such as I_D sharing and V_{GS} sharing. These would create design constraints in the free-sizing space. Moreover, pole-zero cancelation constraints for multistage Op Amps would further impose sizing restrictions on the available g_m/I_D selections, resulting in substantial dimension reduction of the design space. Therefore, it is a highly appealing direction to explore for the class of multistage Op Amps.

Note that there exist inherent relationships between the small-signal parameters and the g_m/I_D quantity via the presimulated sweeping data. As soon as the g_m/I_D values of all transistors and their biasing voltages have been determined, we can immediately derive all small-signal transistor parameter values by table lookup. They can then be substituted to the behavioral design equations (if they are available), such as equations for the DC gain, simplified transfer function (TF), noise, and distortion. As long as the behavioral design equations have maintained good accuracy, most of the performance metrics so estimated would be trustful reflection of the true circuit performance. Accuracy by this method should be much better than what is obtained by *pencil–paper* derivations, although less than that by SPICE simulation. In the analog domain, this kind of behavioral evaluation is usually good enough as approximate small-signal analysis is known to be a powerful tool as long as the dominant circuit behavior is captured. Circuit performance such as the DC gain, slew rate (SR), and unity gain frequency (UGF) can be reliably captured by dominance-based roughly derived equations. In most occasions, macrolevel circuit structures such as gain stages and compensation profile are the main determining factors on these key performance metrics. One may refer to such deductive practice in many modern publications.[30–32] Simplifying symbolic analysis serves the analogous goal.[22]

Of course, a random combination of device g_m/I_D values would not always give rise to a circuit with good performance. But if the coverage of sampling is comprehensive, it is possible that a portion of circuit samples could be generated with satisfactory performance. The reason why sampling-based sizing might be a good strategy is that device parameters that appear in the design equations need not be assigned any a priori order like they are used in *procedural design methods*.[33] Any a priori determination of the order of parameters being calculated in sizing is a biased decision that might not fairly lead to design space exploration. Procedural design actions might at best lead to one or two good candidate designs but would not contribute to a full profiling of the design space.

Design space profiling is supposedly a collective effort, for which a large amount of data are necessary, and this amount of data must be generated without pre-injection of any subjective design bias. This task seems to be daunting at the beginning, but after inspecting the structural properties of many multistage Op Amps, we find that the aforementioned *sharing* property with the advantage of small amount of g_m/I_D samples in the small intervals makes the collective sampling feasible. This tactic is further enhanced by the possibility of behavioral equation-based performance evaluation. As long as sizing data are quickly generated and circuit sizing samples are swiftly evaluated, we can go on to conduct *data mining* to do collective profiling, such as sifting those qualified circuits with admissible performance measures. Furthermore, we may plot the histograms resulting from sifting, providing graphical views on the different aspects of design space. In the meanwhile, Pareto front is a natural consequence of such data mining work. These are

the basic work contents that we shall address in the sequel. The basic workflow of the proposed scheme is summarized in Figure 2.

A number of design metrics are considered in multistage Op Amp design; typical metrics are DC gain, GBW, phase-margin (PM), SR, power supply rejection ratio (PSRR), common mode rejection ratios (CMRR), noise power spectrum density, power consumption, harmonic distortion (HD2, HD3), and others. However, at an initial design stage, what the designers concern the most are usually a few very fundamental performance metrics that influence the adoption of a specific topology. DC gain, PM, GBW, and drivability are often the most fundamental ones. Other more advanced metrics can be accommodated later as the design is refined. Considering all design metrics at once is not common practice, which is also the design style we presume in this work.

2.2 | The g_m/I_D sharing property

It is worth noting that multistage Op Amps bear certain topological structures that naturally embody the g_m/I_D device sharing property. The g_m/I_D value of a device is controlled by the gate-source biasing V_{GS} and the channel length L only. When transistors are interconnected to form an Op Amp circuit, structural constraints like current mirroring and symmetry impose design variable constraints. Hence, identification of such particular topological device grouping boils down to g_m/I_D sharing, reducing the number of free variables in design space configuration.

Shown in Figure 3 are some circuit structures that exhibit g_m/I_D sharing. Figure 3A is a regular current mirror bank that is often employed in biasing multiple branch currents. If these transistors are selected to have equal channel length, they would share the equal g_m/I_D due to equal V_{GS} biasing. The structure of Figure 3B is another common structure seen in multistage designs. It is a current mirror (M_0, M_1) whose drain voltage drives the gate of a common source transistor (M_2). If the mirror is symmetrical, the drain voltages at both terminals are approximately equal, leading to the equal V_{GS} of all three transistors. Then if these transistors employ equal channel length L, they would have equal g_m/I_D.

It is hereby noted that, if full automation is desired, it is necessary to incorporate structural identification of devices that share common g_m/I_D because that would benefit reduction of the number of free variables for optimization.

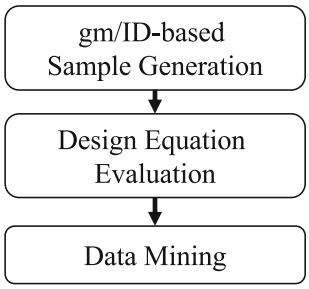

FIGURE 2 Proposed design space exploration workflow.

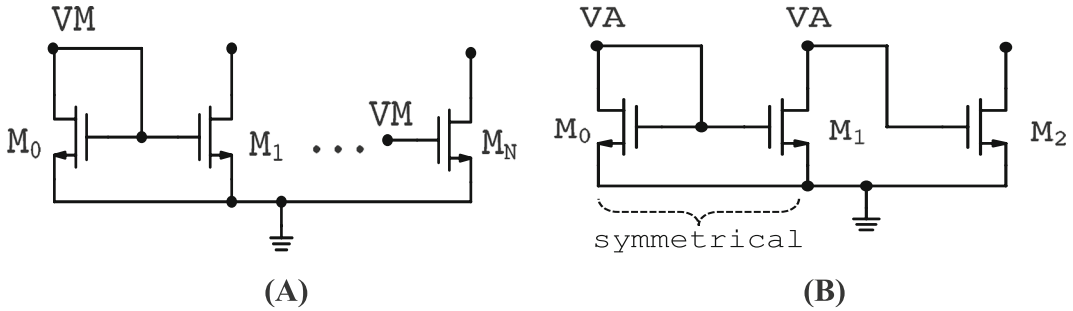

FIGURE 3 Circuit structures exhibiting g_m/I_D sharing: (A) Current mirror bank and (B) current mirror (M_0,M_1) driving a common source transistor M_2.

Preliminary work incorporating this idea in sizing was attempted by Shi[21] but in a different context on design equation generation. Another preliminary effort on auto-recognition of functional cell structures exhibiting common voltage biasing has been addressed recently by Shi,[22] where the g_m/I_D sharing property is considered in a sizing scheme.

2.3 | Behavioral performance evaluation

We note that g_m/I_D is a quantity that combines the small-signal parameter g_m and the large-signal parameter I_D, both are pertinent to the circuit performance metrics. In particular, g_m of a certain device determines GBW or noise while I_D is to do with the power and SR and so on. Such particular connections from the design parameters to the circuit performance are not explicitly offered by using W/L as the design parameters. This is a first benefit of using g_m/I_D instead of W/L as the sizing parameter. Secondly, from the g_m/I_D lookup tables (LUTs) (or curves), we are able to retrieve the binding between the device biasing voltages/currents and the small-signal quantities, such as g_m, R_{ds}, and C_{xy} (denoting parasitic capacitances between arbitrary terminals x and y of a transistor). As soon as the small-signal parameter values are known for all transistors, a number of metrics regarding small-signal (or AC) circuit performance can be measured/calculated directly (which includes DC gain, PM, GBW, and noise, etc.) As a matter of fact, these small-signal metrics often reflect the most important Op Amp performance in the design practice.

Upon knowing all small-signal parameter values by g_m/I_D table lookup (instead of SPICE DC simulation), we may find the circuit AC performance by directly evaluating symbolic design functions that characterize the relevant ac responses. Symbolic methods that can be applied for such purpose have been available by referring to works earlier.[34–36] Proper fidelity in the behavioral characterization is necessary, but we do not require SPICE-precision accuracy for the purpose of design space exploration because correlation between behavioral characterization and SPICE-precision true performance is well-established by symbolic analysis.

To the best of the authors' knowledge, analytical design equations are the most useful means of design aid when proceeding to the numerical performance evaluation phase because they are the neutral characterization of the circuit performance without dependence on the specific circuit biasing and sizing. The only basic requirement is that all transistors have been biased in saturation so that the small-signal model used in symbolic analysis is valid. Unsaturated transistors could lead to failure of the small-signal model, which could lead to loss of accuracy.

TFs are a special class of behavioral models that can be applied to represent a variety of small-signal behaviors of Op Amp circuits. Typically, a TF can be generated in the following s-factored polynomial form by a symbolic tool

$$A(s)=\frac{N(s)}{D(s)}=\frac{N_0+N_1s+N_2s^2+\ldots+N_ms^m}{D_0+D_1s+D_2s^2+D_3s^3+\ldots+D_ns^n}, \tag{1}$$

where the coefficients N_i and D_j are functions of small-signal circuit parameters. As long as all small-signal parameters involved in function (1) are available, we may directly obtain the numerical frequency response by one round of numerical evaluation. From the resulting magnitude and phase response curves (e.g., Bode plot in Figure 4), we may directly find the relevant AC performance measurements. An alternative method is to find the performance measures by directly processing the coefficients of the symbolic TF (1), obtaining more detailed numerical estimates on the poles and zeros. Behavioral performance estimation as such does not require any invocations of costly SPICE simulation, saving the cost of evaluating all circuit samples during design space exploration.

If more behavioral performance measures are wanted, one may apply existing behavioral modeling methods for such purpose beyond small-signal modeling techniques. Issues as such are not further pursued in this paper as it would not affect our purpose of establishing the methodology.

3 | SAMPLING STRATEGY AND COMPUTATION PROCEDURE

Monte Carlo (MC) sampling has been widely used in the VLSI circuit design field. However, when dealing with high dimensional problems, the efficiency of direct MC sampling is a concern due to the curse of dimensionality. Better sampling strategies have been extensively studied in the field of statistics, whose quality can be measured by the notion of *low-discrepancy sequence*,[37] With a meticulous generation method of a low-discrepancy sequence (e.g., *Sobol′* sequence), it was claimed by Singhee and Rutenbar[37] that *quasi-Monte Carlo* (QMC)[38] could be a better sampling strategy than

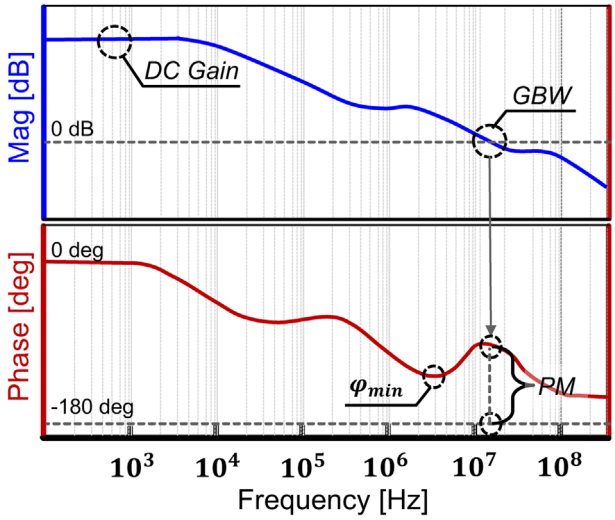

FIGURE 4 Reading circuit AC performance from the Bode plot. (ϕ_{min} denotes the minimum phase before gain-bandwidth product [GBW]).

regular MC or *Latin hypercube sampling* (LHS). However, complexity involved in the implementation of QMC could be an issue, which leaves the QMC method not most favorable in less mission critical applications.

In this work, we choose to use the regular LHS method,[39] which is one of the low-discrepancy sample generation methods that seems to be adequate for our purpose in this work. Properly distributed LHS sequence ensures a good coverage of the sizing parameter space in that the multiple-objective performance boundary can be effectively profiled by visualizing the samples in the performance space.

Another advantage of using LHS is that the independent sampling variables can take their own customized probability density function (PDF) to emphasize certain regional importance in the respective sampling region. Emphasizing different regional importance of a sampling variable is virtually one kind of design knowledge injection.

3.1 | LHS

Figure 5 illustrates a visualization of the LHS procedure. We let x_i be the ith variable, $i = 1,\ldots,K$, whose sampling range is denoted by $\mathcal{R}_{x_i} = [x_{i,min}, x_{i,max}]$. The K sampling variables are arranged in K consecutive rows. Each variable range is divided into N small subintervals, indicated by N small boxes in each row in the figure. In each sampling round, a vector of dimension K is formed, each entry is a sampled value from a randomly selected box in the corresponding row. As soon as M sample vectors have been drawn, the sampled subinterval boxes are crossed out (i.e., *Round 1* to *Round 2* in Figure 5), forbidding subsequent sampling there.

The M samples drawn in each subinterval follow MC sampling. After N rounds of sampling, all boxes in the sampling matrix must have been crossed out. As a result, a total of $\mathcal{S} = N \cdot M$ sample vectors have been generated.

Regarding the distribution of samples in each parameter sampling range, it is unnecessary to evenly distribute the subintervals in each range as illustrated in Figure 5. LHS can easily allow us to unevenly subdivide a sampling interval according to any pre-efined *cumulative distribution function* (CDF) selected for a sampling variable. More specifically, let $F(x_i)$ be the corresponding CDF of variable x_i. Evenly dividing the range of each $F(x_i)$ into M subintervals, mapping them to the x-axis, we may obtain a set of unevenly distributed subintervals customized to this CDF. This principle is illustrated in Figure 6.

3.2 | Data mining strategy for design space exploration

According to the flow chart shown in Figure 2, the parameter values determined by one sample generation action can be used to deduce the circuit sizing. In the meanwhile, all small-signal transistor parameter values can be obtained by

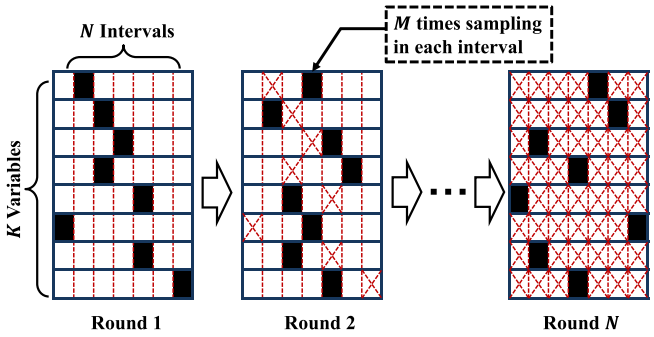

FIGURE 5 Illustration of the Latin hypercube sampling (LHS) procedure. The crossed areas indicate forbidden subintervals after sampling there has been finished.

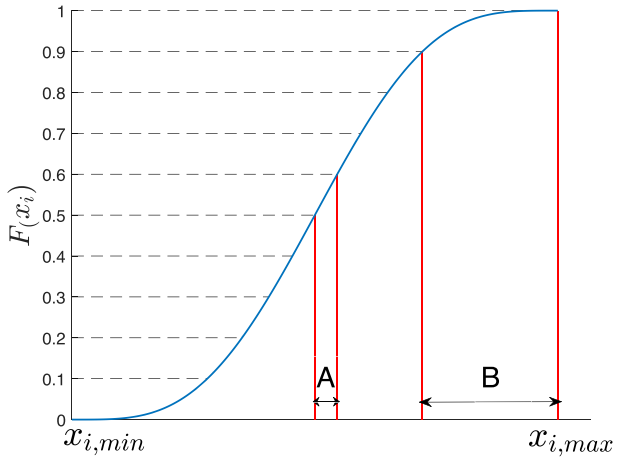

FIGURE 6 Sampling interval subdivision according to a cumulative distribution function (CDF) $F_{x_i}(x)$.

gm/ID table lookup, which can be used to calculate the AC performance of the working circuit. In addition, because the VDD and all quiescent biasing currents are known, we may further calculate the total power of the circuit. Consequently, the following four design metrics {gain, GBW, PM, power} can be determined for each circuit sample. By generating an enough number of samples with performance evaluation this way, a large volume of performance metric vectors can be collected accordingly. Based on this set of samples, we may conduct a data mining procedure to retrieve understanding on the design space. Several design space exploration means are discussed next, which are performance histogram visualization, Pareto front visualization, and performance surface visualization.

3.2.1 | Performance histogram visualization

A useful data mining strategy is by plotting all performance histograms; see the examples shown in Figure 7, where two histograms for two performance metrics are plotted. All histograms corresponding to the measured performance metrics can be plotted. Typically, the upper or lower bound of each performance will be set by the designer. For instance, typically lower bounds are set for the DC gain, PM, and GBW, and upper bounds are set for the power and noise and so on. Hence, by drawing a dividing line at the chosen performance bound in one histogram, the qualified samples can be separated from those disqualified. (See that illustrated in Figure 7.) By screening those samples qualifying all considered performance in each histogram, we may visualize the portion of samples that qualifying all inspected metrics. (See the dark shaded parts in Figure 7.) The intersection set of all qualified samples is supposed to be nonempty for a successful design qualification. Based on these qualified circuit samples, we may further data-mine them to get other visualizations such as Pareto front or the performance surfaces with respect to selected sizing parameters.

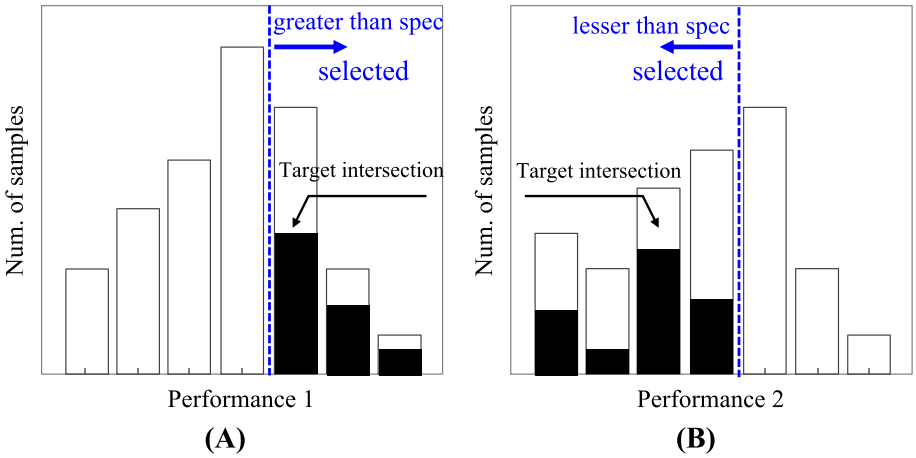

FIGURE 7 Example of performance histograms. (A) Samples qualifying the lower bound are selected. (B) Samples qualifying the upper bound are selected.

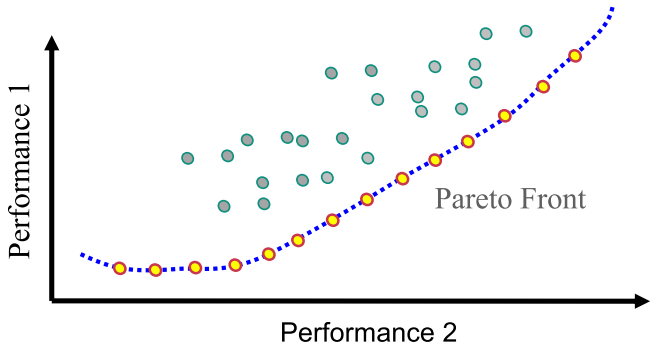

FIGURE 8 Example of Pareto front considering the trade-off between two performance metrics.

3.2.2 | Pareto front visualization

Pareto front is a measure of the design space boundary, where the design trade-off with respect to the selected performance targets at the performance extremities can be visualized; see an illustration in Figure 8. Even though we have not applied any optimization actions during the whole work cycle of sample generation and performance evaluation, the design space coverage by LHS in principle is capable of exposing the performance frontier or boundary. The reason is that we do not weigh the multiple design targets differently by introducing any weight coefficients by formulating a single objective function as done by many other optimization methods.[18] Instead, all performance measures generated by the LHS method are unbiased raw number of all selected performance measures. Although the displayed Pareto front might not be farthest reaching, it is supposedly asymptotically approaching of the true fronts as sampling regions and number of sample combinations are exhaustively extended. Of course, this kind of exhaustion is unnecessary in practice, as a rough profiling of the Pareto fronts is sufficient for reaching design intuitions.

Pareto fronts of a given circuit can be a good indication of the fundamental limitation of the specific type of circuit topology. With this part of information, the designer may consider redesign of circuit topology to further extend the Pareto fronts or adopt trade-off by sacrificing certain unimportant performance targets.

3.2.3 | Performance surfaces visualization

With an ensemble of qualified circuit samples available, we may further plot performance surfaces with respect to selected design variable combinations. Figure 9 presents such a performance surface for one performance metric with

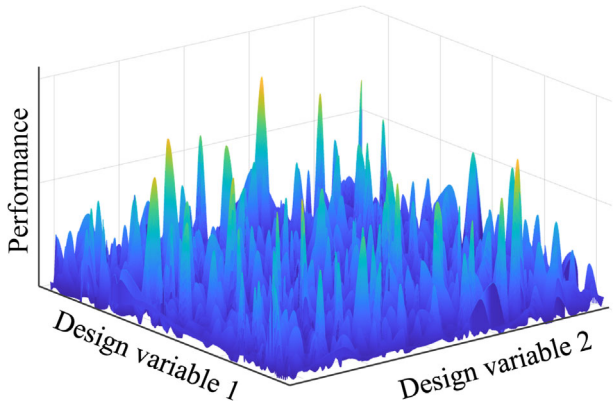

FIGURE 9 Performance surface visualization example with respect to two selected design variables.

respect to the sizing variation of two design parameters. One may also plot equipotential contour charts as done by Zhou et al.[19] A performance surface as such is an indication of the performance sensitivity view, with more spikes implying more severe sensitivity of the performance with respect to the variation of the parameters.

When proceeding to layout, joint parameter sensitivity could be a useful guide for placing those performance sensitive devices. Choosing sizing options with low sensitivity may lead to less vulnerability to process variation.

A few comments on the advantages of the sampling-based method are in order following the preceding discussions.

- Firstly, sample-based performance evaluation puts no bias toward any partial targets so that all performance measures are equally measured; that is, no prioritization of performance has been imposed.
- Secondly, the sample-based evaluation does not require solving a set of simultaneous behavioral performance equations to determine sizing. As we know, nonlinear solving could highly likely lead to local solution, losing the possibility of finding a global profile.
- Thirdly, adding more performance metrics to this computation framework is fairly easy provided that good behavioral performance model can be established.

Other advantages based on this scheme will be elaborated later via circuit examples.

3.3 | Computation procedure

We are now ready to summarize the computation procedure of the *Sizing-based Op Amp Design Aid* (SODA). This tool can be applied to sizing of most CMOS Op Amps and other similar circuits such as low-dropout regulators. The detailed computation flow chart is presented in Figure 10 with the elaboration follows.

1. Given an Op Amp circuit, formulate the set of sizing variables for sampling. Typical sampling variables could include the following: grouped g_m/I_D variables, gain stage biasing currents, biasing nodal voltages, and lumped compensation element values (capacitors and resistors). Let there be K sampling variables.
2. Define the sampling interval and the desired sampling distribution (PDF/CDF) for each independent variable.
3. Run LHS with parameters N (number of subintervals) and M (number of samples drawn from each subinterval). Generate a total of $\mathcal{S} = M \cdot N$ sample vectors of dimension K.
4. Generate the small-signal symbolic design equations for the given circuit using any verified symbolic method for performance evaluation.
5. Derive all small-signal device parameter values from the g_m/I_D LUTs, necessary for evaluating the circuit AC performance metrics. Detailed operations in this step include the following:
   (5a) Calculate parameter values of V_{GS}, g_m, C_{gs}, and C_{gd} corresponding to the sampled variables and the g_m/I_D table lookup values. In this step, we may use the g_m/I_D curve sets of V_{GS}-versus-g_m/I_D and $C_{g(s|d)}/I_D$-versus-g_m/I_D.

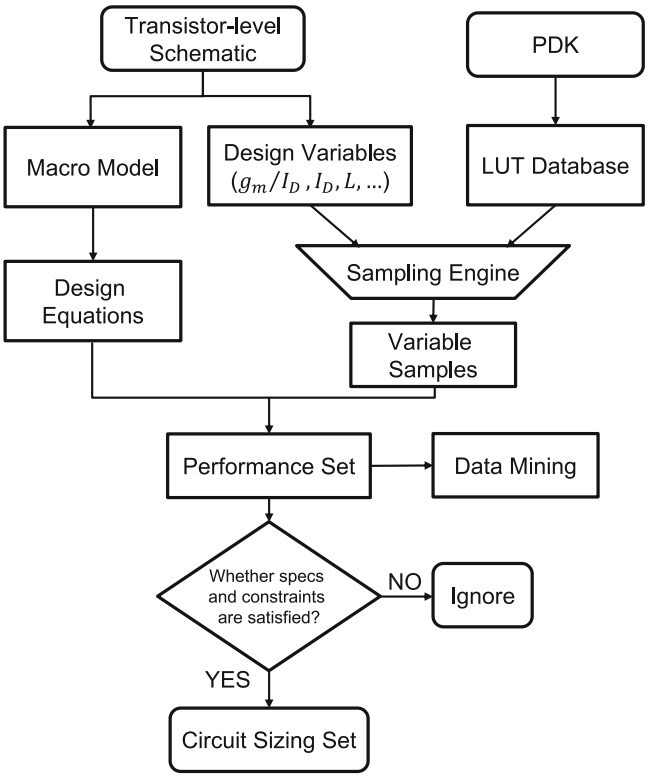

FIGURE 10 Detailed computation flow chart of the Sampling-based Op Amp Design Aid (SODA) program.

(5b) According to the power rails of VDD and VSS/GND, determined a part of nodal voltages by the known V_{GS} with the help of g_m/I_D table lookup considering voltage propagation (e.g., mirror biasing[21]). Undetermined nodal voltages should have already been selected as sampling variables.

(5c) Calculate V_{DS} for all transistors and update all g_{ds} (transistor channel conductance) values. In this step, use the three-dimensional g_{ds}/I_D-versus-$(g_m/I_D, V_{DS})$ LUTs.

(5d) Substitute to the symbolic design equations the obtained small-signal transistor parameter values g_m, g_{ds}, C_{gs}, and C_{gd} and other lumped passive (R and C) element values to calculate the desired AC performance metrics.

6. Generate the statistics of the evaluated samples by drawing performance histograms together with the preset performance bounds.
7. Select all qualified circuit samples passing all performance bound tests.
8. Derive the device sizing of all qualified circuit samples by using the $I_D/(W/L)$-versus-g_m/I_D LUTs.
9. Postprocess the qualified circuits by plotting the Pareto fronts, performance surfaces, and other interested profiles.

Depending on the depth of design space exploration, this procedure can be simplified or enhanced. For example, we may stipulate a priori that all transistors assume an equal channel length to find a quick design space. Assumption on *equal channel length* simplifies gm/ID table lookup greatly. However, if free transistor channel length is allowed, the gm/ID sharing could be broken even some transistors share common V_{GS}. In this case, more design parameters would have to be introduced, and possibly, a larger design space will end up while requiring more computational effort.

4 | EXPERIMENTAL STUDY

The SODA program was written in the C++ language. Currently, we require a text input file describing the content of a circuit, all human recognized sharing of g_m/I_D transistors, and other customized information. Sampling variables are designated by the SODA program according to the description in this file, with the prescribed parameters defining the sampling PDF functions. The input file also includes annotation on transistor symmetry. Symbolic functions are

pregenerated and converted to C++ evaluation routines. Full automation is not the priority at this experimental stage. In the future, circuit recognition routines[22] can be incorporated to fulfill the role of interfacing (i.e., substituting the description file).

All test results reported in this paper were collected from a Macbook Pro laptop (Apple M1 Pro core with 16G memory).

We report experimental results on three multistage Op Amps in this section. The first two cases are variants of a two-stage amplifier, one with the voltage buffer (VB) compensation and the other with the current buffer (CB) compensation. The third case is a three-stage amplifier with reversed nested Miller compensation and feedforward G_m and nulling resistor (RNMCFNR). Different compensations in multistage circuits are mainly created for achieving good high frequency performance with a mild cost.

All circuits were sized using the PTM 45 nm CMOS technology with 1-V supply voltage.[29] Sampling ranges for parameters such as g_m/I_D, biasing voltage, and lumped elements used in the compensation were all tentatively customized (according to technology and other constraints). For testing purpose, we restricted all design parameters to uniform distribution for sampling; in the meanwhile, we presumed that all device lengths were equal.

4.1 | Test 1: Two-stage amplifiers with VB and CB compensations

The two-stage Op Amp is shown in Figure 11, whose compensation block could be VB or CB. The first gain stage is a five-transistor stage containing a source-coupled n-channel input pair M_{N1}-M_{N2} and a p-channel mirror load M_{P0}-M_{P1}. The second gain stage is the single ended common source stage. Both compensation circuits include an active device besides biasing. The two stage biasing currents are denoted by I_{stg1} and I_{stg2}.

Figure 12 shows the small-signal model, which was used by the symbolic tool[35] to generate the voltage TF. The symbolic coefficients are presented in Table 1 together with the symbolic equations for the SR and power. These equations were coded in the SODA program for numerical evaluation.

Design variables and sampling ranges for this test instance are given in Table 2. Uniform sampling was adopted for all sampling variables. The sampling regions chosen for all g_m/I_D variables (NMOS and PMOS) were derived from the V_{GS}-versus-g_m/I_D curves (refer to Figure 1A) for the 45 nm PTM technology. We restricted $V_{GS} = [0.45,0.65]V$ for both NMOS and PMOS, which confined the g_m/I_D values to the intervals given in Table 2. We chose the external biasing current source $I_b = 10\mu A$, voltage bias $V_{B1} = 0.7V$ (for the CB cell whose variation is nonessential), and load capacitor $C_L = 5pF$. These were fixed during design space exploration. All transistors were presumed to have the fixed channel length $L = 200$ nm for PTM 45 nm.

The following five design targets were considered for this design case:

$$\begin{aligned} \text{Gain} &> 70\text{dB}, \\ \text{GBW} &> 5\text{MHz}, \\ \text{PM} &> 55^\circ, \\ \text{SR} &> 1V/\mu S, \\ \text{Power} &< 150\mu W. \end{aligned}$$

We also introduced extra constraints ($\mathcal{K}$) during the data mining phase:

$\mathcal{K}1$: All transistors are operating in the saturation region.
$\mathcal{K}2$: Limiting all device widths to the interval $[0.05,500]\mu$m as per PTM 45 nm.
$\mathcal{K}3$: For the sake of stability, we further require that the circuit step response be *underdamping* whenever conjugate nondominant poles occur. In particular, the Q factor $\left(Q = \sqrt{D_1 D_3}/D_2\right)$ (using coefficients from formulas (1)) is constrained by $Q \le 0.5$, while the real part of the complex poles, that is, $\omega_n = \sqrt{D_1/D_3}$, is constrained by $\omega_n > GBW$.

It is important to emphasize that the determination of transistor operating regions relies on calculating the voltage biases at each node. As mentioned in Section 2.1, once values for $\{g_m/I_D, I_D, length\}$ are assigned, the value of V_{GS} can be obtained. This allows us to propagate the node voltages along the transistor G-S paths starting from known nodes

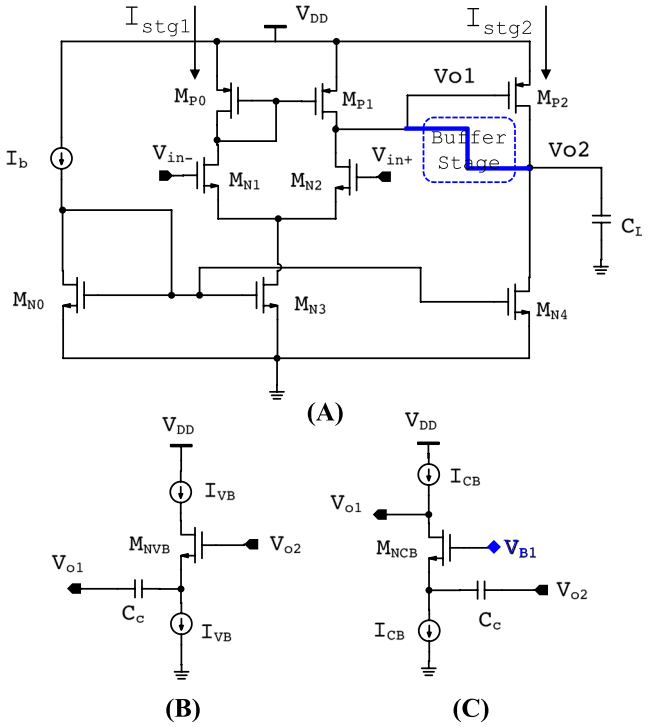

FIGURE 11 (A) Two-stage Op Amp with voltage buffer (VB) or current buffer (CB) compensation. (B) Voltage buffer. (C) Current buffer.

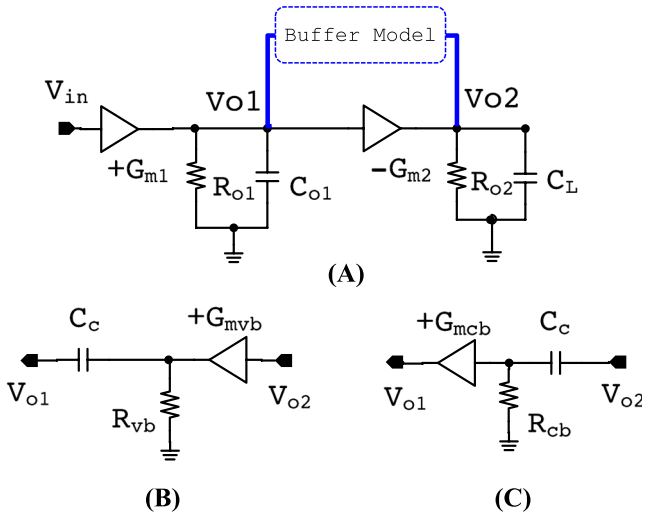

FIGURE 12 (A) Small-signal model of the two-stage Op Amp. (B) Voltage buffer (VB) compensation. (C) Current buffer (CB) compensation.

such as V_{DD}, V_{SS}, or fixed voltage biasing. For example, in Figure 11A, we calculate the value of V_{o1} by subtracting $V_{GS,P2}$ (obtained using gm/ID lookup) from V_{DD}. However, there might be certain nodes like $V_{in,cm}$ and V_{o2} that cannot be determined through this method. To account for perturbations affecting these nodes, we treat them as sampling variables and assign a range of possible values to them.

In this experiment, the total number of samples was chosen to be proportional to the number of sampling variables, say, $\mathcal{S} = K \times 10^4$, where K is the number of sampling variables. Summarized in Table 3 are the reference sampling statistics and the runtime for the two two-stage Op Amps.

TABLE 1 Design equations for the two-stage MCVB and MCCB Op Amps.

Eqn. name	MCVB	MCCB
TF	$N_0 = G_{m1}G_{m2}R_{o1}R_{o2}$	$N_0 = G_{m1}G_{m2}R_{o1}R_{o2}$
	$N_1 = R_{vb}C_c$	$N_1 = R_{cb}C_c$
	$D_0 = 1$	$D_0 = 1$
	$D_1 = G_{m2}R_{o1}R_{o2}C_c$	$D_1 = G_{m2}R_{o1}R_{o2}C_c$
	$D_2 = R_{o1}R_{o2}C_cC_L$	$D_2 = R_{o1}R_{o2}C_{o1}(C_c + C_L)$
	$D_3 = R_{o1}R_{o2}R_{vb}C_{o1}C_cC_L$	$D_3 = R_{o1}R_{o2}R_{cb}C_{o1}C_cC_L$
SR	$\min\left\{\frac{2I_{stg1}}{C_c}, \frac{I_{stg2}}{C_c+C_L}\right\}$	
Power	$V_{DD}\cdot\left(I_{DC} + 2I_{stg1} + I_{stg2} + I_{VB\|CB}\right)$	

Abbreviations: SR, slew rate; TF, transfer function.

TABLE 2 Design variables and sampling regions for the two-stage MCVB and MCCB Op Amps.

Parameter	Range[a]	Parameter	Range
$(g_m/I_D)_{N0,3,4}$	$7.5–17.5V^{-1}$	I_{stg1}	$1–20\mu A$
$(g_m/I_D)_{N1,2}$	$7.5–17.5V^{-1}$	I_{stg2}	$1–100\mu A$
$(g_m/I_D)_{P0,1,2}$	$7.4–17V^{-1}$	$I_{VB\|CB}$	$1–100\mu A$
$(g_m/I_D)_{NVB\|CB}$	$7.4–17V^{-1}$	$V_{in,cm}$	$0.5–0.8V$
C_c	$0.5–5pF$	V_{o2}	$0.2–0.8V$

[a]Selected g_m/I_D ranges correspond to the $V_{GS} = 0.45 \sim 0.65V$ for both NMOS and PMOS.

TABLE 3 Experimental results and runtime for the two-stage test cases.

Statistic name	MCVB	MCCB
Num. of sampling variables	10	10
Total num. of samples	100,000	100,000
Num. of samples satisfying $\mathcal{K}1 - \mathcal{K}3$	19,839	17,159
Num. of qualified samples	4271	4221
SODA runtime (s)	64.5	65.8

Abbreviation: SODA, Sampling-based Op Amp Design Aid.

The performance histograms generated by SODA are plotted in Figure 13, where the three metrics, gain, GBW, and power, were considered for both compensation options (with the white bars counting the samples qualifying conditions $\mathcal{K}1 - \mathcal{K}3$ and the shaded bars for those qualifying all specifications). We note that the two circuits with different compensations have roughly equal symbolic dc gain expression (equal to the coefficient N_0 in Table 1), meanwhile, the two power measures coincide as well. The key performance difference lies in the GBW (i.e., high frequency performance), which can be seen from the distribution of the qualified samples in the GBW histograms (the shaded bars). This fact is also reflected by the symbolic expressions of D_2 in Table 1, the main factor determining the position of the second pole of each circuit.

In Figure 14, the Pareto fronts of power versus GBW are plotted together with all qualified samples. We observe that the Pareto front by the Miller capacitor and VB (MCVB) compensation is pushed farther out than that of Miller capacitor and VB (MCCB) compensation, which indicates that the MCVB compensation can achieve the same GBW as MCCB by consuming less power.

The above observation seems to be contradictory to the conventional perception that CB compensation is superior for achieving higher GBW due to the shielding effect of the Miller capacitor from the first stage. However, owing to the possible pole-zero cancelation, using the VB compensation can still achieve a better GBW while consuming less power.

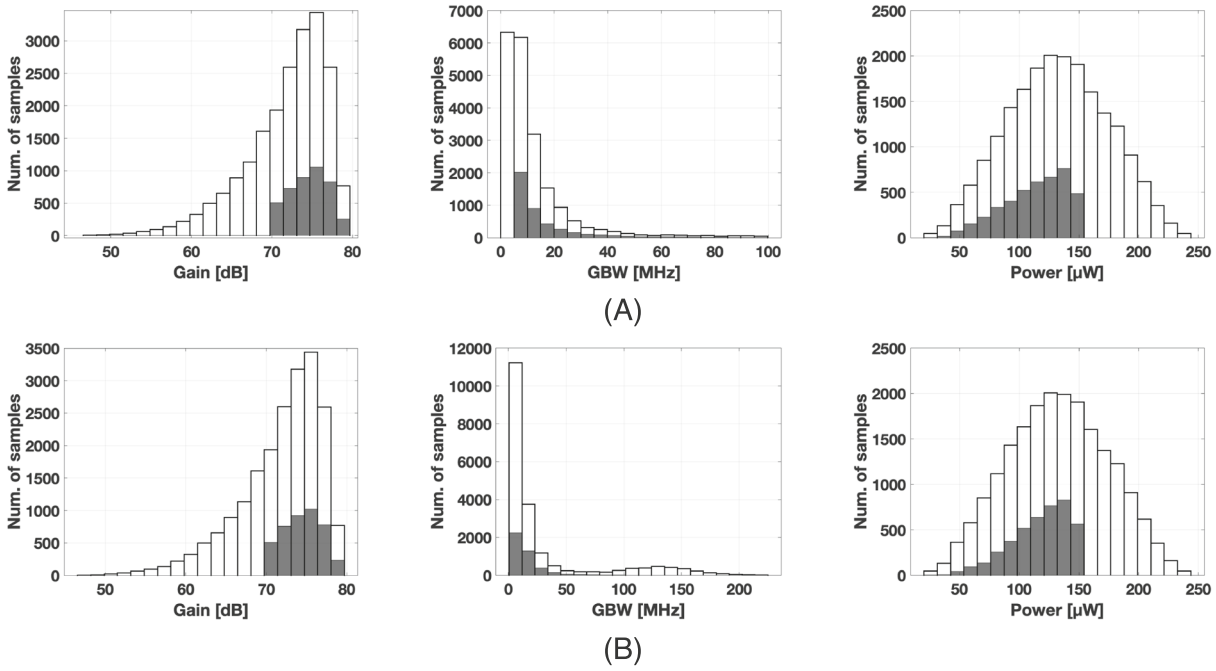

FIGURE 13 Performance histograms of gain, gain-bandwidth product (GBW), and power for the two-stage Op Amp: (A) MCVB. (B) MCCB. Shaded areas indicate the counts of the qualified samples.

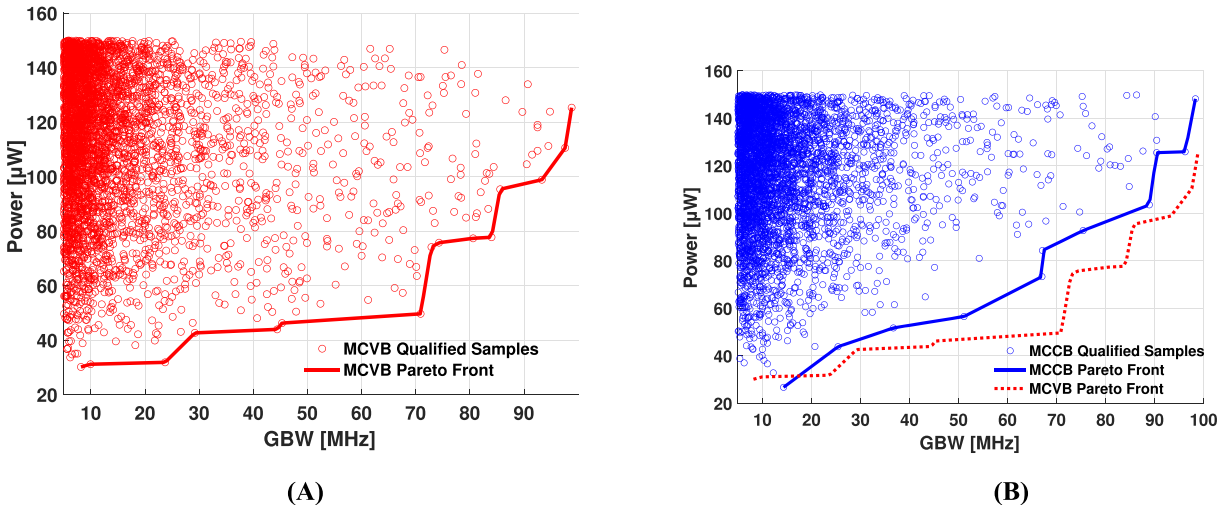

FIGURE 14 Scatter plots of qualified samples with Pareto fronts showing the trade-off between power and gain-bandwidth product (GBW) for the two two-stage amplifiers. (A) MCVB. (B) MCCB.

We note that it would be hard to do a comprehensive design space comparison as such if not using the developed sampling-based performance profiling method, which is more powerful in exhibiting global performance aspects of a given circuit topology, rather than local or biased performance view of a design. In the meanwhile, the computational cost is mild.

In addition, we may inspect the figure of merit (FoM) defined by

$$FoM_{2stg} = \frac{GBW}{Power \cdot C_c} \left[\frac{\text{MHz}}{\mu\text{W} \cdot \text{pF}}\right]$$

TABLE 4 Performance comparison for the two two-stage circuits with the best FoM_{2stg}.

	MCVB		MCCB	
Name	**SODA**	**SPICE**	**SODA**	**SPICE**
Gain (dB)	76.3	76.3	74.8	75.2
GBW (MHz)	58.33	56.14	52.3	50.9
PM (°)	57.1	56.09	88.6	85.1
Power (μW)	66.9	65.7	73.4	71.7
SR ($V/\mu S$)	3.3	29.9	3.5	12.6
MOS area (μm^2)	9.82		4.16	
C_c (pF)	0.6		0.56	

Abbreviations: SODA, Sampling-based Op Amp Design Aid; SR, slew rate.

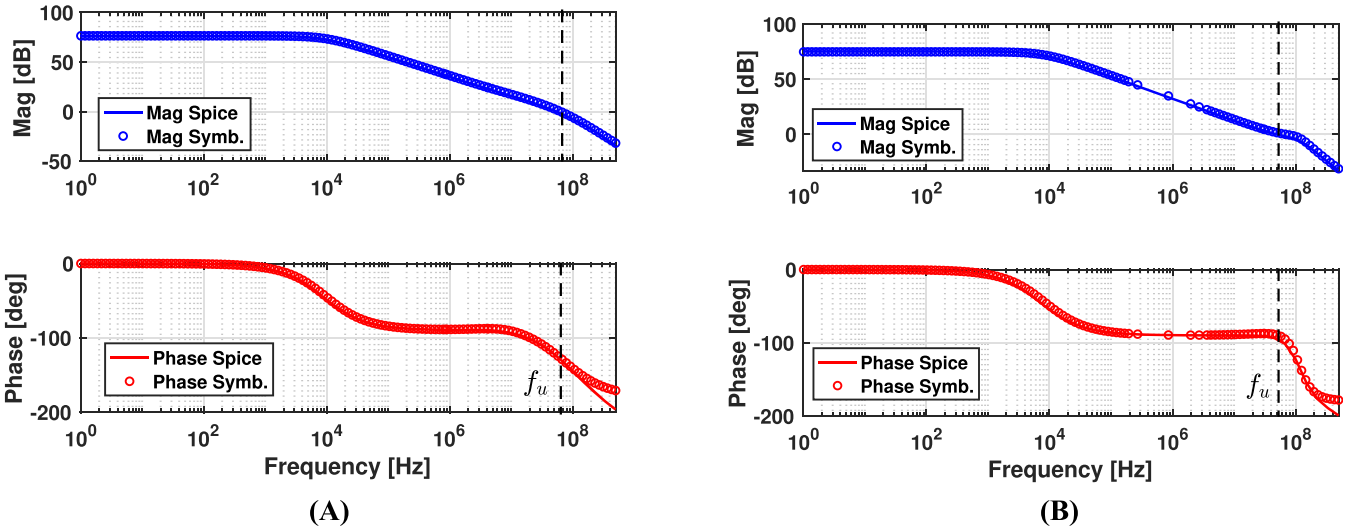

FIGURE 15 AC responses generated by behavioral evaluation and SPICE simulation for the two two-stage Op Amp with the best FoM_{2stg}. (A) MCVB. (B) MCCB.

to assess the combined metric convolving speed, power, and Miller capacitance size. By sorting all qualified samples, we find that the best FoM for MCVB was 1.45, while for MCCB, it was 1.27. In Table 4, we list the performance values for two cases attaining the best FoM_{2stg}, evaluated by both SODA and SPICE simulation, showing the trustfulness of behavioral evaluation on these metrics. The frequency responses of these two special cases are compared in Figure 15, showing the closeness of behavioral prediction from the SPICE precision. Table 5 lists the sizing results of both circuit cases.

In particular, we also plot the performance surfaces with respect to the design variables $(g_m/I_D)_{P0,1,2}$ and $(g_m/I_D)_{N1,2}$ in Figure 16. We see densely distributed spikes in the inspected area of two g_m/I_D parameters (indicating the sensitivity), in addition to the heights of the FoM. Overall, the MCCB compensation outperforms MCVB in terms of FoM but seems to be more sensitive to the variation of the g_m/I_D pair.

4.2 | Test 2: Three-stage RNMCFNR Op Amp

The second test case is the three-stage Op Amp employing the RNMCFNR compensation,[40] whose transistor-level schematic is shown in Figure 17. The first gain stage is the standard five-transistor differential stage with the *p*-type input pair. The second gain stage is the single-ended common source stage, which is succeeded by the third stage which is noninverting and contains a current mirror gain. This Op Amp includes a feedforward path implemented by the active

TABLE 5 Sizing details of the two two-stage cases with the best FoM_{2stg}.

Parameter	MCVB	MCCB	Parameter	MCVB	MCCB
W_{N0} (μm)	1.64	0.95	$W_{P0,1}$ (μm)	11.9	3.07
$W_{N1,2}$ (μm)	1.13	1.01	W_{P2} (μm)	12.2	5.4
W_{N3} (μm)	5.95	2.08	L_{ALL} (μm)	0.2	0.2
W_{N4} (μm)	3.03	1.77	C_c (pF)	0.6	0.56
$W_{NVB\|CB}$ (μm)	0.13	2.41	$V_{in,cm}$ (V)	0.71	0.675

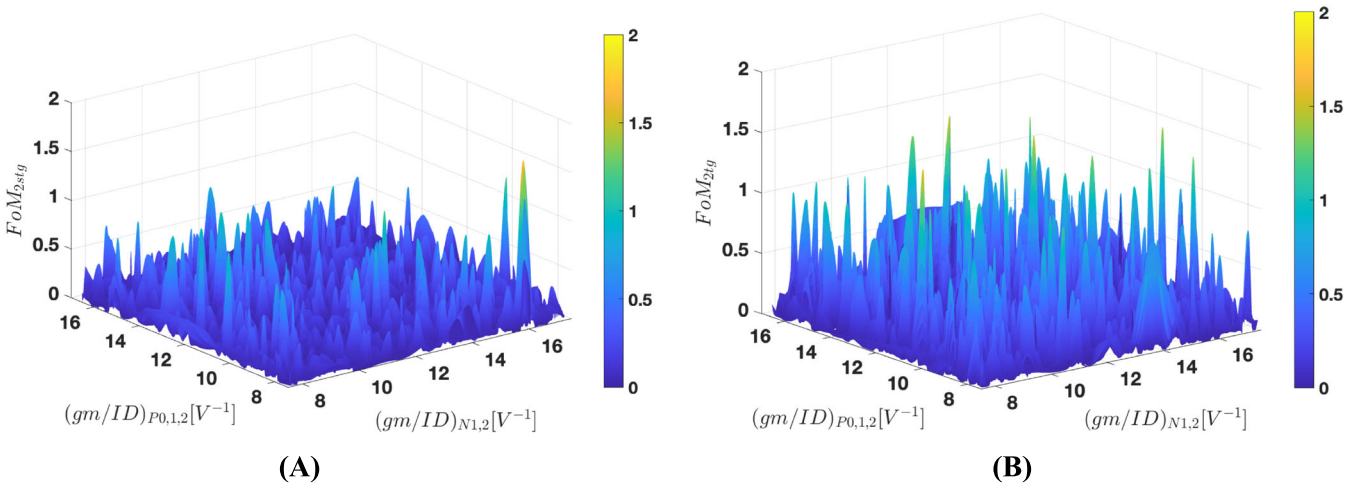

FIGURE 16 Performance surfaces of FoM_{2stg} versus $(g_m/I_D)_{P0,1,2}$ and $(g_m/I_D)_{N1,2}$ for the two two-stage Op Amps. (A) MCVB. (B) MCCB.

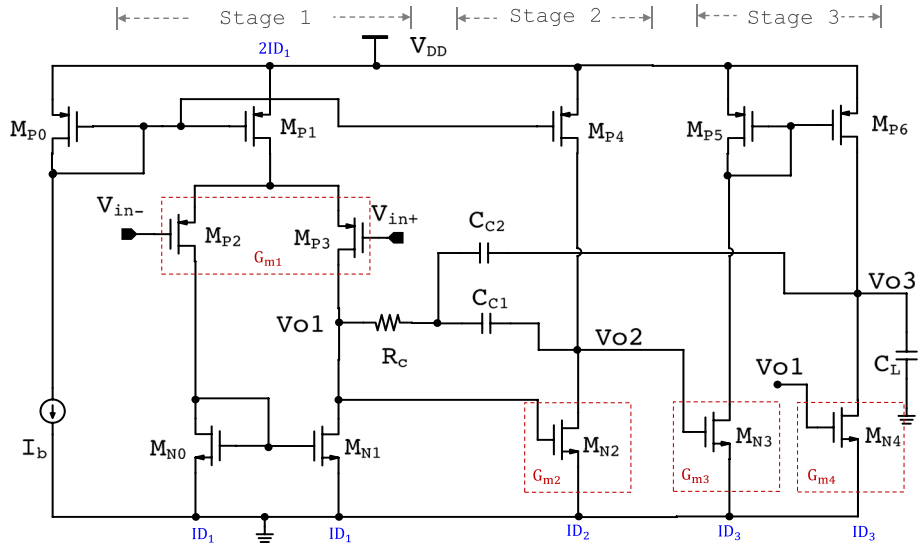

FIGURE 17 Schematic of the three-stage reversed nested Miller compensation and feedforward G_m and nulling resistor (RNMCFNR Op Amp.

load transistor M_{N4} in the output column of the third stage. Moreover, two feedback paths employing two Miller capacitors C_{C1} and C_{C2} and a common nulling resistor R_C are introduced for frequency compensation.

The circuit is known to be able to drive a large capacitive load. We chose a large $C_L = 500pF$ in our test. Similar to the previous case study, the design variables were chosen according to the circuit topology and were sampled by

assuming uniform distributions. The technology library used in sizing was again PTM 45 nm. The sampling variable information is provided in Table 6. The biasing currents are annotated in Figure 17 with $I_b = 10\mu A$. The channel lengths of all transistors were fixed at $L = 200\text{nm}$.

The small-signal model shown in Figure 18 was used for generating the symbolic TF. Dominance relations were assumed as $C_L \gg C_{c1}, C_{c2} \gg C_{o1}, C_{o2}$, and $G_{mi}R_{oi} \gg 1$ for $i = 1, \ldots, 3$. Generated symbolic TF expressions are given in Table 7.

The sampling and runtime information for this case are collected in Table 8. Design targets for this circuit in the qualification test were chosen as

$$\begin{aligned} \text{Gain} &> 100\text{dB}, \\ \text{GBW} &> 1\text{MHz}, \\ \text{PM} &> 60^\circ, \\ \text{Power} &< 150\mu\text{W}. \end{aligned}$$

Moreover, the constraints adopted in qualification test were the same as in the previous case study, that is, conditions $\mathcal{K}1 - \mathcal{K}3$. Performance histograms for this circuit are presented in Figure 19 (with the white bars standing for the samples sifted by satisfying $\mathcal{K}1 - \mathcal{K}3$ while the shaded bars for those qualifying all specification). Moreover, the scatter plot and the Pareto front of power versus GBW are displayed in Figure 20.

We also define a composite FoM for this circuit by

TABLE 6 Design variables and sampling settings for the three-stage RNMCFNR Op Amp.

Parameter	Range[a]	Parameter	Range
$(g_m/I_D)_{N0,1,2,3,4}$	7.5–17.5V^{-1}	I_{D1}	1–20μA
$(g_m/I_D)_{P0,1,4}$	7.4–17V^{-1}	I_{D2}	1–100μA
$(g_m/I_D)_{P2,3}$	7.4–17V^{-1}	I_{D3}	1–100μA
$(g_m/I_D)_{5,6}$	7.4–17V^{-1}	C_{c1}	0.5–5pF
$V_{in,cm}$	0.2–0.5V	C_{c2}	0.5–5pF
V_{o3}	0.2–0.8V	R_c	1–100kΩ

[a]The g_m/I_D ranges correspond to $V_{GS} = 0.45 \sim 0.65V$ for both NMOS and PMOS.
Abbreviation: RNMCFNR, reversed nested Miller compensation and feedforward G_m and nulling resistor.

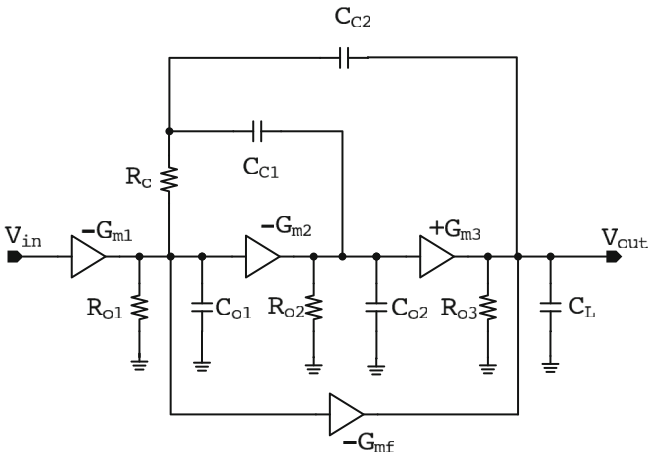

FIGURE 18 Small-signal model of the three-stage reversed nested Miller compensation and feedforward G_m and nulling resistor (RNMCFNR) Op Amp.

TABLE 7 Design equations for the three-stage RNMCFNR Op Amp.

Eqn. name	RNMCFNR
TF	$N_0 = G_{m1}G_{m2}G_{m3}R_{o1}R_{o2}R_{o3}$
	$N_1 = N_0R_c(C_{c1}+C_{c2})$
	$N_2 = G_{m1}R_{o1}R_{o2}R_{o3}C_{c1}C_{c2}\{R_C(G_{m2}+G_{mf})-1\}$
	$D_0 = 1$
	$D_1 = G_{m2}G_{m3}R_{o1}R_{o2}R_{o3}C_{c2}$
	$D_2 = R_{o1}R_{o2}R_{o3}C_{c1}\{(G_{mf}+G_{m2}-G_{m3})C_{c2}+G_{m2}C_L\}$
	$D_3 = R_{o1}R_{o2}R_{o3}C_{c1}C_{c2}C_L$
	$D_4 = R_{o1}R_{o2}R_{o3}R_cC_{o1}C_{c1}C_{c2}C_L$
Power	$V_{DD}\cdot(I_b+2I_{D1}+I_{D2}+2I_{D3})$

Abbreviations: RNMCFNR, reversed nested Miller compensation and feedforward G_m and nulling resistor; TF, transfer function.

TABLE 8 Experimental results and runtime for the three-stage test circuit.

Statistic name	Number
Num. of sampling variables	12
Total num. of samples	120,000
Num. of samples satisfying $\mathcal{K}1-\mathcal{K}3$	15,349
Num. of qualified samples	712
SODA runtime (s)	100.6

Abbreviation: SODA, Sampling-based Op Amp Design Aid.

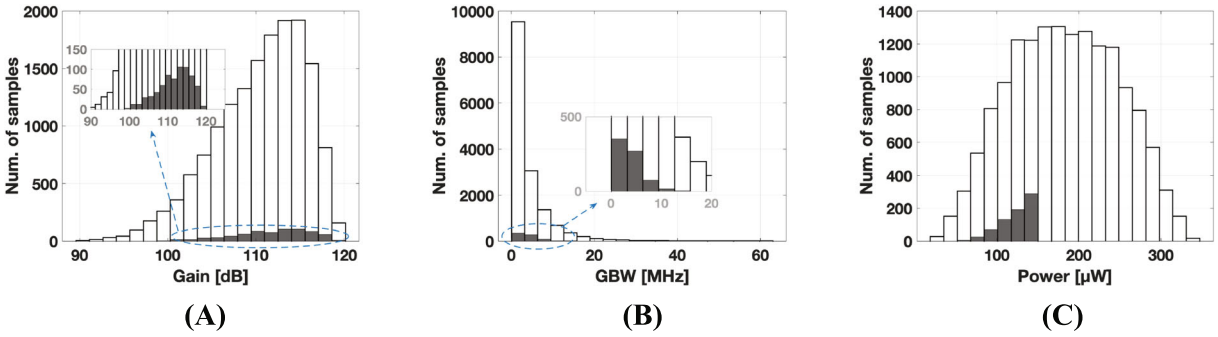

FIGURE 19 Performance histograms for the three-stage reversed nested Miller compensation and feedforward G_m and nulling resistor (RNMCFNR) Op Amp: (A) Gain. (B) Gain-bandwidth product (GBW). (C) Power.

$$FoM_{3stg} = \frac{GBW\cdot C_L}{Power}\left[\frac{\text{MHz}\cdot\text{pF}}{\text{m W}}\right]$$

where the drivability is manifested by the inclusion of the load capacitor C_L. The best FoM_{3stg} discovered by the SODA program turned out to be 52,975, which was corrected to be 48,661 by SPICE simulation. More information about performance cross-validation is presented in Table 9. We found that the superior designs picked by SODA could achieve a FoM nearly 10X of that by the manual design reported by Grasso et al..[40] Sizing results are listed in Table 10, and the AC response cross-validation is plotted in Figure 21. Furthermore, Figure 22 discloses the performance surface of FoM_{3stg} versus the pair of parameters $(g_m/I_D)_{P2-3}$ and $(g_m/I_D)_{N0-4}$. We found that a less sensitive FoM area is when the pair of parameters is restricted to $(g_m/I_D)_{P2-3}$ in $[8,10]V^{-1}$ and $(g_m/I_D)_{N0-4}$ in $[12,14]V^{-1}$.

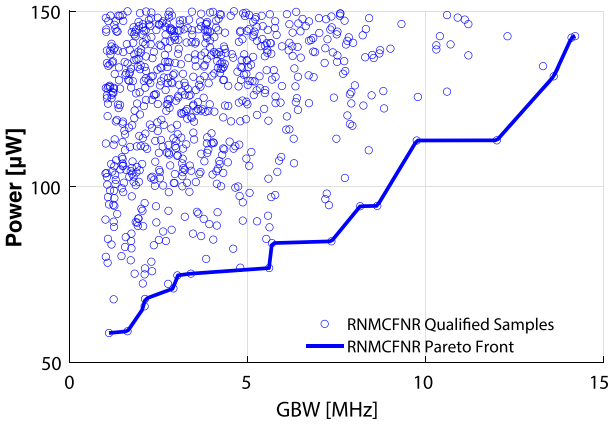

FIGURE 20 The sample scatter plot with Pareto front showing power-versus-gain-bandwidth product (GBW) trade-off for the three-stage reversed nested Miller compensation and feedforward G_m and nulling resistor (RNMCFNR) Op Amp.

TABLE 9 Comparison of the circuit performance for the three-stage case with the best FoM_{3stg}.

Name	SODA	SPICE	Grasso's work[40]
C_L (pF)	500		
Tech @ VDD	PTM 45nm @ 1V		CMOS 0.5μm @ 3V
Gain (dB)	118.5	117.8	109
GBW (MHz)	12.0	10.9	2.4
PM (°)	71.5	64.9	58
Power (mW)	0.113	0.112	0.255
SR (V/μS)	/	1.16	1.29
C_{c1} (pF)	0.55		0.35
C_{c2} (pF)	4.27		11.5
FoM_{3stg} $\left(\frac{\text{MHz}\cdot\text{pF}}{\text{mW}}\right)$	52,975	48,661	4706

TABLE 10 Sizing information of the three-stage cases with the best FoM_{3stg}.

Parameter	Value	Parameter	Value
$W_{N0,1}$ (μm)	1.7	W_{P4} (μm)	14.76
W_{N2} (μm)	4.98	$W_{P5,6}$ (μm)	2.77
$W_{N3,4}$ (μm)	2.86	L_{all} (μm)	0.2
W_{P0} (μm)	4.09	R_c ($k\Omega$)	68.76
W_{P1} (μm)	10.5	I_b (μA)	10
$W_{P2,3}$ (μm)	10.5	$V_{in,cm}$ (V)	0.294

Before concluding this section, we would like to highlight several benefits observed from experimenting the sampling-based design space exploration method:

1. With the small-signal transistor parameter values derived from the g_m/I_D data, the circuit performance metrics calculated by the behavioral symbolic design equations are reliable (without losing too much accuracy) in comparison with SPICE simulation.

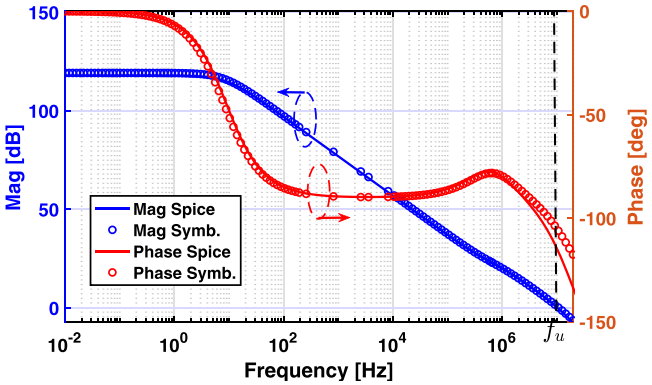

FIGURE 21 AC response compared between symbolic transfer function (TF) and SPICE for the three-stage Op Amp carrying the best FoM_{3stg}.

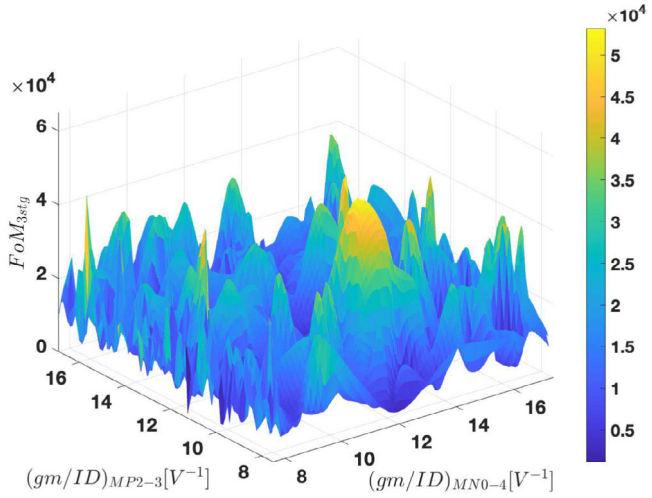

FIGURE 22 Performance surface of FoM_{3stg} versus $(g_m/I_D)_{P2-3}$ and $(g_m/I_D)_{N0-4}$ for the three-stage reversed nested Miller compensation and feedforward G_m and nulling resistor (RNMCFNR) amplifier.

2. The sampling-based data mining method is especially appealing in visually disclosing circuit performance profiles from different perspectives, including histograms, Pareto fronts, and performance surfaces. The developed method can quickly generate this set of visual information without incurring formidable computation cost.
3. Prior knowledge on the nuance of topological circuit alteration can now be visually appraised given a process technology, the comprehensiveness as such has not been attempted by previous research.
4. By the presented method, we are able to visually inspect the global optimization boundary in the multidimensional performance space without invoking any optimization scheme. Such a global design space profiling method did not exist before this work.
5. With certain effort in graphical user interface development, this tool can be applied to significantly improve the productivity in the multistage Op Amp design.

4.3 | Comparison with multi-objective GA-based optimization

Some authors have applied multi-objective optimization methods for analog circuit sizing. A widely applied heuristic optimization method is NSGA-II (nondominated sorting GA II).[17] Cuautle and Borbon[12] applied this algorithm to Op Amp sizing based on the gm/ID method. Traditionally, application of a heuristic algorithm like NSGA-II often incorporates repeated SPICE simulation for evaluating all intermediate circuit sizing solutions. Hence, the computational cost

is high. Other than that, behavioral modeling and symbolic equations are not judiciously applied in most heuristic schemes. Lack of computational efficiency has prevented these methods from exhaustively exploring the design space. Although certain quality of circuit sizing (meeting the design specs) could be generated with proper engineering on the GA parameters, the quality of optimization boundary (such as Pareto front for example) is not easily appreciated. On the other hand, because NSGA-II is a stochastic algorithm, it means that its search performance is case-by-case, with the quality of search results highly uncertain.

In this section, we make a selective comparison of our sampling-based design space exploration method to NSGA-II based optimization. Experiment shows that NSGA-II could only explore a subset of the design space that is exposed by our sampling method, and the subsets reached by NSGA-II were uncertain. It indicates that the broadness and certainty achieved by the sampling-based method are the most favorable properties.

For reference purpose, we list the pseudo-code of NSGA-II in Algorithm 1. Conventional notations such as parent population ($\mathbf{P_p}$), child population ($\mathbf{P_c}$), and the maximum number of iterations T are used. Experiment was conducted by using the PyMoo package.[41]

Algorithm 1 NSGA-II pseudo-code

Prepare necessary input files describing a circuit and its design targets.
Initialize a population $\mathbf{P_p}$ at random.
Evaluate fitness values for all solutions in $\mathbf{P_p}$.
for $t = 1 : T$ **do**
  Initiate an empty new population $\mathbf{P_c}$.
  while $size(\mathbf{P_c}) \leq size(\mathbf{P_p})$ **do**
    Select solutions from $\mathbf{P_p}$ based on the tournament selection.
    Perform the crossover operation on selected parents to generate offspring p.
    Perform mutation operation on the generated offspring to get p'.
    Update the circuit using p'.
    Evaluate the circuit fitness by SPICE simulation.
    Add p' to the child population $\mathbf{P_c}$.
  end while
  Merge $\mathbf{P_c}$ and $\mathbf{P_p}$ to form $\mathbf{R}$.
  Apply nondominated sorting to $\mathbf{R}$.
  Calculate the crowding distance for each solution in each rank.
  Select the new population $\mathbf{P_p}$ from $\mathbf{R}$ based on the nondomination ranking and the crowding distance.
end for

Note that, in the comparison, we used aspect ratios (W/L) of all transistors as the optimization parameters by NSGA-II. This is out of the consideration of studying whether our gm/ID-based design space setup (by assuming preselected gm/ID ranges and biasing conditions) could accommodate the design space defined by the traditional W/L-based design space.

In Table 11, we have shown the ranges chosen for the transistor channel widths (W) (with the channel length fixed as before) for the two two-stage circuits (used as the test cases in Section 4.1). We applied NSGA-II to find the Pareto front of power versus GBW. For each of the two-stage circuit, we collected four test run results by choosing different population sizes and iteration numbers. The GA settings and runtime are summarized in Table 12. Among the four runs, we intentionally increased either the GA population size or the number of iterations to inspect the variation of Pareto fronts at runtime.

Pareto fronts generated by the four runs of NSGA-II and our SODA program are compared in Figure 23. Observations are made as follows.

For the MCVB two-stage circuit, as the population size or the number of iterations increases, the Pareto fronts found by NSGA-II expand outwards; however, they remain in the territory predicted by SODA (whose Pareto front stays at the outmost). Note that it took NSGA-II quite some time (about 40 minutes) to reach the outmost Pareto front among the four while it took SODA only one minute.

For the MCCB case, NSGA-II was also capable of searching a far reaching Pareto front as the cost of computation is increased. We note that three Pareto fronts found by NSGA-II extended farther than the SODA front which was

TABLE 11 Design variable setting when apply NSGA-II regarding the two-stage circuits in Figure 11.

Parameter	Range	Parameter	Range
W_{N0}	0.15–1.96μm	W_{P2}	0.12–133.3μm
$W_{N1,N2}$	0.05–3.92μm	$W_{NVB,NCB}$	0.05–19.6μm
W_{N3}	0.05–7.84μm	$I_{VB\|CB}$	1–100μA
W_{N4}	0.05–19.6μm	$V_{in,cm}$	0.5–0.8V
$W_{P0,P1}$	0.12–26.7μm	C_c	0.5–5pF

TABLE 12 Experimental results and runtime for applying NSGA-II method to two-stage cases.

	MCVB				MCCB			
Test group number	I	II	III	IV	I	II	III	IV
Population size	50	100	100	200	50	100	100	200
Max. iterative times	100	100	300	500	100	100	300	500
Runtime (s)	132.9	267.8	805.97	2586.3	133.2	281.5	872.5	3154.2

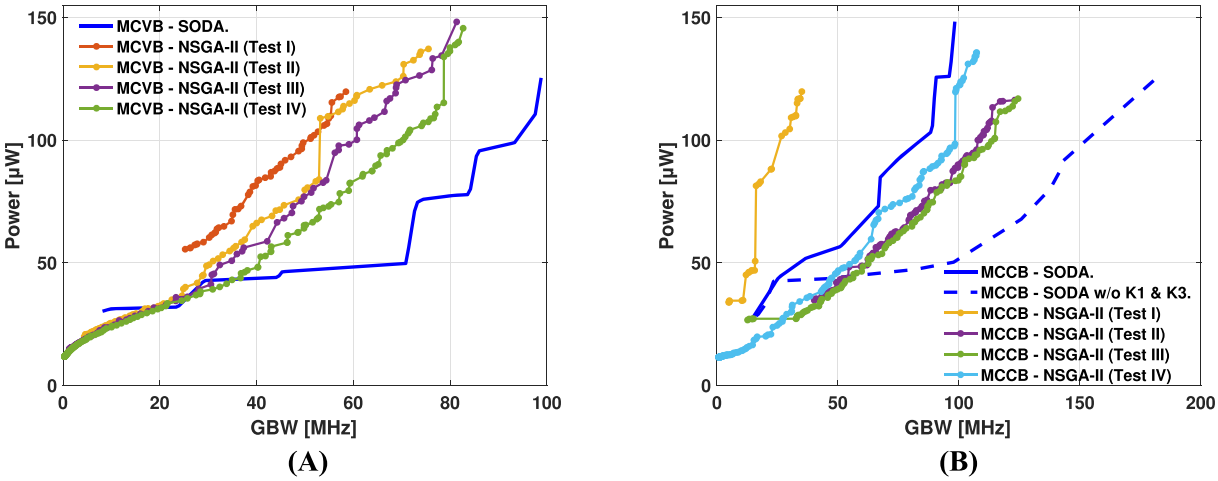

FIGURE 23 Pareto fronts of power versus gain-bandwidth product (GBW) computed by NSGA-II (four runs) and the Sampling-based Op Amp Design Aid (SODA) program for the two two-stage amplifiers. (A) MCVB. (B) MCCB.

enforced by checking conditions $\mathcal{K}1$ and $\mathcal{K}3$ mentioned in Section 4.1 (whereas in NSGA-II we did not impose such conditions). It implies that some solutions at the Pareto fronts by NSGA-II could possibly be ill-conditioned (in the sense of nonsaturation or bad Q-factor). By removing the $\mathcal{K}$-conditions in the SODA program, the predicted Pareto front was then pushed to farther right in Figure 23B. This experiment also exposes a fact that the Q-factor optimization could be a tricky issue when applying the MCCB compensation due to the existence of conjugate nondominant poles.

We also applied NSGA-II to the three-stage RNMCFNR circuit studied in Section 4.2. Sizing space is defined in Table 13 where transistor channel widths were part of the optimization variables. Four test run settings of NSGA-II are given in Table 14. Test run I of NSGA-II failed due to a small population size. Pareto fonts found by the three successful NSGA-II runs are plotted in Figure 24 and compared with the result by the SODA program. It seems that the Pareto fronts by SODA partly coincide to that by NSGA-II, but apparently SODA searched the design space at the lower power region (probably due to the pre-allocation of lower biasing currents). In contrast, NSGA-II searched the design space in the high power region, thus extended the GBW performance boundary to the far higher region.

In summary, by making a comparison between the design space exploration capabilities by SODA and NAGA-II, we have established further confidence in the foundation of SODA. Mainly, it has been validated with solid evidence

TABLE 13 Design variable setting for applying NSGA-II to the three-stage RNMCFR circuit in Figure 17.

Parameter	Range	Parameter	Range
$W_{N0,N1}$	0.05–3.92μm	W_{P4}	0.12–133.3μm
W_{N2}	0.05–19.6μm	$W_{P5,P6}$	0.12–133.3μm
$W_{N3,N4}$	0.05–19.6μm	C_{c1}	0.5–5pF
W_{P0}	1.28–13.3μm	C_{c2}	0.5–5pF
W_{P1}	0.12–53.3μm	Rc	1–100kΩ
$W_{P2,P3}$	0.12–26.7μm	$V_{in,cm}$	0.2–0.5V

TABLE 14 Experimental results of applying NSGA-II to the three-stage RNMCFNR circuit.

Statistic name	NSGA-II result			
Test number	I	II	III	IV
Population size	50	100	100	200
Max. num. iterations	100	100	300	500
Runtime (s)	Failed	332.8	737.9	2485.3

Abbreviation: RNMCFNR, reversed nested Miller compensation and feedforward G_m and nulling resistor.

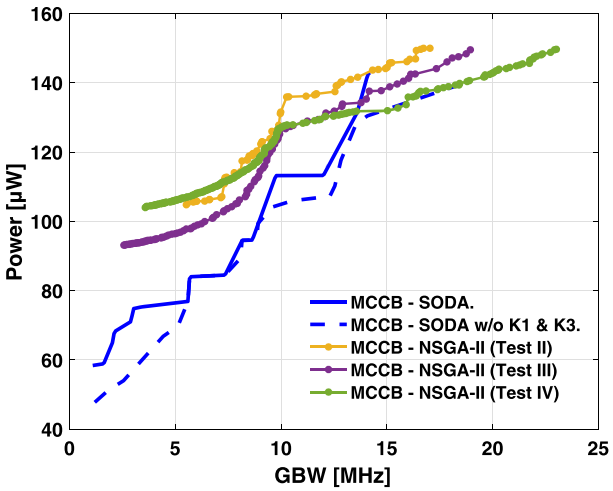

FIGURE 24 Pareto fronts regarding power-versus-gain-bandwidth product (GBW) trade-off computed by NSGA-II and Sampling-based Op Amp Design Aid (SODA) for the three-stage reversed nested Miller compensation and feedforward G_m and nulling resistor (RNMCFNR) Op Amp.

that the introduction of customized gm/ID design space and behavioral circuit performance evaluation does not sacrifice the design space fidelity. The design space portfolios generated by SODA do factually reflect the true circuit performance limitation as soon as the circuit topology is determined.

5 | CONCLUSION

We have proposed a gm/ID-based sampling method for the sizing design space exploration of multistage Op Amps. Sampled circuit instances are evaluated by using symbolic behavioral design equations for a number of key performance metrics. The sampling-based sizing process is very fast, enabling efficient design space exploration by means of data mining. With assurance of the quality of behavioral performance estimation, global images of the sizing parameter

space with performance trade-offs can be acquired, circumventing severe limitations existing with other localized optimization methods. Sampling also enables nonprioritized use of a set of design equations, guaranteeing nonbiased assessment of multiple design targets inherent with a given circuit. Consequently, the proposed design space exploration method can overcome the limitations existing with other optimization-based sizing methods that require cumulative objective functions. Experimental tests and comparison with a multi-objective GA have justified that the SODA program is not only computationally efficient (taking only couple of minutes on a personal computer for million samples) but also bears good fidelity comparing to NSGA-II that runs much slower with nondeterministic results depending on the population size and number of iterations.

Future research can be directed toward leveraging the level of automation by developing auto-recognition of circuit module and auto-generation of the sampling variables in the SODA tool. Support in graphical visualization of the design space portfolios could be a worthwhile development subject as well in order to make such a tool useful in practice.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Qixu Xie https://orcid.org/0000-0002-6179-1331
Guoyong Shi https://orcid.org/0000-0002-8655-3487

REFERENCES

1. Mallya S, Nevin J. Design procedures for a fully differential folded-cascode CMOS operational amplifier. *IEEE J Solid-State Circ*. 1989; 24(6):1737-1740.
2. Palmisano G, Palumbo G, Pennisi S. Design procedure for two-stage CMOS transconductance operational amplifiers: a tutorial. *Analog Integr Circ Sig Process*. 2001;27:179-189.
3. Cannizzaro SO, Grasso AD, Mita R, Palumbo G, Pennisi S. Design procedures for three-stage CMOS OTAs with nested Miller-compensation. *IEEE Trans Circ Syst - I: Reg Pap*. 2007;54(5):933-940.
4. Leung KN, Mok PKT. Analysis of multistage amplifier – frequency compensation. *IEEE Trans Circ Syst – I: Fundam Theory Appl*. 2001; 48(9):1041-1056.
5. Riad J, Estrada-Lopez JJ. Classification and design space exploration of low-power three-stage operational transconductance amplifier architectures for wide load ranges. *Electronics*. 2019;8(11):1-32. doi:10.3390/electronics8111268
6. Degrauwe M, Nys O, Dijkstra E, et al. IDAC: an interactive design tool for analog CMOS circuits. *IEEE J Solid-State Circ*. 1987;22(6): 1106-1116. doi:10.1109/JSSC.1987.1052861
7. Harjani R, Rutenbar R, Carley L. OASYS: a framework for analog circuit synthesis. *IEEE Trans Comput-Aided Des Integr Circ Syst*. 1989; 8(12):1247-1266. doi:10.1109/43.44506
8. Koh H, Sequin C, Gray P. OPASYN: a compiler for CMOS operational amplifiers. *IEEE Trans Comput-Aided Des Integr Circ Syst*. 1990; 9(2):113-125. doi:10.1109/43.46777
9. Shi G. Toward automated reasoning for analog IC design by symbolic computation—a survey. *Integr, the VLSI J*. 2018;60:117-131. doi:10.1016/j.vlsi.2017.08.005
10. Settaluri K, Liu Z, Khurana R, Mirhaj A, Jain R, Nikolic B. Automated design of analog circuits using reinforcement learning. *IEEE Trans Comput-Aided Des Integr Circ Syst*. 2022;41(9):2794-2807. doi:10.1109/TCAD.2021.3120547
11. Lourenço N, Martins R, Canelas A, Póvoa R, Horta N. AIDA: layout-aware analog circuit-level sizing with in-loop layout generation. *Integr, the VLSI J*. 2016;55:316-329.
12. Tlelo-Cuautle E, Sanabria-Borbon AC. Optimising operational amplifiers by evolutionary algorithms and gm/ID method. *Int J Electron*. 2016;103(10):1665-1684.
13. Lyu W, Xue P, Yang F, et al. An efficient Bayesian optimization approach for automated optimization of analog circuits. *IEEE Trans Circ Syst – I: Reg Pap*. 2018;65(6):1954-1967.
14. Liao T, Zhang L. Efficient parasitic-aware g_m/I_D-based hybrid sizing methodology for analog and RF integrated circuits. *ACM Trans Des Autom Electron Syst*. 2020;26(2):10.1-10.31.
15. Kennedy J, Eberhart R. Particle swarm optimization. In: In: Proc, of International Conference on Neural Networks (ICNN); 1995:1942-1948.
16. Zhang Q, Li H. MOEA/D: a multiobjective evolutionary algorithm based on decomposition. *IEEE Trans Evol Comput*. 2007;11(6):712-731. doi:10.1109/TEVC.2007.892759
17. 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. doi:10.1109/4235.996017

18. De Smedt B, Gielen G. WATSON: design space boundary exploration and model generation for analog and RF IC design. *IEEE Trans Comput-Aided Des Integr Circ Syst*. 2003;22(2):213-224. doi:10.1109/TCAD.2002.806598
19. Zhou R, Poechmueller P, Wang Y. An analog circuit design and optimization system with rule-guided genetic algorithm. *IEEE Trans Comput-Aided Des Integr Circ Syst*. 2022;41(12):5182-5192. doi:10.1109/TCAD.2022.3166637
20. Jespers PGA, Murmann B. *Systematic Design of Analog CMOS Circuits Using Pre-computed Lookup Tables*. Cambridge UK: Cambridge University Press; 2017.
21. Shi G. Sizing of multi-stage Op Amps by combining design equations with the gm/ID method. *Integr, the VLSI J*. 2021;79:48-60. doi:10.1016/j.vlsi.2021.03.003
22. Shi G. Automatic generation of macromodels and design equations for application to Op Amp design. *Int J Circ Theory Appl*. 2023;2023:51.
23. Xie Q, Li Z, Shi G. Design space exploration of multi-stage Op Amps by symbolic modeling and gm/ID sampling. In: In: Proc. IEEE International Symposium on EDA (ISEDA); 2023:1-6.
24. Sansen W. Analog CMOS from 5 micrometer to 5 nanometer. In: ISSCC Digest of Technical Papers, 22-27; 2015.
25. Sabry MN, Omran H, Dessouky M. Systematic design and optimization of operational transconductance amplifier using gm/ID design methodology. *Microelectron J*. 2018;75:87-96. doi:10.1016/j.mejo.2018.02.002
26. Konishi T, Inazu K, Lee JG, Natsui M, Masui S, Murmann B. Design optimization of high-speed and low-power operational transconductance amplifier using gm/ID lookup table methodology. *IEICE Trans Electron*. 2011;E94-C(3):334-345.
27. Murmann B. Gm/ID starter kit last accessed; 2023.
28. Tajalli A. Power-speed trade-offs in design of scaled FET circuits using C/ID methodology. *IEEE Trans Circ Syst I: Reg Pap*. 2021;68(2):631-640. doi:10.1109/TCSI.2020.3036683
29. Zhao W, Cao Y. New generation of predictive technology model for sub-45 nm early design exploration. *IEEE Trans Electron Dev*. 2006;53(11):2816-2823. doi:10.1109/TED.2006.884077
30. Yan Z, Mak PI, Law MK, Martins RP. A 0.016-mm^2 144-μ W three-stage amplifier capable of driving 1-to-15 nF capacitive load with >0.95-MHz GBW. *IEEE J Solid-State Circ*. 2013;48(2):527-540. doi:10.1109/JSSC.2012.2229070
31. Tan M, Ki WH. A cascode Miller-compensated three-stage amplifier with local impedance attenuation for optimized complex-pole control. *IEEE J Solid-State Circ*. 2015;50(2):440-449. doi:10.1109/JSSC.2014.2364037
32. Qu W, Singh S, Lee Y, Son YS, Cho GH. Design-oriented analysis for Miller compensation and its application to multistage amplifier design. *IEEE J Solid-State Circ*. 2017;52(2):517-527. doi:10.1109/JSSC.2016.2619677
33. Schweikardt M, Scheible J. Expert design plan: a toolbox for procedural automation of analog integrated circuit design. In: Proc. 18th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD).; 2022. Villasimius; Italy:1-4.
34. Shi G, Tan SXD, Tlelo-Cuautle E. *Advanced Symbolic Analysis for VLSI Systems – Methods and Applications*. New York, NY USA: Springer; 2014.
35. Shi G. Topological approach to symbolic pole-zero extraction incorporating design knowledge. *IEEE Trans Comput-Aided Des Integr Circ Syst*. 2017;36(11):1765-1778. doi:10.1109/TCAD.2017.2664065
36. Shi G. Automatic stage-form circuit reduction for multistage Op Amp design equation generation. *ACM Trans Des Autom Electron Syst*. 2019;25(1):2:1-26.
37. Singhee A, Rutenbar RA. Why quasi-Monte Carlo is better than Monte Carlo or Latin hypercube sampling for statistical circuit analysis. *IEEE Trans Comput-Aided Des Integr Circ Syst*. 2010;29(11):1763-1776. doi:10.1109/TCAD.2010.2062750
38. Acworth PA, Broadie M, Glasserman P, Carlo M, Carlo Q-M. A comparison of some Monte Carlo and quasi Monte Carlo techniques for option pricing. *Monte Carlo and Quasi-Monte Carlo Methods*, Vol. 1996. New York, NY: Springer New York; 1998:1-18.
39. McKay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 1979;21(2):239-245.
40. Grasso A, Palumbo G, Pennisi S. Advances in reversed nested Miller compensation. *IEEE Trans Circ Syst–I: Reg Pap*. 2007;54(7):1459-1470.
41. Blank J, Deb K. PYMOO: multi-objective optimization in Python. *IEEE Access*. 2020;8:89497-89509. doi:10.1109/ACCESS.2020.2990567

How to cite this article: Xie Q, Shi G. Multistage Op Amp design space exploration by g_m/I_D sampling and symbolic design equations. *Int J Circ Theor Appl*. 2024;52(3):1358-1383. doi:10.1002/cta.3825