ORACLE: Optimization with Recourse of Analog Circuits including Layout Extraction *

Yang Xu, Lawrence T. Pileggi
Electrical & Computer Engineering Department
Carnegie Mellon University
Pittsburgh, PA 15213
{yangx, pileggi}@ece.cmu.edu

Stephen P. Boyd
Electrical Engineering Department
Stanford University
Stanford, CA 95602
boyd@stanford.edu

ABSTRACT

Long design cycles due to the inability to predict silicon realities is a well-known problem that plagues analog/RF integrated circuit product development. As this problem worsens for technologies below 100nm, the high cost of design and multiple manufacturing spins causes fewer products to have the volume required to support full custom implementation. Design reuse and analog synthesis make analog/RF design more affordable; however, the increasing process variability and lack of modeling accuracy remains extremely challenging for nanoscale analog/RF design. We propose an analog/RF circuit design methodology ORACLE, which is a combination of reuse and *shared-use* by formulating the synthesis problem as an *optimization with recourse* problem. Using a two-stage geometric programming with recourse approach, ORACLE solves for both the globally optimal shared and application-specific variables. Concurrently, we demonstrate ORACLE for novel metal-mask configurable designs, where a range of applications share common underlying structure and application-specific customization is performed using the metal-mask layers. We also include the silicon validation of the metal-mask configurable designs.

Categories and Subject Descriptors

B.7.2 [**Integrated Circuits**]: Design Aids

General Terms

Design, Algorithm

Keywords

Optimization with recourse

1. INTRODUCTION

The IC design and manufacturing costs are increasing to the point that fewer products have the volume required to amortize the large upfront NRE (nonrecurring engineering) costs. This is especially the case for mixed-signal ICs that are designed in sub-100nm technologies, where the technology advances are making application-specific SoC (system-on-chip) designs technically feasible, but the economic realities require even higher product volumes. Design reuse and analog synthesis methodologies [1]-[3], have substantially addressed the design cost and risk challenges. For a given circuit topology and specifications, *simulation based* optimization[1] and *equation based* optimization [2, 3] have been effective for automating the design process. However, the large process parameter variability that is evident for nanoscale technologies along with the complex nature of parasitic coupling can cause the design risk, hence cost, to remain quite high, even for the best synthesis approaches.

For this reason there have been some proposals for configurable analog circuits [4] and software radio circuits [5]. Importantly, such circuits can be pre-characterized for the subtle device properties and coupling parasitics that are difficult to predict prior to layout and manufacturing. These configurable analog/RF circuits reduce the design risk and accommodate the tight time-to-market windows. While the configurable design cost can be high, the cost is shared over multiple applications. The impact of parasitics remains significant, and most importantly, the nonideal behavior of the configuration switches further degrades performance.

We propose a methodology ORACLE, which incorporates the shared-use and reuse benefits of configurable circuits, while offering performance that is comparable to a fully customized design. Instead of a flow to optimize a circuit for a single application, we propose an optimization framework that supports a methodology for configurable designs that "share" common structures. These common structures can then be pre-characterized for subsequent application-specific customization, thereby allowing the second stage of optimization to accommodate extracted layout realities. We formulate our configurable design problem as an *optimization with recourse* problem. If we can formulate each of the sample problems (scenarios) as geometric programming (GP) problem [2, 3], the optimization with recourse problem can be then reduced to a two-stage geometric programing with recourse (GPR) problem and solved efficiently.

Unlike optimization for a single application, the shared common structure can be well characterized via simulation or measurement before it is configured for multiple applications, thereby providing the predictability that is needed for a risk-free robust design. The proposed optimization infrastructure is applicable to configurable designs in general, but here is applied in metal-mask configurable designs, as shown in Figure 1, to produce performance comparable to

*This work was funded in part by C2S2, the MARCO Focus Center for Circuit & System Solution, under MARCO contract 2003-CT-888.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
DAC 2004, June 7-11, 2004, San Diego, California, USA.
Copyright 2004 ACM 1-58113-828-8/04/0006 ...$5.00.

a fully customized application-specific design. The metal-mask configurable circuits are constructed as a two stage process: 1) designing the underlying regular structure, or *implementation fabric*, that can be configured for multiple designs, and extracting the device properties and parasitics; then 2) configuring the implementation fabric by customizing back-end-of-line (BEOL) metal-masks for a particular application. Numerical examples are shown for the metal-mask configurable Low Noise Amplifier (LNA) designs in SiGe and CMOS process. This methodology is validated via the silicon implementation of a BEOL metal-mask configurable RF front-end circuit.

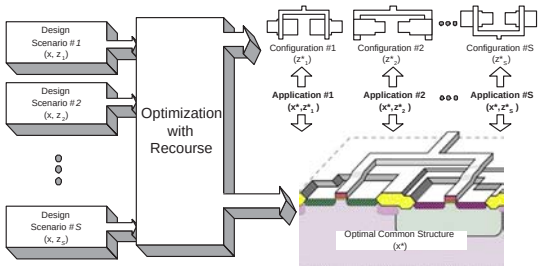

Figure 1: Metal-mask configurable demonstration of ORACLE design methodology.

2. OPTIMIZATION WITH RECOURSE

2.1 Problem definition

We formulate the configurable circuit design as an optimization problem with recourse, also called *two-stage optimization*[6], as shown in Figure 1. In the two-stage optimization, we are to choose the values of two variables, $x \in \mathbf{R}^n$ and $z \in \mathbf{R}^q$, which in conjunction are the design variables for each of S applications, or *scenarios*. The variable x must be chosen *before* the particular scenario s is known; the variable z, however, is chosen *after* the value of the scenario random variable is known. In other words, z is a function of the scenario random variable s. To describe our choice for z, we list the values we would choose under the different scenarios; *i.e.*, we list the vectors $z_1, \ldots, z_S \in \mathbf{R}^q$. Here z_3 is our choice of z when $s = 3$ occurs, and so on. The set of values, $x \in \mathbf{R}^n, \quad z_1, \ldots, z_S \in \mathbf{R}^q$, is called the *policy*, since it tells us what choice to make for x (independent of which scenario occurs), and also, what choice to make for z in each possible scenario. The variable z is called the *recourse variable* (or *second-stage variable*), since it allows us to take some action or make a choice after we know which scenario occurred. In contrast, our choice of x (which is called the *first-stage variable*) must be made without any knowledge of the scenario.

The cost function and constraints depend not only on our choice of variables, but also on a discrete random variable $s \in \{1, \ldots, S\}$, which is interpreted as specifying which of S scenarios occurred. The cost function is given by $f : \mathbf{R}^n \times \mathbf{R}^q \times \{1, \ldots, S\} \to \mathbf{R}$, where $f(x, z_i, i)$ gives the cost when the first-stage choice x is made, second-stage choice z_i is made, and scenario i occurs.

2.2 Two-stage geometric programming with recourse approach

A *geometric program* is an optimization problem which can minimize *posynomial* cost function subject to *posynomial* inequality and *monomial* equality constraints. It can be reformulated as a convex optimization problem: i.e., the problem of minimizing a convex function subject to convex inequality constraints and linear equality constraints. This special type of convex optimization can be *globally* solved with *great* efficiency. We can use efficient interior-point methods to solve the problem, and there is a complete and useful duality, or sensitivity theory for it. Many analog/RF circuit design problems have been successfully formulated as GP [2, 3] and solved with great efficiency.

Once individual optimization problems are formulated as GP, optimization with recourse can be reformulated as a two-stage geometric programming with recourse problem. Suppose that the objective and constraint functions f are posynomial functions of (x, z), for each scenario $i = 1, ..., S$. In order to find an optimal policy, we must solve a *geometric program with recourse* of the form

$$\begin{array}{ll} \text{minimize} & F_0(x, z_1, \ldots, z_S) \\ \text{subject to} & F_j(x, z_i) \le 1, \quad i = 1, \ldots, S, \quad j = 1, \ldots, m, \\ & G_j(x, z_i) = 1, \quad i = 1, \ldots, S, \quad j = 1, \ldots, p, \\ & x_i > 0, \quad i = 1, \ldots, n, \\ & z_i > 0, \quad i = 1, \ldots, q, \end{array}$$

where $F_0, \ldots, F_m$ are posynomial functions and $G_1, \ldots, G_p$ are monomial functions.

The new objective is the expected value of the total cost, and the new constraints are the union of all individual design constraints. The two stage GPR problem can be treated as a much larger GP problem, since for each i, $f(x, z, i)$ can be transformed to be convex in (x, z_i), and $\pi_i \ge 0$, and sums preserve convexity.

The variables in the problem are $x, z_1, ..., z_s$, i.e., the policy. The total dimension of the variables is $n + Sq$ compared with $n + q$ as in one scenario case. The computational burden of solving the large geometric program equivalent for the original problem can be quite prohibitive. This is because we need to solve the set of $n + Sq$(symmetric, positive definite) linear equations $\nabla^2 F \Delta_{nt} = -\nabla F$, where $F(x, z, i) = \begin{pmatrix} F_0 & F_1 & \cdots & F_m \end{pmatrix}^T$, so the cost is around $(1/3)(n+Sq)^3$ flops. As a function of the number of scenarios, this grows like S^3.

Since posynomial function f is twice differentiable function of (x, z) for each scenario $i = 1, \ldots, S$, however, we can exploit the structure to greatly simplify the complexity. This is because the Hessian of $F(x)$ has the block-arrow form, which we can exploit to compute the Newton step efficiently. Therefore, the overall complexity grows *linearly* [6] in S, and this scalability is an important feature of GPR.

3. EXAMPLE: CONFIGURABLE LNA

3.1 Metal-mask configurable case

For analog/RF circuits, metal-mask configurability provides tight interaction between circuit design and layout to achieve performance comparable to fully customized application-specific designs. We choose such BEOL metal-mask configurability to illustrate our ORACLE methodology.

A common implementation fabric is shared by multiple applications through different configurations of metal-mask layers. We select device design variables and metal-mask design variables as first and second stage design variables, and the *scenario* is an application corresponding to a set of specifications. The design is accomplished in two stages: 1) optimal implementation fabric design, and 2) optimal individual metal-mask design. In the first stage, we optimize the structure of the implementation fabric over a domain of multiple applications. Then device and component properties are characterized via post-simulation or potentially

on-wafer measurement. By doing this, we can use the extracted information to center the final design. In the second stage, accurate device and component models are plugged into the original problem and re-solved to achieve the individual metal-mask designs.

In the formulated optimization problem, the new constraints include all individual design constraints. While selecting the new objectives, we have several choices depending on the design goal. We can minimize the expected objective (average or weighted average), or the maximum objective among all the scenarios, which would leave large margin for most scenarios. Another choice is to minimize the maximum *design surcharges*, which is defined as the performance difference between mask configurable design and independent design. The independent design is a full custom design for each design scenario, and this is the best we can get for each scenario. In this way, we can minimize the cost to achieve mask configurability.

Since the GP formulation of each individual design is the basis for the GPR formulation of mask configurable design, in the following example we will describe the separation of design variables and list design constraints used in a single design GP formulation. Once the GP formulation of each design is obtained, the GPR formulation of the entire metal-mask configurable design can be readily obtained as discussed in §2.2.

3.2 SiGe LNA design

A Low Noise Amplifier is an important building block for any RF or wireless receiver. SiGe is a promising process for wireless communication ICs due to the quality of having a higher performance compared to CMOS and the ability of integration with standard CMOS. One drawback of SiGe is its high cost of masks and manufacturing, which is exactly addressed by our BEOL mask configurable methodology.

The specific SiGe LNA topology we consider in this paper is shown in Figure 2. This topology has been widely used due to its lower noise performance compared with other topologies. This circuit consists of an input tune loop followed by a cascode common-emitter transconductance stage with tuned output loop. Since the LNA is part of an RF front-end, it is also required to match the impedance with input and output to maximize power transfer.

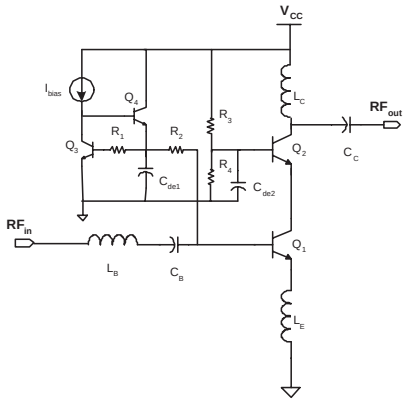

Figure 2: Simplified SiGe LNA schematic.

3.3 Design variables

There are 12 independent physical variables in a single design that we would like to optimize for the simplified SiGe LNA design. These design variables are related to the sizing and biasing of the input BJT and inductors. The cascode transistor Q_2, the DC biasing circuitry (R_1, R_2, R_3, R_4, Q_3 and Q_4) and decoupling capacitors (C_{d1} and C_{d2}) are heuristically sized for best matching and power consumption performance.

The 12 independent design variables are divided into two categories: *device design variables* and *metal-mask design variables*, which correspond to the front-end and the back-end of the SiGe fabrication processes. The 5 device design variables are: the emitter length l_E and width w_E of input transistor Q_1, and the outer dimension D_1, D_2, D_3 of three inductors L_E, L_B and L_C. These variables are restricted to take values on a discrete grid. Since the layout grid in a modern technology is very small, we ignore the grid constraints in this paper and consider these variables to be *positive real numbers*. The 7 metal-mask design variables are: m_1 is the number of devices of the the same geometry used in parallel for input BJT, which should be *integers*; the number of turns n_1, n_2, n_3 of three inductors, which should be *integer multiples* of 0.25 (quarter turns); the value of the input and output tune capacitors C_B and C_C, and the collector current I_C, which are considered as *positive real numbers*.

The supply voltages $V_{\rm cc}$ and *gnd*, and the various process and technology parameters associated with the SiGe models are assumed to be fixed value.

3.4 Design specifications and parameters

In order to cast the design of LNAs as GP, we need to show that the LNA design specifications can be posed as posynomial functions of the design variables. Being able to write circuit equations in posynomial form is the key to use geometric programming to design analog circuits. To achieve equations in posynomial form, one needs to make reasonable approximations. Since our equations agree with simulated results very closely, we conclude that our approximations are valid.

The LNA was designed to achieve simultaneous noise and power match using the method reported in [7, 8]. Under the power consumption constraint, it is desirable to achieve gain with input and output impedance match, while maintaining the minimum noise and distortion level, also minimizing the silicon area. Therefore, when formulating the GP problem, we minimize area subject to the following constraints:

- Noise match ($F \leq F_{\max}$)
- Input impedance match ($S_{11} \leq S_{11}^{\max}$)
- Gain requirement ($G \geq G_{\min}$)
- Output impedance match ($S_{22} \leq S_{22}^{\max}$)
- Nonlinear distortion requirement ($I \geq I_{\min}$)
- Power constraint ($P \leq P_{\max}$)

Where F is the noise figure, S_{11} is the input reflection coefficient, G is the gain, S_{22} is the output reflection coefficient, I is the third-order intercept point (IIP3) and P is the power consumption. The corresponding minimum or maximum are the upper bound or lower bound of each specification.

We use the Gummel-Poon BJT model to derive the initial design equations, where electrical elements in this model are monomial expressions of physical design variables. A simple monomial fitting [6] technique can be employed to fit the VBIC95 BJT model to achieve better accuracy. For on-chip inductors, all the elements in the lumped electrical model can be expressed as monomial or posynomial function of layout variables, as shown in [3]. Due to space limitation, derivations for the posynomial design equations are not included, readers can refer to [8] for details.

The design parameters considered in the LNA design is the center frequency ω_0 and source impedance R_s, which is usually 50Ω in RF systems.

4. NUMERICAL RESULTS

4.1 Independent design and verification

We use the design variables described in §3.3 and design constraints listed in §3.4 for a SiGe LNA example with 2.1GHz center frequency. The resulting geometric program has 12 variables, and 28 inequality constraints. The formulated GP problem was solved efficiently by the MOSEK toolbox in the order of a millisecond.

The target specifications and the performance achieved by this design, as predicted by the program, are summarized in Table 1. For a given circuit topology and a set of design specifications, this is the best we can get and used as the benchmark. Note that some constraints are tight (power consumption, center frequency and gain), while some constraints are not (Noise figure, S-parameters and IIP3).

Specification	Targeted Value	GP Achieved	Simulated
S_{11}	$\leqslant -10$dB	-19dB	-13.25dB
$Gain$	$\geqslant 20$dB	20dB	21dB
S_{22}	$\leqslant -10$dB	-21dB	-11dB
NF	$\leqslant 2$dB	1.61dB	1.84dB
IIP3	$\geqslant -5$dBm	-2.9dBm	-3.6dBm
Power	$\leqslant 12.5$mW	12.5mW	12.5mW

Table 1: GP optimized results and SpectreRF simulated performance.

The simulation results are also shown in Table 1. We used Cadence's SpectreRF as the simulator with advanced device models. We conclude that GP optimization results and simulation results have good agreement with each other.

4.2 Center frequency configurable designs

Next we consider a mask configurable SiGe LNA design at 13 different center frequencies, ranging from 900MHz to 2.1GHz with separation of 200MHz. Other specifications are the same as listed in Table 1. We use the average noise figure of all mask configurable designs as the cost function. The resulting problem has 96 variables and 364 constraints. The optimization process generates 13 metal-mask configurable designs with the same first-stage design results and 13 sets of second-stage design results.

Independent designs are obtained for each scenario for comparison. The achieved noise figure for independent designs and configurable designs as shown in Figure 3. We can see that the NF of configurable designs are very close to independent designs and the maximum NF surcharge is less than 0.1dB.

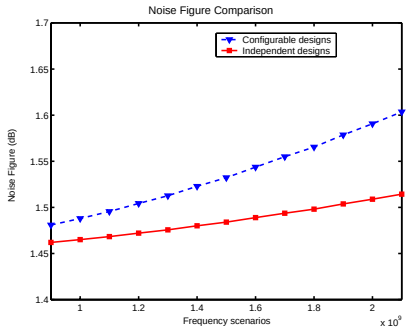

Figure 3: Noise performance comparison.

Since the circuit performance of configurable designs is very close to independent designs, the only penalty we pay for such flexibility is the silicon area. This is inevitable because the area of the implementation fabric would be larger than the maximum of all independent designs. The extent of the area penalty is a trade-off with the amount of design risk. For example, in an LNA, the inductor areas are quite dominant. Reserving a fixed area for all inductors is the lowest risk approach, but incurs the largest area penalty.

4.3 Power and gain configurable designs

As a second example we vary the power and gain specifications and observe the design space trade-offs for mask configurability. The center frequency is fixed at 5.25GHz. The power spec varies from 12.5mW to 20mW by every 0.5mW, while the gain spec varies from 10dB to 24dB by every 2dB. There are 8 different gain requirements and 16 different power constraints. Therefore, in total we generate 128 design instances using the average NF of all designs as the cost function.

The achieved noise performance and NF surcharge of the mask configurable design are shown in Figure 4. The plots show that the noise performance is more sensitive to power consumption. The NF surchage of mask configurable design is less than 0.1dB. In this way, design space exploration can be achieved in the early stages for the entire system design.

It is worth mentioning that the 128 designs with 901 variables and 3584 constraints are solved in 1.5 seconds, using the 1.4GHz 256MB memory Pentium PC.

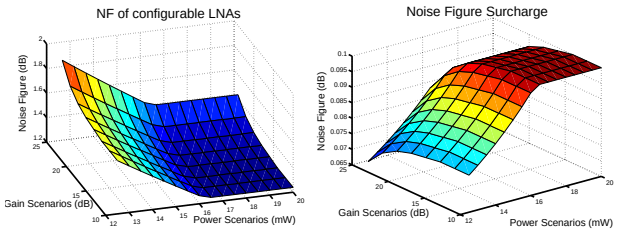

Figure 4: NF of configurable LNAs and NF surcharges.

5. CONCLUSIONS

Configurable analog/RF circuits can be used to reduce design risk and manufacturing cost. In this paper, we proposed a novel design methodology and supporting optimization infrastructure for such circuits. Our methodology and optimization procedure is applied in a set of metal-mask configurable LNA designs. Numerical examples demonstrate that comparable performance can be achieved with little design surcharge. The silicon implementation to validate the metal-mask configurable designs can be found in [8].

6. REFERENCES

[1] H. Liu, A. Singhee, R. Rutenbar, L. Carley. *Remembrance of Circuits Past: Macromodeling by Data Mining in Large Analog Design Spaces* IEEE DAC, June 2002.

[2] M. Hershenson, S. Boyd, and T. H. Lee, *Optimal design of a CMOS op-amp via geometric programming* IEEE TCAD March 2001

[3] M. Hershenson, S.S. Mohan, S. P. Boyd and T. H. Lee *Optimization of Inductor Circuits via Geometric Programming* IEEE DAC, June 1999

[4] G. Gulati, H. Lee *A Low-Power Reconfigurable ADC* IEEE JSSC Vol. 36, No.12, Dec. 2001

[5] J. Mitola *Software Radio Architecture: Object-Oriented Approaches to Wiresless Systems Engineering* New York: Willey, 2000.

[6] S. Boyd, L. Vandenberghe, *Convex Optimization* Cambridge University Press, 2003.

[7] O. Shana'a, I. Linscott, L. Tyler, *Frequency-Scalable SiGe Bipolar RF Front-End Design* IEEE JSSC Vol. 36, No. 6, June 2001.

[8] Y. Xu, C. Boone and L. Pileggi *Metal-mask configurable RF Front-end* 2004 RFIC symposium June, 2004