Inverse Sizing and Exploration Based on Diffusion Models and Structural Knowledge

Filipe Azevedo* , Markus Leibl*, Helmut Graeb and Ricardo Martins

***Abstract*—In the realm of analog integrated circuit sizing, quick and effortless exploration of design spaces is of major importance in the face of rapid development cycles, varying specifications and increasingly complex circuits. It is also well known, that the exploitation of slack in specifications by making adjustments to transistor device sizes can lead to improved yield. We address those challenges by combining two state of the art machine learning approaches to achieve accurate generation of optimal performance points, while also allowing for the possibility to explore nearby sizings that might lead to more robust designs. To this end, a diffusion model is combined with an algorithm that analyzes the circuit and splits the problem into simpler subproblems. The models are tested on a set of typical operational amplifiers.**

I. Introduction

Since the development of SPICE, the industry-accepted method for designing analog circuits has not changed much. While the SPICE models keep becoming more complex [1], the approach of manual device tuning persists [2]. Key challenges of making analog CAD tools useful in practice are robustness, ease of use and applicability to a high range of use cases. All those aspects are not easy to grasp on a fundamental level. On the other hand, machine learning approaches excel in modeling nonlinear black box phenomena. Their combination with already existing conventional techniques has a high potential to bridge remaining gaps that prevent the adoption of tools in industry.

Analog design is inherently an inverse problem. While the calculation of performances, yield etc., is satisfactorily possible via simulation, the inverse, i.e., getting a circuit with optimal parametrization, requires experience, trial and error and complex numerical routines. As shown in previous works [3], [4], machine learning (ML) methods can be used to accurately reproduce sizings on (Pareto) optimal fronts. Furthermore, it has been shown [5] that diffusion models are well adaptable to the inverse sizing problem. Other ML based works attempted to solve this same problem. In [6], artificial neural networks (ANNs) are used to guess circuit sizings, utilizing only one ANN for modeling an entire netlist, leading to a higher learning task complexity and thus large sample sizes (in [6] 16600 or roughly $16\,x$ the sample of this work) are required to learn the inverse problem within a smaller performance range. In [7], the problem is broken down into smaller tasks, and one ANN is trained to sequentially size a single device, in the same vein as [3], feeding the outputs of the previous ANNs to the input of the next, producing a cascade of networks that can be sorted in different ways until an optimal order is found. The drawback is that for more complex circuits of e.g. 15 devices, optimally ordering the ANNs for sizing leads to a complexity of 15!.

Other works follow the traditional approach of tackling the direct sizing problem, treating it as an optimization problem [8]–[11]. Reinforcement learning was also studied, with [12]–[14] replacing the optimizer with an artificial neural network (ANN), leading to improvements in prediction speed, accuracy, and scalability. However, to the best of our knowledge, none of them show strengths in all four categories, speed, accuracy, sample efficiency and exploration.

In this work, we aim for a combination of all those virtues by applying denoising diffusion probabilistic models (DDPM) to the sizing of OpAmps, utilizing the approach in [3] of subdividing the circuits into simpler structures, which has shown high accuracy for low sample sizes. DDPMs have proven their potential for generating strongly varied, but realistic results in other fields, OpenAI's DALL-E and Stable Diffusion being among the most prominent examples. [5] proved these model's efficacy in the inverse sizing problem.

Unlike most machine learning models, DDPMs can suggest different sizings for the same problem by repeatedly sampling them. As shown in Section IV, this approach to the inverse problem can produce a more varied set of possible sizings, which is very helpful when exploring sizing options in the vicinity of performance optima. This work contributes a machine learning approach that:

- can suggest valid sizings with resulting performance in the vicinity of the required metrics, allowing for exploration of sizings while still fulfilling specifications,
- is highly data efficient,
- has an accurate expectation of the generated sizings,
- generates sizing solutions at push-button speed.

In section II, we provide an overview of the proposed method and its connection to structural analysis, as well as describing the sizing process in detail. In section III, a short background of DDPMs and our implemented models are presented. We conclude with a presentation of results for two OpAmps, in section IV.

Filipe Azevedo and Ricardo Martins are with ..., (email) Markus Leibl and Helmut Graeb are with Technical University of Munich, Chair of Electronic Design Automation (email: markus.leibl@tum.de, helmut.graeb@tum.de).
* *Both authors contributed equally to this work.*
Parts of this work have been done within the collaborative project HoLoDEC funded by the Federal Ministry of Education and Research Germany (BMBF) under the funding code 16ME0705.

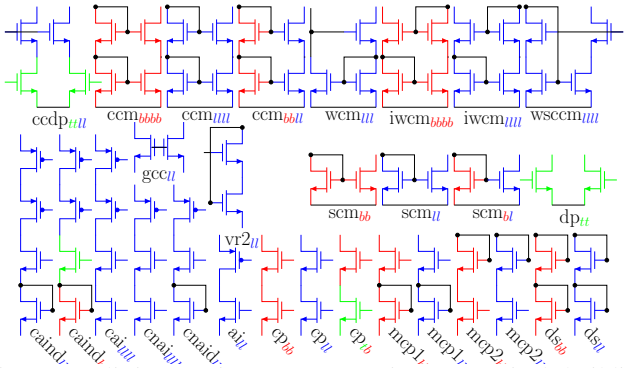


Fig. 1: Splitting various OpAmps into canonical building blocks with different functional characteristics. Indices b, l and t represent *bias*, *load* and *transconductance* functionality of devices. Adapted from [3].

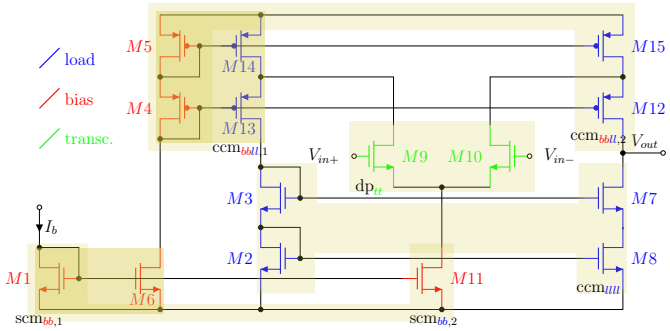


Fig. 2: Folded-Cascode OpAmp (FCOA). The circuit is decomposed into its structural and functional blocks with structural blocks shaded and functionality of the devices represented by colours.

II. From Structural Analysis to Circuit Sizing via Machine Learning

A. Basics of Structural and Functional Blocks

As detailed in [3], the method relies on the decomposition of circuits into subcircuits, which we call structural building blocks. While the structure of the blocks is defined by their devices and interconnections, we identify three functional properties, *bias, load* or *transconductance*, for each individual transistor. This leads to further discrimination among the structural blocks. But because not every combination of functional properties is useful, the total set of possible structures remains small. All possible variants of blocks can be found in figure 1.

An example of decomposing a Folded-Cascode OpAmp (FCOA) into blocks can be seen in figure 2. Its structural building blocks are shaded, the names annotated close by. Structural and functional block recognition is fully automated. The time required for this is much less than one second.

B. Overview of the Method

In contrast to [3] we use a simplified version of the model. We skip the pretraining step and directly finetune the networks. Furthermore, instead of using gain boosting for sizing of L, we employ the same model type. The sizing process is then straightforward:

- Split the opamp into its building blocks.
- Train a neural network for each of the buidling blocks, where we use the OpAmp performance as guidance for the DDPM model and train to recover only the sizings of the blocks.
- Finally, at evaluation time, recombine the predicted sizings. In case of overlap of two blocks, we take the average value.

Like in the previous work, we enforce symmetry that can be automatically detected by deterministic algorithms. For symmetry within blocks, this is done during training, for symmetry outside of blocks, this is done during recombination.

III. DDPM Background and Implementation

Several DDPMs are trained to learn the sizing and respective performance distribution of the different building blocks that make up an OpAmp. DDPMs were first introduced in [15] as a new diffusion model parametrization, which are a class of NNs mostly used in an imaging context, like data generation and restauration. These models are composed of 3 different processes, which are briefly highlighted here:

1) Forward Process: First, the training data x_0 is systematically destroyed by the addition of noise over T sequential steps, in the literature called timesteps, until a terminal distribution x_T is reached. This noise is sampled from a Gaussian distribution and at the end of the T timesteps, the resulting training data distribution resembles itself a Gaussian distribution. For example, if $T = 10$ then noise is sampled at each timestep and added to the data 10 times. This process is demonstrated by equation 1, where β_t is the variance schedule, a hyperparameter that controls how much noise is added to the data x at each time timestep t.

$$\begin{aligned} q(x_t|x_{t-1}) &:= \mathcal{N}(x_t; \mu, \Sigma) \\ &:= \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \end{aligned} \tag{1}$$

2) Reverse Process: After the data is destroyed until x_t, an ANN is trained to try to reverse the forward process, either by predicting the original data x_0 up front, or by predicting another value that can be used to reconstruct the original data, like the added noise ϵ, etc. This process is represented by equation 2, where $\alpha_t = 1 - \beta_t$ and $\overline{\alpha}_t = \prod_{s=1}^{t} \alpha_s$.

$$\begin{aligned} &\rho_\theta(x_{t-1}|x_t) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \\ &:= \mathcal{N}(x_t; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)), \beta_t I) \end{aligned} \tag{2}$$

The correlation between the forward and reverse process to noise and denoise images can be seen in fig. 3. An overview of the training phase for the FCOA of fig. 2 can be seen in fig. 4.

3) Sampling Process: Finally, after the model is trained, the sampling process is used to generate new data. The model learned to reverse the forward process, and so, when giving it pure Gaussian noise, it tries to reverse the noise addition until arriving at the original data. Except this time, there is no

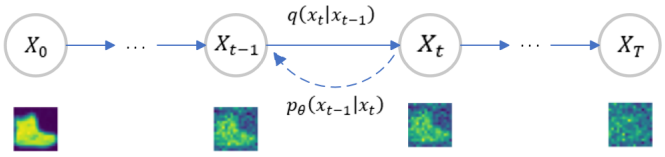


Fig. 3: Correlation between forward and reverse processes [5].

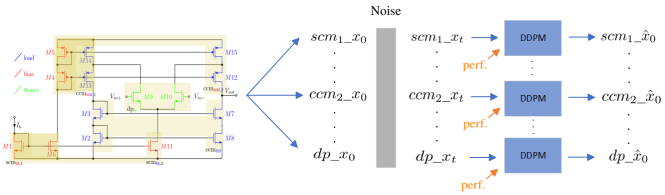


Fig. 4: Training phase for sizing an OpAmp. After the decomposition into sub-blocks, noise is added to the respective sizings. A DDPM is trained for each sub-block to predict the original sizing values. The respective performances of the original sizings are given as guidance.

actual data to begin with, marking this a form of generative AI. This stage is similar to the training phase shown in figure 4, except random noise is given as input to the network instead of the noisy sizings.

It was shown by [5] that DDPMs produce interesting results when applied to the inverse sizing problem. In this work, a similar implementation is considered, with a cosine schedule for β and classifier-free guidance [16], but with 2 key differences:

- instead of predicting the noise ϵ, we implement a model that predicts a velocity equation $v_t = \alpha_t \epsilon - \sigma_t x$ which has been shown to enable a true signal-to-noise ratio of 0 at the last timestep T (meaning that at the last timstep the input of the model is pure Gaussian noise), as opposed to other DDPM parametrizations. This removes an important discrepancy between training and inference [17];
- the backbone of the model implemented is a transformer, with an architecture similar to the "adaLN-Zero" model of [18], as opposed to the simple multi-layer perceptrons implemented by [5]. To predict a more complex value in v, a stronger architecture was required.

IV. Experimental Results

We trained our approach on two different OpAmps, a Folded-Cascode OpAmp (FCOA) previously studied in [3] shown in figure 2 and the Miller OpAmp in figure 5. After

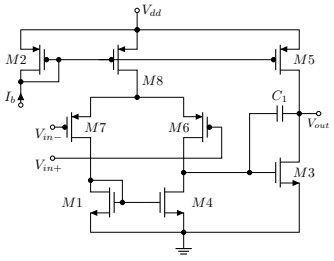
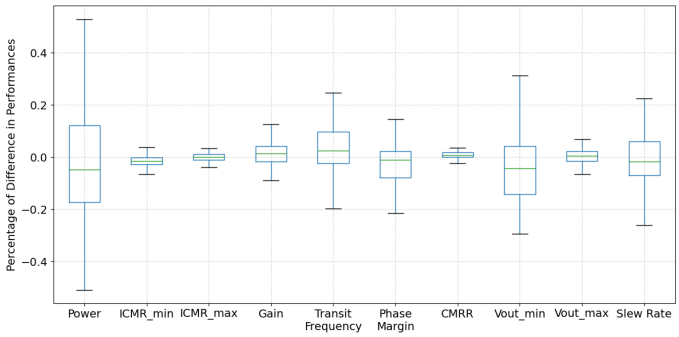
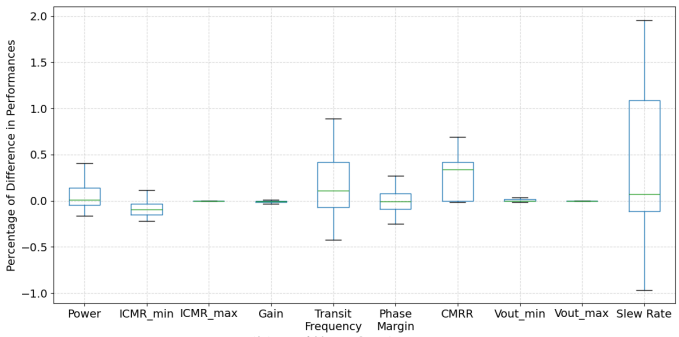


Fig. 5: Miller OpAmp (MOA)



(a) Folded-Cascode OpAmp.



(b) Miller OpAmp.

Fig. 6: Boxplots of the percentage of differences in performances for the test set of both tested OpAmps, after removal of outliers. The innermost line represents the median, the box 50% of the data.

training, we sampled the models for all ground truth sizings in the test sets of both OpAmps, corresponding to 118 for the FCOA and 77 for the Miller OpAmp. After simulating the sampled values, the resulting performances were compared with simulations of the ground truth values. All models have been trained with a sample size between 700 and 1000 elements. While slightly less efficient as [3], the price for added flexibility, the sample sizes are very low.

Boxplots of the relative performance differences are presented in fig. 6, bar outliers. For the FCOA in fig. 6a, most boxes are centered around 0 (i.e. the ground truth), but some performances show a tendency of improvement. For Power, ICMR_{min} and Vout_{min} median and box extend downwards (the predictions' performances are lower than the labels), while Gain and Transit Frequency median and box extend upwards (the predictions' performances are greater than the ground truth sizings). Similar observations can be made for the Miller OpAmp in fig. 6b, with the CMRR and Slew Rate having the biggest improvement. Here, the only performance that showed a tendency toward deterioration was Power. This shows that the models are able to accurately learn the distribution of the circuits' sizing and respective performances, sometimes finding interesting improvements that might be attributed to learned trade-offs between sizes of the different sub-blocks.

In a second experiment we sampled the model for the FCOA 100 times, while keeping the same performance specification. The results can be seen in figure 7 and table I. Most distributions in the performance space in fig. 7 exhibit a gaussian

TABLE I: Estimated mean value and standard deviation of W parameter sizes of the FCOA for 100 samples and same performance specification. It can be seen that the standard deviation varies heavily depending on the device. It appears that the model learns to scale the effects of parameter deviations.

device	mean value	standard deviation
0	8.52	2.0
1	43.26	19.0
2	41.74	17.0
3	14.4	1.0
4	90.98	8.0
5	88.47	23.0
6	41.74	17.0
7	43.26	19.0
8	97.88	55.0
9	97.88	55.0
10	114.78	35.0
11	367.21	22.0
12	367.21	22.0
13	244.11	13.0
14	244.11	13.0

behaviour, centered around the ground truth performance marked by a red star. Table I indicates that the model learns to estimate the different contributions and sensitivities for the individual parameters.

V. Conclusion

We have shown a machine learning approach to not only reproduce optimal sizings for a given specification, but also to help explore points in the nearby performance space. In the scenario of sampling multiple times for the same specification, the sampling can be interpreted as adding random noise to the ground truth sizing. The model learns the corresponding covariance matrix. As the standard deviation is not equal for all devices, this behaviour cannot be mimicked by simply adding noise to an optimal sizing.

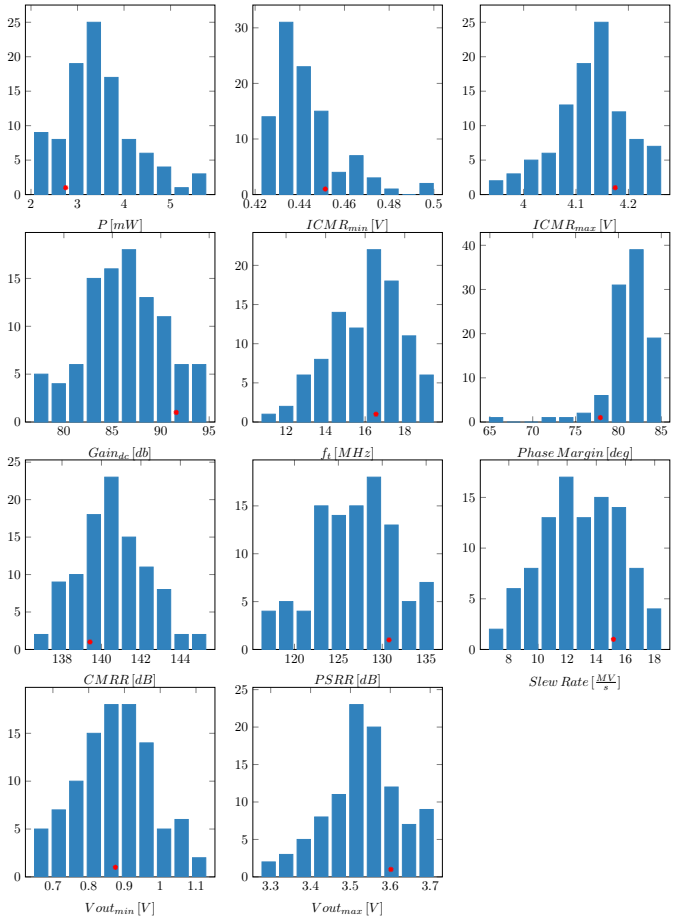


Fig. 7: Histograms for 100 samples of the trained FCOA model with the **same** performance specification. Most distributions are well shaped around the ground truth from the test set. The performance value of the ground truth sizing is marked by a red star.

References

[1] C. Gatermann and R. Sommer, "Teaching the mosfet: A circuit designer's view," in *2022 18th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, 2022, pp. 1–4.

[2] G. Gielen, "Analog synthesis 3.0: Ai/ml to synthesize and test analog ics: hope or hype ?" in *2023 ACM/IEEE 5th Workshop on Machine Learning for CAD (MLCAD)*, 2023, pp. 1–1.

[3] M. Leibl and H. Graeb, "Optimizer-free sizing of opamps leveraging structural and functional properties," in *2024 20th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, 2024, pp. 1–4.

[4] N. Lourenço, E. Afacan, R. Martins, F. Passos, A. Canelas, R. Póvoa, N. Horta, and G. Dundar, "Using polynomial regression and artificial neural networks for reusable analog ic sizing," in *2019 16th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, 2019, pp. 13–16.

[5] P. Eid, F. Azevedo, R. Martins, and N. Lourenço, "Solving the inverse problem of analog integrated circuit sizing with diffusion models," in *2024 20th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, 2024, pp. 1–4.

[6] N. Lourenço, J. Rosa, R. Martins, H. Aidos, A. Canelas, R. Póvoa, and N. Horta, "On the exploration of promising analog ic designs via artificial neural networks," in *2018 15th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, 2018, pp. 133–136.

[7] P.-O. Beaulieu, E. Dumesnil, F. Nabki, and M. Boukadoum, "Analog rf circuit sizing by a cascade of shallow neural networks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 1–1, 2023.

[8] M. Fayazi, M. T. Taba, E. Afshari, and R. Dreslinski, "Angel: Fully-automated analog circuit generator using a neural network assisted semi-supervised learning approach," *IEEE Transactions on Circuits and Systems I: Regular Papers*, pp. 1–14, 2023.

[9] A. F. Budak, D. Smart, B. Swahn, and D. Z. Pan, "Apostle: Asynchronously parallel optimization for sizing analog transistors using dnn learning," in *2023 28th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2023, pp. 70–75.

[10] G. Wolfe and R. Vemuri, "Extraction and use of neural network models in automated synthesis of operational amplifiers," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, no. 2, pp. 198–212, 2003.

[11] K. Hakhamaneshi, M. Nassar, M. Phielipp, P. Abbeel, and V. Stojanovic, "Pretraining graph neural networks for few-shot analog circuit modeling and design," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 7, pp. 2163–2173, 2023.

[12] J. Gao, W. Cao, and X. Zhang, "Rose: Robust analog circuit parameter optimization with sampling-efficient reinforcement learning," in *2023 60th ACM/IEEE Design Automation Conference (DAC)*, 2023, pp. 1–6.

[13] K. Settaluri, Z. Liu, R. Khurana, A. Mirhaj, R. Jain, and B. Nikolic, "Automated design of analog circuits using reinforcement learning," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 9, pp. 2794–2807, 2022.

[14] J. Zhang, J. Bao, Z. Huang, X. Zeng, and Y. Lu, "Automated design of complex analog circuits with multiagent based reinforcement learning," in *2023 60th ACM/IEEE Design Automation Conference (DAC)*, 2023, pp. 1–6.
[15] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020. [Online]. Available: https://arxiv.org/abs/2006.11239
[16] J. Ho and T. Salimans, "Classifier-free diffusion guidance," 2022. [Online]. Available: https://arxiv.org/abs/2207.12598
[17] S. Lin, B. Liu, J. Li, and X. Yang, "Common diffusion noise schedules and sample steps are flawed," 2024. [Online]. Available: https://arxiv.org/abs/2305.08891
[18] W. Peebles and S. Xie, "Scalable diffusion models with transformers," 2023. [Online]. Available: https://arxiv.org/abs/2212.09748