CMOS Op-Amp Sizing Using a Geometric Programming Formulation

Pradip Mandal and V. Visvanathan, *Member, IEEE*

***Abstract*—The problem of CMOS op-amp circuit sizing is addressed here. Given a circuit and its performance specifications, the goal is to automatically determine the device sizes in order to meet the given performance specifications while minimizing a cost function, such as a weighted sum of the active area and power dissipation. The approach is based on the observation that the first order behavior of a MOS transistor in the saturation region is such that the cost and the constraint functions for this optimization problem can be modeled as posynomial in the design variables. The problem is then solved efficiently as a convex optimization problem. Second order effects are then handled by formulating the problem as one of solving a sequence of convex programs. Numerical experiments show that the solutions to the sequence of convex programs converge to the same design point for widely varying initial guesses. This strongly suggests that the approach is capable of determining the globally optimal solution to the problem. Accuracy of performance prediction in the sizing program (implemented in MATLAB) is maintained by using a newly proposed MOS transistor model and verified against detailed SPICE simulation.**

***Index Terms*—Cell-generation, device-models, optimization, transistor-sizing, VLSI.**

I. Introduction

THE CURRENT trend in microelectronics is to integrate a complete system that previously occupied one or more boards on one or a few chips. Although most of the functionality in an integrated system is implemented in digital circuitry, analog circuits are needed to interface between the core digital system and the real world. Therefore, to realize an integrated system on a single chip, the digital and analog circuits are combined together. This integration of analog and digital circuits results in so called mixed-signal integrated circuits which have a large market of applications in the telecom, consumer products, computing, and automotive sectors.

Increase of design complexity and, at the same time, demand of design cycle time reduction due to highly competitive market can be managed only by the use of computer aided design. Though in an integrated system, the analog circuitry occupies a small physical area compared to the digital counterpart and becomes the bottleneck in design time reduction.

Manuscript received July 20, 1998; revised July 18, 1999 and April 7, 2000. An earlier version of this paper was presented at the 12th International Conference on VLSI Design, Goa, India, January 1999. This paper was recommended by Associate Editor K. Mayaram.

P. Mandal was with the Indian Institute of Science, Bangalore, India. He is now with Philips Semiconductors, Bangalore, India (e-mail: pradip.mandal@philips.com).

V. Visvanathan was with the Indian Institute of Science, Bangalore, India. He is now with Cadence Design Systems, Allentown, PA 18195 USA (e-mail: vish@cadence.com).

Publisher Item Identifier S 0278-0070(01)00349-9.

The main reason for this is that the number of performance functions in an analog circuit is much larger than that in a digital circuit. Further, analog performances are very sensitive to the design variables and variation in the performance across the design space is quite high. In other words, the analog design problem is a complex tradeoff problem that is knowledge intensive. However, the research community has been aggressively working for computer-aided analog design. A good survey of recent analog synthesis techniques is available in [1].

Existing approaches of automatic circuit sizing are broadly classified into three main categories, namely *knowledge-based optimization*, *simulation-based optimization*, and *analytical equations-based optimization*. In this context, we note that, unlike in the digital domain, the standard cell based approach [2] is quite restrictive in the analog domain.

Since analog design requires detailed circuit knowledge, a major approach of implementing an analog synthesis tool has been the knowledge-based approach. Some of the existing tools which follow this approach are BLADES [3], OASYS [4], and IDAC [5] and [6]. However, the application of this approach has been limited due to requirement of having to codify extensive circuit knowledge and design heuristics.

On the other hand, DELIGHT.SPICE [7], ASTRX/OBLX [8], FRIDGE [9], MAELSTROM [10], and ANACONDA [11] use the simulation-based optimization approach. This approach does not require much circuit knowledge. Hence, the main advantage of this approach is that a wide range of circuits can be synthesized. However, the basic limitation comes from the requirement of costly circuit simulation in each iteration of the optimization algorithm.

To reduce the CPU time of optimization-based techniques, the third approach is analytical equations-based optimization, where the circuit performances are evaluated using analytical equations. OPASYN [12] uses simple analytical equations of op-amp performance. OPTIMAN [13] uses a symbolic simulator, ISAAC [14], to get the analytical models of the ac performances of a circuit. However, in [13], analytical models for dc and transient performances have to be provided by an expert designer. In circuit sizing, the use of a single weighted cost function [12] is inadequate since sizing is a constrained optimization problem with complex tradeoffs among the constraints. With simulated annealing [13], the drawback is that it is computationally intensive (even with the use of analytical equations of performances) and cannot be realistically used in an interactive setting.

It therefore appears that an analytical equation based constrained optimization method is the most promising approach for automatic circuit sizing. However, the existing technique that uses this approach [15] suffers from the drawback that it needs expert designer knowledge to sequentially introduce the

constraints. If this is not done, the method may fail to provide even a feasible design point. Further, any optimal point that is provided is only a local optimum design point. Thus, the objective of this work is to propose an analytical equation-based constrained optimization method that is fast, robust, and provides the globally optimal design point.

The main purpose of developing an automatic sizing tool is that the design space of a circuit will be quickly explored to find a design point at which the circuit satisfies the required specification and, at the same time, some cost (e.g., area, power) is minimized. This places a number of stringent requirements on the tool, which are enumerated below.

- In order to support an iterative design methodology, the tool should be fast.
- The final solution point should be independent of the initial guess.
- The optimization technique should be extremely robust. In particular, the designer should not be burdened with the task of tuning the optimizer.

In this paper, a circuit sizing method for CMOS op-amps is proposed, which meets these three requirements as verified via a prototype implementation applied to a number of two-stage op-amps. The effectiveness of the method is due to the use of convex optimization techniques via a geometric programming formulation. It is similar to the recent work described in [16]–[19], which appear to have been done independently and in parallel to this work [20]. However, compared to [16], our technique addresses second-order effects and is far less restrictive in terms of the range of MOS models it supports. This is because, unlike [16]—where the problem is solved as a single geometric program—we show that it is necessary to formulate and solve the problem as a sequence of (convex) geometric programs. In other words, the formulation of [16] is just the first step of our approach. This difference is crucial in the context of submicrometer technology, since the models necessary to support our methodology need to support the geometric programming paradigm only in their first order effects, while the models of second order behavior can be arbitrary. The sequential convex programming formulation is made possible by judiciously combining various techniques, most notably relaxed dc formulation [15] and casual dc analysis [22] with the powerful geometric programming paradigm [23]–[25].

The organization of the rest of this paper is as follows. In Section II, the basic approach of op-amp sizing is described. In Section III, with the first-order Shichman–Hodges (S–H) model, the op-amp sizing problem is formulated as a convex programming problem. Section IV describes how the higher order effects can be captured through an iterative approach. In Section V, the sizing technique is applied on a large number of op-amp circuits. Experimental results are provided in Section VI. By the use of a new accurate MOS model, the accuracy of performance prediction has been increased in Section VII. Section VIII provides a discussion, while Section IX summarizes the work that is described in this paper.

II. The Basic Approach

The proposed op-amp sizing technique is based on three basic ideas. The first one is that of extrapolating the saturation region characteristic of a transistor into its linear region of operation. The use of the extrapolated transistor characteristic results in a design formulation that is simple and robust.

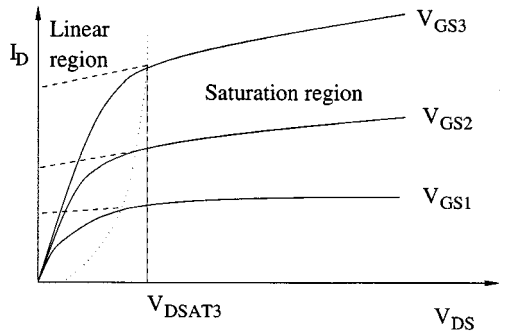

Fig. 1. MOS transistor characteristic.

The second idea on which our sizing technique is based is that the first-order and higher order behavior of the MOS transistor have been separated. With the first-order model, it is shown that the constraint functions and the objective function for the design optimization problem that has been addressed are *posynomials* in the design variables, namely transistor sizes and biases. In other words, with the first-order model, the op-amp synthesis problem is formulated as a geometric programming problem. With a logarithmic transformation, a geometric programming problem becomes a convex optimization problem.

Finally, the higher order effects are handled by iteratively updating the first-order model parameters using the higher order transistor models, and by solving a sequence of convex programs. The technique of iteratively updating the first-order model parameters is elaborated in Section IV. The parameter update is based on iteratively refining the dc operating point. Section III-A describes a systematic approach for finding the dc operating point.

A. Extrapolated Transistor Characteristic

Usually transistors in an analog circuit are biased in the saturation region where the drain conductance is low, which helps to get high ac performance. Therefore, for analog design analysis, one can use only the saturation region characteristic. Systematic use of the saturation region characteristic is pictorially illustrated in Fig. 1. The continuous curves are the actual I_D–$V_{\rm DS}$ characteristic curves of a transistor with different $V_{\rm GS}$. The dotted line divides the whole region of operation into two parts, linear (left) and saturation (right) regions. The dotted line cuts a characteristic curve at a point where $V_{\rm DS}$ is equal to the drain saturation voltage $V_{D\rm SAT}$. By extrapolating the saturation region characteristic curves into the linear region, as shown by the dashed lines in Fig. 1, a set of artificial characteristic curves are obtained. These characteristic curves are simple and smooth over the entire region of operation. For analog design analysis, rather than using the actual characteristic curves, these artificial characteristic curves can be used. However, to ensure that at the final design point the transistors are actually in saturation, it is necessary to satisfy the constraint $V_{\rm DS} \geq V_{D\rm SAT}$ for all transistors. In the following section, we propose a systematic approach of circuit analysis for finding a set of design space constraints which helps to satisfy the constraint on $V_{\rm DS}$.

If the S–H model is used for the extrapolated device characteristic curves, over the entire region of operation transconductance g_m, drain conductance g_d, and effective gate-to-source voltage V_{GT} ($=V_{\rm GS}–V_T$) are, respectively,

$$\left.\begin{aligned} g_m &= 2\left[k' I_D \frac{W}{L}(1+\lambda V_{\rm DS})\right]^{1/2} \\ g_d &= \frac{\lambda_c I_D}{L(1+\lambda V_{\rm DS})}, \qquad \lambda_c \text{ proportionality constant} \\ \text{and } V_{GT} &= \left[\frac{I_D}{k'\dfrac{W}{L}(1+\lambda V_{\rm DS})}\right]^{1/2} \end{aligned}\right\} \tag{1}$$

Note that, assuming constant $(1+\lambda V_{\rm DS})$, the device model parameters g_m, g_d, and $(V_{\rm GS}–V_T)$ are product of power (PoP) functions of W/L, L, and I_D. A PoP function is the product of a positive coefficient and various variables that are raised to some power (any real number). This is the key information on which our sizing formulation is based.

The extrapolated device characteristic curves and the design space constraints are used in our proposed sizing technique.

B. Geometric Programming Problem

A geometric programming problem is of the following form:

$$\left.\begin{array}{ll} \text{minimize} & f_o(x) \\ \text{subject to} & g_i(x) \le 1 \\ \text{and} & x_j > 0 \end{array}\right\}. \tag{2}$$

The objective function $f_o(x)$ and each one of the constraint function $g_i(x)$ are posynomials of the design variables x_j's. In equation form, a posynomial function is

$$f(x) = \sum_i C_i \prod_j x_j^{\alpha_{ij}} \tag{3}$$

where C_i's are positive coefficients and α_{ij}'s are arbitrary real numbers. Note that in the above equation there are a number of PoP terms. The coefficient C_i is referred to as the constant of the corresponding PoP term. For any i, the corresponding α_{ij}'s are referred to as the powers of the ith PoP term. With a logarithmic transformation on the design variables, the posynomial function is

$$f(z) = \sum_i C_i e^{(\sum_j \alpha_{ij} z_j)} \tag{4}$$

where $z_j = \ln x_j$. It can be easily shown that $f(z)$ is a convex function [23] of the transformed variables z_j's. Therefore, with the logarithmic transformation on the design variables, the geometric programming problem becomes a convex programming problem. In the following, it is shown that the op-amp sizing problem is a geometric programming problem. As a result, the op-amp design problem becomes a convex programming problem. The well known property of a convex programming problem is that any of its local minima is also a global minimum.

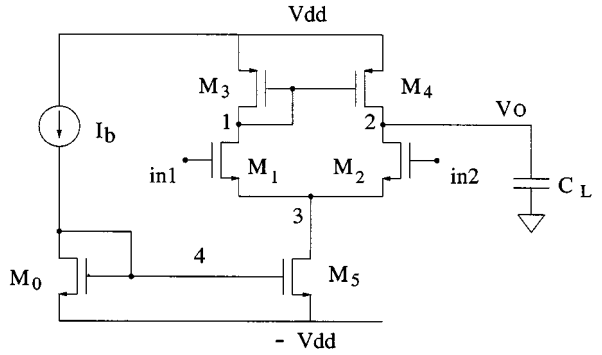

Fig. 2. Simple op-amp, a running example.

III. Sizing Formulation with a Running Example

Various steps of op-amp sizing are described here by using a simple running example, shown in Fig. 2. Consider the design optimization problem as

minimize a weighted sum of gate area and power while

$$\begin{aligned} \text{low frequency gain, } A(0) &\ge A_{\rm SPEC} \\ \text{unity gain frequency, } UGF &\ge UGF_{\rm SPEC} \\ \text{slew rate, } SR &\ge SR_{\rm SPEC} \\ \text{common mode range, CMR} &\ge \text{CMR}_{\rm SPEC} \\ W_{\min} \le W_i &\le W_{\max} \\ L_{\min} \le L_i &\le L_{\max}. \end{aligned}$$

In this design problem, L and W/L of the transistors and the bias current I_b are the design variables. In the op-amp, M_1, M_2 and M_3, M_4 are matched pairs. The various intermediate steps of the design formulation follow.

A. DC Analysis

In this step, the circuit is analyzed to determine analytical equations that give its dc operating point. These equations are solved to get the dc operating point, which is then used for predicting its performance.

It is observed that in a CMOS analog circuit, various node voltages can be defined by the gate-to-source voltages of various transistors. Further, the gate-to-source voltages of the transistors can be determined by its size and its dc current. Finally, the dc current through all the transistors can be essentially determined by only a few transistors, which we refer to as current source transistors.

An exception to the approach described above is the output node, whose voltage usually cannot be determined from the $V_{\rm GS}$'s of various transistors. However, in actual applications, op-amps are used in a closed loop configuration, where the output node voltage is stable at a predetermined value (usually zero) [22].

In the example op-amp, the bias stage current is I_b and the drain current of the current source transistor M_5 is

$$I_{D5} = k_5'\left(\frac{W_5}{L_5}\right)(1+\lambda_5 V_{\rm DS5})\left[\frac{I_b}{k_0'(1+\lambda_0 V_{\rm DS0})}\left(\frac{L_0}{W_0}\right)\right]. \tag{5}$$

TABLE I
$V_{\rm DS}$ AND $V_{\rm SB}$ OF VARIOUS TRANSISTORS

Transistors	V_{DS}	V_{SB}
M_1, M_2	$V_1 - V_3$	$V_3 + V_{dd}$
M_3, M_4	$V_{dd} - V_1$	0.0
M_5	$V_3 - (-V_{dd})$	0.0
M_0	$V_4 - (-V_{dd})$	0.0

The current I_{D5} determines the current through the other transistors. Since the transistor pairs M_1, M_2 and M_3, M_4 are matched, their drain currents are

$$I_{D1} = I_{D2} = I_{D3} = I_{D4} = \frac{I_{D5}}{2}. \tag{6}$$

From (1), the gate-to-source voltage of all transistors can be written as

$$V_{{\rm GS}i} = V_{Ti} + \left[\frac{I_{Di}}{k_i'(1+\lambda V_{{\rm DS}i})}\left(\frac{L_i}{W_i}\right)\right]^{1/2}. \tag{7}$$

Now, using a casual dc analysis which is described in [22], various node voltages are expressed in terms of $V_{{\rm GS}i}$'s and various biases.

The analysis starts from the positive supply and moves toward the negative supply rail. Consider the transistor M_3. The drain-to-source and gate-to-source voltages of the transistor M_3 are[1]

$$V_{\rm DS3} = V_{\rm GS3} = V_{\rm dd} - V_1.$$

Therefore,

$$V_1 = V_{\rm dd} - V_{\rm GS3}. \tag{8}$$

Note that the node voltages V_1 and V_2 are the same, since $V_{\rm in1}$ and $V_{\rm in2}$ are assumed to be at the same dc bias. Now considering the transistors M_1 and M_5, respectively,

$$V_3 = V_{\rm in1} - V_{\rm GS1} \tag{9}$$

and

$$V_4 = -V_{\rm dd} + V_{\rm GS5}. \tag{10}$$

These node voltages are now used to express the $V_{{\rm DS}i}$'s and $V_{{\rm SB}i}$'s of the various transistors as given in the Table I.

The above analysis shows that (5)–(10) and Table I represent a complete set of nonlinear equations whose solution provides accurate node voltages. An effective way of solving this set of nonlinear equations is through a fixed-point scheme. The overall method of finding node voltages is shown in Fig. 3.

In the first step, channel length modulation factor, $(1 + \lambda V_{{\rm DS}i})$, and threshold voltage, V_T, of various transistors are updated based on the $V_{{\rm DS}i}$ and $V_{{\rm SB}i}$ values in the previous iteration. Note that in the first iteration, $(1+\lambda V_{{\rm DS}i})$'s are assumed to be one and all the V_{Ti}'s are taken to be V_{To}. Next, the drain current of all current source transistors in the circuit (e.g., I_{D5} in the example circuit) are found. These currents are then used to find the current through the remaining transistors in the circuit.

[1] $V_{\rm DS3}$ and $V_{\rm GS3}$ represent, respectively, the magnitudes of drain-to-source and gate-to-source voltages of the p-transistor M_3.

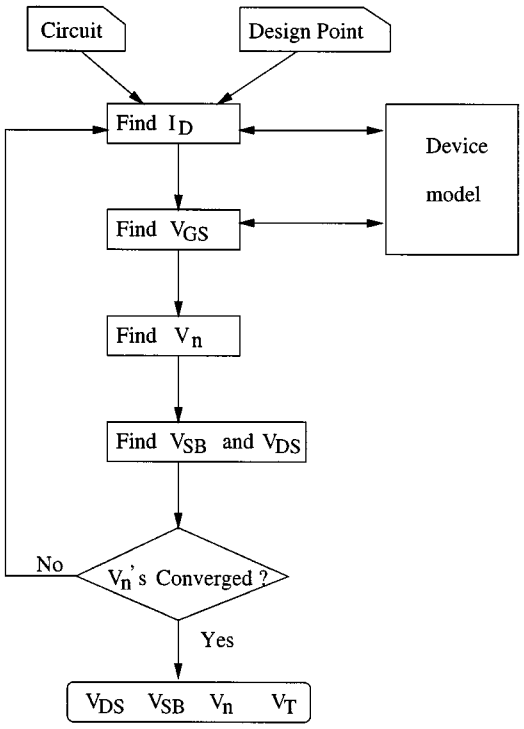

Fig. 3. Iterative approach of finding dc operating point.

In the next step, $V_{{\rm GS}i}$'s are determined from I_{Di}'s, W_i/L_i's, $(1 + \lambda V_{{\rm DS}i})$'s, and V_{Ti}'s of the transistors. In the subsequent step, from the $V_{{\rm GS}i}$'s, various node voltages are determined by using (5)–(10). In the final step, $V_{{\rm DS}i}$'s and $V_{{\rm SB}i}$'s are evaluated using their equations, which are given in Table I.

The values of the $V_{{\rm DS}i}$'s and $V_{{\rm SB}i}$'s, which are obtained at the end of each iteration, are then used in the next iteration to get a more accurate estimate of the dc operating point. The terminating condition is that all node voltages in two consecutive iterations are very close. It is found that this formulation results in a highly contractive fixed-point scheme that converges very fast. This is because, across the design space, the value of $(1 + \lambda V_{{\rm DS}i})$'s are close to one and the back bias dependency function of the threshold voltage is strictly monotonic with small slope.

B. Design Space Constraints

Here we find the design space constraints by which all the transistors are kept in saturation, i.e., away from subthreshold and linear regions. To keep all the transistors away from the subthreshold region with a margin of $\epsilon_{\rm SUB}$, the constraints to be satisfied are

$$\frac{\epsilon_{\rm SUB}}{V_{GTi}} \leq 1. \tag{11}$$

On the other hand, to keep a transistor away from the linear region, we require

$$\left.\begin{array}{r} V_D \geq V_G - V_{T\rm SAT} \text{ for n-type} \\ \text{or } V_D \leq V_G + V_{T\rm SAT} \text{ for p-type} \end{array}\right\} \quad \text{where } V_{T\rm SAT} = V_{\rm GS} - V_{D\rm SAT}. \tag{12}$$

Note that in the S–H MOS model, $V_{T\mathrm{SAT}}$ is equal to V_T. However, this new notation is introduced for ease of extending the design formulation for any other MOS model, such as the one described in [20].

In the example circuit, the gate and drain voltages of the transistors M_3 and M_4 are the same. Therefore, these two transistors are always in saturation. However, to keep M_1 (and M_2) in saturation, we require $V_1 \geq V_{\mathrm{in1}} - V_{T\mathrm{SAT1}}$. In quiescent condition, $V_{\mathrm{in1}} = 0$. Further, from (8), $V_1 = V_{\mathrm{dd}} - V_{\mathrm{GS3}}$. Therefore, the design inequality is

$$(V_{\mathrm{dd}} - V_{T3} + V_{T\mathrm{SAT1}}) \geq V_{GT3}$$

or

$$\frac{1}{(V_{\mathrm{dd}} - V_{T3} + V_{T\mathrm{SAT1}})} \cdot V_{\mathrm{GT3}} \leq 1. \tag{13}$$

To keep transistor M_5 away from linear region, we require $V_3 \geq V_4 - V_{T\mathrm{SAT5}}$. Using the expressions of V_3 and V_4 in (9) and (10), we get

$$\begin{aligned} V_{\mathrm{in1}} - V_{\mathrm{GS1}} &\geq -V_{\mathrm{dd}} + V_{\mathrm{GS5}} - V_{T\mathrm{SAT5}} \\ V_{GT5} + V_{GT1} &\leq V_{\mathrm{dd}} + V_{\mathrm{in1}} - V_{T1} - V_{T5} + V_{T\mathrm{SAT5}} \\ \frac{(V_{GT5} + V_{GT1})}{V_{\mathrm{dd}} + V_{\mathrm{in1}} - V_{T1} - V_{T5} + V_{T\mathrm{SAT5}}} &\leq 1. \end{aligned} \tag{14}$$

Note that, assuming constant V_{Ti}'s and $V_{T\mathrm{SAT}i}$'s, (11), (13), and (14) are posynomial functions of V_{GTi}'s, which are constrained to be less than or equal to one. As will be made clear in the following sections, this is an important step toward our formulation.

Finally, to keep the transistor sizes within the specified limit, the following inequalities should be satisfied:

$$\left.\begin{aligned} \left(\frac{W_i}{L_i}\right)\frac{L_i}{W_{\max}} &\leq 1 \\ \left(\frac{L_i}{W_i}\right)\frac{W_{\min}}{L_i} &\leq 1 \\ \frac{L_i}{L_{\max}} &\leq 1 \\ \frac{L_{\min}}{L_i} &\leq 1 \end{aligned}\right\}. \tag{15}$$

C. Performance Constraints and Objective Function

The low frequency gain of the op-amp is $g_{m1}/(g_{d1} + g_{d3})$. Therefore, to meet the gain specification

$$A_{\mathrm{SPEC}} \cdot \frac{g_{d1}}{g_{m1}} + A_{\mathrm{SPEC}} \cdot \frac{g_{d3}}{g_{m1}} \leq 1. \tag{16}$$

The unity gain frequency of the op-amp is g_{m1}/C_L. So to get the specified UGF

$$\frac{C_L \cdot UGF_{\mathrm{SPEC}}}{g_{m1}} \leq 1. \tag{17}$$

Slew rate of the op-amp is I_{D5}/C_L. So the constraint to get the specified slew rate is

$$\frac{C_L \cdot SR_{\mathrm{SPEC}}}{I_{D5}} \leq 1. \tag{18}$$

The negative common mode range of the op-amp is $\mathrm{CMR}^- = -V_{\mathrm{dd}} + V_{\mathrm{GS0}} + V_{\mathrm{GS1}} - V_{T\mathrm{SAT5}}$. Note that CMR^- has negative value. Therefore, to get the specified CMR

$$-V_{\mathrm{dd}} + V_{\mathrm{GS0}} + V_{\mathrm{GS1}} - V_{T\mathrm{SAT5}} \leq -\mathrm{CMR}_{\mathrm{SPEC}}$$

or

$$\frac{(V_{GT0} + V_{GT1})}{(V_{\mathrm{dd}} - V_{T0} - V_{T1} + V_{T\mathrm{SAT5}} - \mathrm{CMR}_{\mathrm{SPEC}})} \leq 1. \tag{19}$$

On the other hand, the positive common mode range is $\mathrm{CMR}^+ = V_{\mathrm{dd}} - V_{\mathrm{GS3}} + V_{T\mathrm{SAT1}}$, which should be greater than $\mathrm{CMR}_{\mathrm{SPEC}}$. In other words

$$\frac{V_{GT3}}{(V_{\mathrm{dd}} - V_{T3} + V_{T1} - \mathrm{CMR}_{\mathrm{SPEC}})} \leq 1. \tag{20}$$

The objective function is a weighted sum of total effective gate area (in micrometers $\times$ micrometers) and total quiescent current of the op-amp (in microamperes), which is given by

$$f_0 = w_1 \sum_i \left(\frac{W_i}{L_i}\right) L_i^2 + w_2(I_b + I_{D5}). \tag{21}$$

In this equation, w_1 and w_2 are two specified weights.

To summarize the formulation which is given in this section, an op-amp design problem is a constrained optimization problem of the following form:

$$\left.\begin{aligned} \text{minimize} \quad & f_0 \\ \text{subject to} \quad & g_i \leq 1 \end{aligned}\right\}. \tag{22}$$

The objective function is given in (21) while various constraint functions are provided in (11) and (13)–(20). It is observed that, assuming constant V_T, $V_{T\mathrm{SAT}}$ and $(1 + \lambda V_{\mathrm{DS}})$, the constraint functions g_i's are *posynomials* of V_{GT}, g_m, g_d, and I_D of various transistors. Further, the objective function f_0 is a posynomial of W/L, L, and I_D of the transistors. With the same assumption, it is also found that I_D of the transistors are PoP functions of the design variables (W/L, L, and the bias current I_b). Therefore, with the extrapolated S–H model where the device parameters g_m, g_d, and V_{GT} are also PoP's in the design variables, the objective and the constraint functions are posynomials of the design variables. With logarithmic transformation, the optimization problem becomes a convex programming problem.

IV. Formulation of Op-Amp Design as a Sequence of Convex Programming Problems

In the last section, assuming constant V_T, $V_{T\mathrm{SAT}}$, and $(1 + \lambda V_{\mathrm{DS}})$, the op-amp synthesis problem is formulated as a convex programming problem. Here, to account for the effect of variations in V_T, $V_{T\mathrm{SAT}}$, and $(1 + \lambda V_{\mathrm{DS}})$, the overall method is shown in Fig. 4. The various steps of the method are as follows.

Step 0) Accept the designer specifications and the circuit to be sized. The user may provide an initial design point, which is optional. By default, the initial design point is taken as the minimum feature size for all the transistors. Using two iterations of the dc analysis technique described in Section III-A, an approximate dc operating point at the initial design point is found.

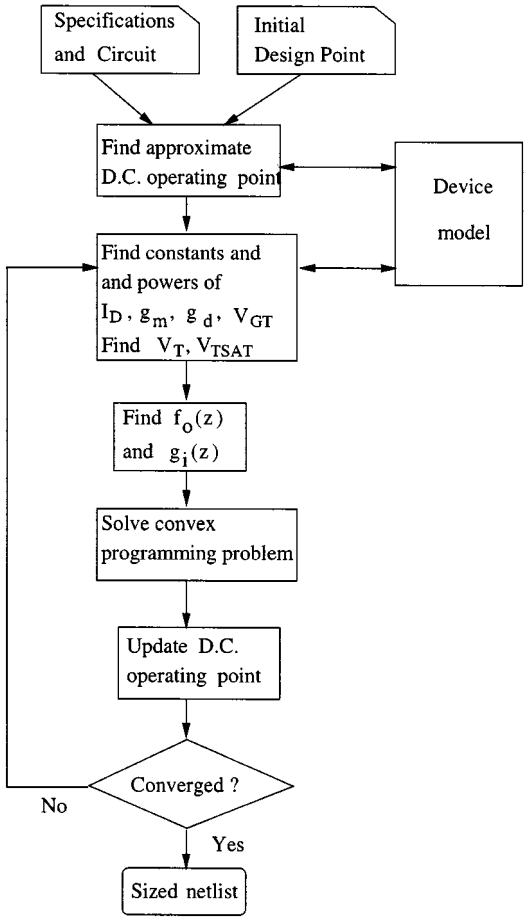

Fig. 4. Op-amp sizing method through sequential convex programming.

Step 1) In this step, the constants and powers of the PoP representation of the various I_{Di}'s are determined. The values of $V_{\mathrm{DS}i}$'s, which are used to determine the constant, come from the previous iteration. For the example circuit, refer to (5) and (6) for the relevant PoP functions. Similarly, the constants and powers of the PoP representation of the various g_{mi}'s, g_{di}'s, and V_{GTi}'s are also determined. Using the values of $V_{\mathrm{SB}i}$'s, which come from the previous iteration, the threshold voltage of the transistors are also evaluated here.

Step 2) From the constants and powers of g_{mi}'s, g_{di}'s, V_{GTi}'s, I_{Di}'s, and the values of V_{Ti}'s, the objective function and the constraint functions in the log transformed design space are derived. Both the objective function and the constraint functions are convex functions.

Step 3) The convex programming problem is then solved to find the global optimal solution of the current iteration.

Step 4) In this step, the dc operating point is updated for the new solution design point. The method of finding the dc operating point is the same as the iterative method described in Section III-A. The key difference, however, is that only one pass of the flow graph shown in Fig. 3 is executed. In the single pass, to get the updated values of I_{Di}'s, V_{GTi}'s, and V_{Ti}'s, the values of $V_{\mathrm{DS}i}$'s and $V_{\mathrm{SB}i}$'s are taken from the operating point at the solution point in the previous iteration.

Step 5) In this step, the convergence of the sequence of solution design points and node voltages are checked. If the design points and the node voltages of the last two iterations are very close, then the op-amp netlist with sized transistors is provided. Otherwise, we go back to *Step 1*.

Note that within an iteration, the values of $V_{\mathrm{DS}i}$'s and V_{Ti}'s, which are used in *Step 1* to determine the convex programming problem, are not mathematically consistent with the solution point. However, as the iterations proceed and when the sequence of solution points converges and the node voltages converge, the inconsistency becomes negligible. Such an approach is in spirit similar to the relaxed dc formulation, which is described in [15].

V. Two-Stage CMOS Op-Amp Sizing

The op-amp sizing technique, which is described in the last two sections, is applied here for designing a number of two-stage op-amps. For ease of referencing, in Section V-A, the op-amp topologies are characterized by five binary variables. While details of the design formulation of the two-stage op-amps are available in [20], a summary of dc analysis of the op-amps, design space constraints, and the performance constraints are provided, respectively, in Sections V-B–V-D. The design formulation is similar to that of the running example op-amp, which is given in Section III.

A. Characterization of Two-Stage Op-Amps

In the class of two-stage op-amps, there is a basic structural similarity, namely the hierarchical structure of different configurations is the same. It is only the subcircuits, which are the leaf cells of the hierarchy, that are different across the various topologies [29].

A two-stage op-amp consists of an input stage, a second stage, and a compensating circuit. The input stage has three parts: current source, differential pair, and current mirror. The second stage has two parts: transconductance amplifier and active load. Each one of the four subcircuits, namely differential pair, current mirror, transconductance amplifier, and active load, can be either simple or cascoded. For a cascoded current mirror, a level shifter is required between the input stage and the second stage. The compensating circuit consists of a capacitor and a resistor. Further, the transistors in the differential pair can either be n-type or p-type. The choice of polarity of the transistors in the differential pair also determines the polarity of the transistors in the other subcircuits. All the possible op-amp topologies are characterized by five binary variables, which are defined in Table II. With n-type differential pair transistors, the supercircuit [29] of the considered set of op-amps is given in Fig. 5. In the same figure, the relation among the binary variables and the subcircuits is also indicated by naming the enclosed subcircuits by the corresponding binary variables. So, from the supercircuit, depending on the values of the binary variables, other topologies can be constructed by selectively including the enclosed subcircuits.

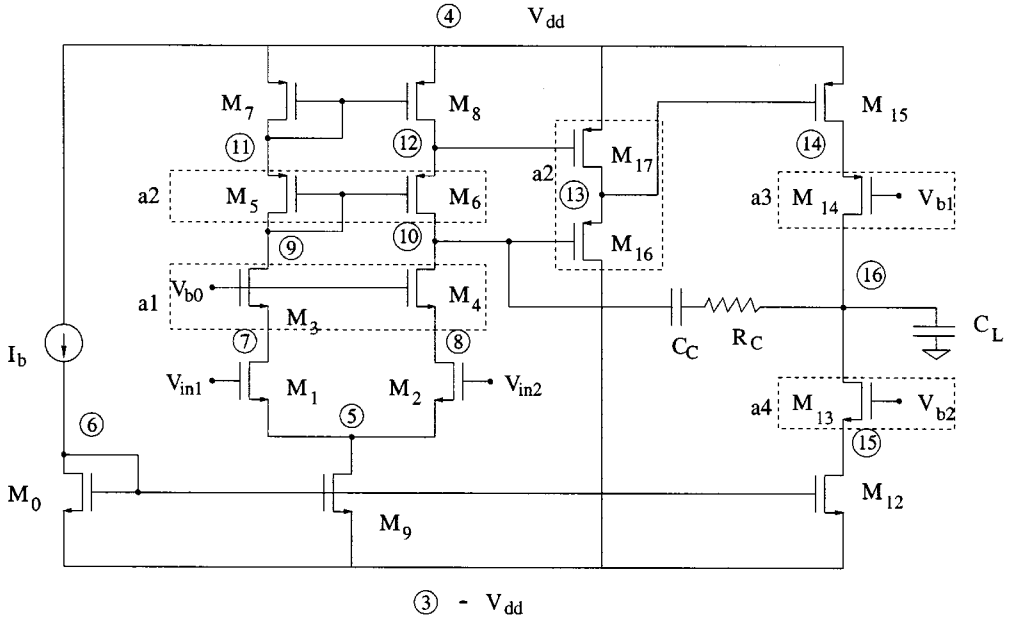

Fig. 5. Supercircuit of the two-stage op-amps.

TABLE II
DEFINITION OF THE TOPOLOGY CHARACTERIZING BINARY VARIABLES

Variables	'0'	'1'
a_1	simple input differential pair	cascode input differential pair
a_2	simple input stage load	cascode input stage load
a_3	simple second stage trans. amp.	cascode second stage trans. amp.
a_4	simple second stage load	cascode second stage load
a_5	input transistors are n-type	input transistors are p-type

B. DC Analysis

DC analysis includes finding drain currents through various transistors, determining $V_{\mathrm{GS}i}$'s of the transistors, expressing various node voltages in terms of $V_{\mathrm{GS}i}$'s, and, finally, writing $V_{\mathrm{DS}i}$'s and $V_{\mathrm{SB}i}$'s in terms of the node voltages. The current through the input stage and the second stage are, respectively,

$$\left.\begin{aligned} I_{D9} &= k_9'\left(\frac{W_9}{L_9}\right)(1+\lambda_9 V_{\mathrm{DS9}}) \\ &\quad\cdot\left[\frac{I_b}{k_0'(1+\lambda_0 V_{\mathrm{DS0}})}\left(\frac{L_0}{W_0}\right)\right] \\ \text{and } I_{D12} &= k_{12}'\left(\frac{W_{12}}{L_{12}}\right)(1+\lambda_{12} V_{DS12}) \\ &\quad\cdot\left[\frac{I_b}{k_0'(1+\lambda_0 V_{DS0})}\left(\frac{L_0}{W_0}\right)\right] \end{aligned}\right\}. \tag{23}$$

The current through the other transistors are

$$\left.\begin{aligned} I_{Di} &= \frac{I_{D9}}{2}, \qquad \text{for } i=1,\,2,\,\cdots,\,8 \\ I_{Di} &= I_{D12}, \qquad \text{for } i=13,\,14,\,15 \end{aligned}\right\}. \tag{24}$$

The sizes of transistors in the level shifter are so chosen that the node voltages V_{13} and V_{12} are equal. With $V_{13}=V_{12}$, the balance equation (39), which helps to achieve small systematic offset, remains simple even with fully cascode first stage. There are a number of possible ways by which this can be achieved. However, a simple way to get $V_{13}=V_{12}$ is that the sizes of transistors M_{17} and M_7 are equal, the length of the two transistors M_{16} and M_5 are the same, and their widths satisfy the following constraint:

$$\frac{W_5}{L_5}(1+\lambda_5 V_{\mathrm{DS5}}) = \frac{W_{16}}{L_{16}}(1+\lambda_{16} V_{\mathrm{DS16}}). \tag{25}$$

Now with $V_{13}=V_{12}$, the current through the transistors M_{16} and M_{17} are

$$I_{D16} = I_{D17} = I_{D9}/2. \tag{26}$$

The gate-to-source voltages of all the transistors are determined by

$$V_{\mathrm{GS}i} = V_{Ti} + V_{GTi}$$

where

$$V_{GTi} = \left[\frac{I_{Di}}{k_i'(1+\lambda_i V_{\mathrm{DS}i})\left(\frac{W_i}{L_i}\right)}\right]^{1/2}. \tag{27}$$

Through a casual dc analysis, which is discussed in Section III, various node voltages are expressed in terms of $V_{\mathrm{GS}i}$'s and bias voltages

$$\left.\begin{aligned} V_5 &= V_{\mathrm{in1}} - V_{T1} - V_{GT1} \\ V_6 &= -V_{\mathrm{dd}} + V_{T0} + V_{GT0} \\ V_7 &= V_8 = V_{b0} - V_{T3} - V_{GT3} \\ V_{15} &= V_{b2} - V_{T13} - V_{GT13} \\ V_{11} = V_{12} &= V_{13} = V_{\mathrm{dd}} - V_{T7} - V_{GT7} \\ V_9 = V_{10} &= V_{11} - V_{T5} - V_{GT5} \\ &= V_{\mathrm{dd}} - V_{T7} - V_{GT7} - V_{T5} - V_{GT5} \\ V_{14} &= V_{b1} + V_{T14} + V_{GT14} \end{aligned}\right\}. \tag{28}$$

Node 16 is the output node of the op-amp. As described in Section III, this node does not have a well defined equation. However, in actual application, op-amps are used in close loop con-

figuration, and in this configuration the output node voltage is, usually, stabilized to zero. It may be noted that $V_{\text{in1}} = V_{\text{in2}} = 0.0$.

The back bias of the transistors in the op-amp in terms of various node voltages are as follows:

$$\left.\begin{array}{rcl} V_{\text{SB}i} &=& 0.0, \qquad \text{for } i = 0,\ 7,\ 8,\ 9,\ 12,\ 15,\ 17 \\ V_{\text{SB1}} &=& V_{\text{SB2}} = V_5 - V_3 \\ V_{\text{SB3}} &=& V_{\text{SB4}} = V_7 - V_3 \\ V_{\text{SB5}} &=& V_{\text{SB6}} = V_4 - V_{11} \\ V_{\text{SB13}} &=& V_{15} - V_3 \\ V_{\text{SB14}} &=& V_4 - V_{14} \\ V_{\text{SB16}} &=& V_4 - V_{13} \end{array}\right\}. \quad (29)$$

Finally, drain-to-source voltages of the transistors are expressed in terms of the node voltages as follows

$$\left.\begin{array}{rcl} V_{\text{DS0}} &=& V_6 - V_3 \\ V_{\text{DS1}} &=& V_{\text{DS2}} = (1-a_1)(1-a_2)V_{11} \\ && + a_2(1-a_1)V_9 + a_1 a_2 V_7. - V_5 \\ V_{\text{DS3}} &=& V_{\text{DS4}} = (1-a_2)V_{11} + a_2 V_9 - V_7 \\ V_{\text{DS5}} &=& V_{\text{DS6}} = V_{11} - V_9 \\ V_{\text{DS7}} &=& V_{\text{DS8}} = V_4 - V_{11} \\ V_{\text{DS9}} &=& V_5 - V_3 \\ V_{\text{DS12}} &=& (1-a_4)V_{16} + a_4 V_{15} - V_3 \\ V_{\text{DS13}} &=& V_{16} - V_{15} \\ V_{\text{DS14}} &=& V_{14} - V_{16} \\ V_{\text{DS15}} &=& V_4 - (1-a_3)V_{16} + a_3 V_{14} \\ V_{\text{DS16}} &=& V_{13} - V_3 \\ V_{\text{DS17}} &=& V_4 - V_{13} \end{array}\right\}. \quad (30)$$

Note that the expressions of some $V_{\text{DS}i}$'s include the binary variables a_i's. This is because these $V_{\text{DS}i}$'s are topology dependent. Therefore, the binary variables are used to define the $V_{\text{DS}i}$'s across the various op-amp topologies.

C. Design Space Constraints

The design space constraints include the technology limits of the transistor sizes and the constraints on design variables, which help to bias the transistors in saturation. The constraints corresponding to the technology limits are straight forward, and hence they are not explicitly given here. Recall from Section III that the inequalities that keep the transistors away from the linear region are given by

$$\left.\begin{array}{r} V_D \geq V_G - V_{T\text{SAT}} \quad \text{for n-type} \\ \text{or } V_D \leq V_G + V_{T\text{SAT}} \quad \text{for p-type} \end{array}\right\} \quad (31)$$
$$\text{where } V_{T\text{SAT}} = V_{\text{GS}} - V_{\text{DS}AT}.$$

Using the appropriate values of the node voltages from (28) in the above inequality, we get the required constraints.

Consider the input stage of the supercircuit. To keep the transistor M_1 (and M_2) in saturation, the required inequality for simple differential pair is

$$V_{\text{dd}} - V_{GT7} - V_{T7} - a_2 V_{GT5} - a_2 V_{T5} \geq -V_{T\text{SAT3}}$$

i.e.

$$\frac{1}{(V_{\text{dd}} - V_{T7} - a_2 V_{T5} + V_{T\text{SAT3}})}(V_{GT7} + a_2 V_{GT5}) \leq 1 \quad (32)$$

and for cascode differential pair, the required inequality is

$$(V_{b0} - V_{GT3} - V_{T3}) \geq -V_{T\text{SAT1}}$$

i.e.

$$(V_{T3} - V_{T\text{SAT1}})/V_{b0} + V_{GT3}/V_{b0} \leq 1. \quad (33)$$

In cascode differential pair, the transistors M_3 (and M_4) are kept in saturation by the inequality

$$V_{\text{dd}} - V_{GT7} - V_{T7} - a_2 V_{GT5} - a_2 V_{T5} \geq V_{b0} - V_{T\text{SAT3}}$$

i.e.

$$\frac{1}{(V_{\text{dd}} - V_{T7} - a_2 V_{T5} + V_{T\text{SAT3}})} \times (V_{GT7} + a_2 V_{GT5} + V_{b0}) \leq 1. \quad (34)$$

The transistor M_9 is kept in saturation by the inequality

$$-V_{GT1} - V_{T1} \geq -V_{\text{dd}} + V_{GT0} + V_{T0} - V_{T\text{SAT9}}$$

i.e.

$$\frac{1}{(V_{\text{dd}} - V_{T1} - V_{T0} + V_{T\text{SAT9}})}(V_{GT0} + V_{GT1}) \leq 1. \quad (35)$$

In the second stage, the transistor M_{12} is kept in saturation by the inequality

$$a_4(-|V_{b2}| - V_{GT13} - V_{T13}) \geq -V_{\text{dd}} + V_{GT0} + V_{T0} - V_{T\text{SAT12}}$$

i.e.

$$\frac{1}{(V_{\text{dd}} - a_4 V_{T13} - V_{T0} + V_{T\text{SAT12}})} \times (a_4|V_{b2}| + V_{GT0} + a_4 V_{GT13}) \leq 1 \quad (36)$$

and the transistor M_{15} is kept in saturation by the inequality

$$\frac{1}{(V_{\text{dd}} - a_3 V_{T14} - V_{T7} + V_{T\text{SAT15}})} \times (a_3 V_{b1} + V_{GT7} + a_3 V_{GT14}) \leq 1. \quad (37)$$

Equation (37) is derived based on the assumption that the $V_{13} = V_{11}$. The design of the level shifter to get this biasing condition is already discussed in Section V-B. Note, finally, that the remaining transistors in the supercircuit are automatically biased in the saturation region.

All the transistors in the op-amps are kept away from the subthreshold region by the inequalities,

$$\frac{\varepsilon_{\text{SUB}}}{V_{GTi}} \leq 1. \quad (38)$$

Along with the various inequality design space constraints, there are two equality constraints also. One of them is given in (25), which is related to designing the level shifter to get $V_{12} = V_{13}$. The other equality constraint is the balance equation, which is given by

$$\frac{(1+\lambda_{15}V_{\text{DS15}})(1+\lambda_9 V_{\text{DS9}})}{(1+\lambda_{12}V_{\text{DS12}})(1+\lambda_7 V_{\text{DS7}})} \times \left(\frac{W_{15}}{L_{15}}\right)\left(\frac{W_9}{L_9}\right)\left(\frac{L_{12}}{W_{12}}\right)\left(\frac{L_7}{W_7}\right) = 1. \quad (39)$$

The balance equation is necessary to keep the systematic offset small [27].

The concluding remarks about the design space constraints are that they are posynomial in V_{GTi}'s if the V_{Ti}'s and V_{TSAT}'s are assumed to be constant. With a further simplifying assumption, i.e., constant $(1+\lambda_i V_{DSi})$, the V_{GTi}'s are PoP functions of the design variables for the S–H model. Therefore, the design space constraints are posynomials of the design variables. Finally, the equality constraints are PoP functions of design variables. Therefore, each one of the equality constraints can be translated into two PoP terms less than or equal to one.

D. Performance Constraints and Objective Function

This section provides analytical expressions of various performance functions of the op-amps. Performance constraints are then derived by constraining the performance expressions by their respective specified value. The analytical expression of the objective function is also given here.

1) DC and Transient Performances: For simple input differential pair ($a_1 = 0$), the positive common mode range is $\text{CMR}^+ = V_{\text{dd}} - V_{\text{GS7}} - a_2 V_{\text{GS5}} + V_{\text{TSAT1}}$. Therefore, to meet its specification, the constraint is

$$\frac{1}{(V_{\text{dd}} - V_{T7} - a_2 V_{T5} + V_{\text{TSAT1}} - \text{CMR}^+_{\text{SPEC}}) \times (V_{GT7} + a_2 V_{GT5}) \le 1}. \quad (40)$$

On the other hand, for cascoded input differential pair (i.e., $a_1 = 1$), $\text{CMR}^+ = V_{b0} - V_{GT3} - V_{T3} + V_{\text{TSAT1}}$ and the corresponding performance constraint is

$$(\text{CMR}^+_{\text{SPEC}} - V_{\text{TSAT1}} + V_{T3})/V_{b0} + V_{GT3}/V_{b0} \le 1. \quad (41)$$

The negative common mode range is given by $\text{CMR}^- = -V_{\text{dd}} + V_{\text{GS0}} + V_{\text{GS1}} - V_{\text{TSAT9}}$, and to meet its specification

$$\frac{1}{(V_{\text{dd}} - V_{T0} - V_{T1} + V_{\text{TSAT9}} - |\text{CMR}^-_{\text{SPEC}}|) \times (V_{GT0} + V_{GT1}) \le 1}. \quad (42)$$

With simple output transconductance amplifier ($a_1 = 0$), the positive output swing is $OS^+ = V_{\text{dd}} - V_{\text{GS7}} + V_{\text{TSAT15}}$ and the required constraint to meet specification is

$$\left[\frac{1}{V_{\text{dd}} - V_{T7} + V_{\text{TSAT15}} - OS^+_{\text{SPEC}}}\right] V_{GT7} \le 1. \quad (43)$$

On the other hand, when the transconductance amplifier is cascoded (i.e., $a_3 = 1$), the positive output swing is $OS^+ = V_{b1} + V_{\text{TSAT14}}$ and to meet its specification we require

$$(OS^+_{\text{SPEC}} - V_{\text{TSAT14}})/V_{b1} \le 1. \quad (44)$$

With a simple second-stage active load, to meet the negative output swing specification, we require

$$\left[\frac{1}{V_{\text{dd}} - V_{T0} + V_{\text{TSAT12}} - |OS^-_{\text{SPEC}}|}\right] V_{GT0} \le 1. \quad (45)$$

On the other hand, when the active load is cascoded ($a_4 = 1$), the negative output swing constraint is

$$(|OS^-_{\text{SPEC}}| - V_{\text{TSAT13}})/|V_{b2}| \le 1. \quad (46)$$

Note that the magnitude of the bias voltage V_{b2}, rather than its actual value, is taken as the design variable.

To meet the slew rate specification, the following two inequalities should be satisfied

$$\left.\begin{aligned} \frac{SR_{\text{SPEC}} C_C}{I_{D9}} &\le 1 \\ \frac{SR_{\text{SPEC}} C_L}{I_{D12}} &\le 1 \end{aligned}\right\}. \quad (47)$$

2) AC Performances: The low frequency gain of an op-amp is

$$A_0 = \frac{g_{m1} g_{m15}}{g_{o1} g_{o2}}$$

where output conductances of the two stages are

$$\left.\begin{aligned} g_{o1} &= g_{d1}\left[\frac{g_{d3}}{g_{m3}}\right]^{a_1} + g_{d7}\left[\frac{g_{d5}}{g_{m5}}\right]^{a_2} \\ g_{o2} &= g_{d12}\left[\frac{g_{d13}}{g_{m13}}\right]^{a_4} + g_{d15}\left[\frac{g_{d14}}{g_{m14}}\right]^{a_3} \end{aligned}\right\}. \quad (48)$$

To satisfy the gain specification, we require

$$\frac{A_{\text{SPEC}} g_{o1} g_{o2}}{g_{m1} g_{m15}} \le 1$$

i.e.

$$t_1 + t_2 + t_3 + t_4 \le 1 \quad (49)$$

where

$$\begin{aligned} t_1 &= \frac{A_{\text{SPEC}} g_{d1} g_{d12}}{g_{m1} g_{m15}} \left[\frac{g_{d3}}{g_{m3}}\right]^{a_1} \left[\frac{g_{d13}}{g_{m13}}\right]^{a_4} \\ t_2 &= \frac{A_{\text{SPEC}} g_{d1} g_{d15}}{g_{m1} g_{m15}} \left[\frac{g_{d3}}{g_{m3}}\right]^{a_1} \left[\frac{g_{d14}}{g_{m14}}\right]^{a_3} \\ t_3 &= \frac{A_{\text{SPEC}} g_{d7} g_{d12}}{g_{m1} g_{m15}} \left[\frac{g_{d5}}{g_{m5}}\right]^{a_2} \left[\frac{g_{d13}}{g_{m13}}\right]^{a_4} \\ t_4 &= \frac{A_{\text{SPEC}} g_{d7} g_{d15}}{g_{m1} g_{m15}} \left[\frac{g_{d5}}{g_{m5}}\right]^{a_2} \left[\frac{g_{d14}}{g_{m14}}\right]^{a_3}. \end{aligned}$$

Transfer function of a two-stage compensated op-amp can be well approximated with its low frequency gain, one zero and three poles. With appropriate choice of the compensating resistor R_c ($=1/g_{m15}$), the zero of the transfer function of the op-amp can be placed at very high frequency [27]. The three poles of the transfer function are well approximated by

$$\left.\begin{aligned} P_1 &= \frac{g_{o1} g_{o2}}{g_{m15} C_C} \\ P_2 &= \frac{g_{m15}}{C_L} \\ \text{and } P_3 &= \frac{g_{m15}}{C_{gs15}} \end{aligned}\right\}. \quad (50)$$

In order to have a transfer function that is well approximated by a single pole system in the frequency range of interest, the

second pole should be beyond the gain bandwidth product ($=g_{m1}/C_C$), i.e., the following should be satisfied

$$\frac{g_{m1\bar{5}}}{C_L} > \frac{g_{m1}}{C_C}. \tag{51}$$

For a true single pole behavior, unity gain frequency (UGF) is equal to g_{m1}/C_C. However, with the approximate single pole behavior that we get by (51), g_{m1}/C_C well approximates the UGF. Hence, the specified UGF can be achieved by

$$\frac{\text{UGF}_{\text{SPEC}} C_C}{g_{m1}} \leq 1. \tag{52}$$

The phase margin is

$$\text{PM} = 90 - \arctan(\text{UGF}/P_2) - \arctan(\text{UGF}/P_3).$$

To meet phase margin specification

$$\arctan(\text{UGF}/P_2) + \arctan(\text{UGF}/P_3) \leq 90^\circ - \text{PM}_{\text{SPEC}}$$

or

$$\arctan\left[\frac{\dfrac{\text{UGF}}{P_2} + \dfrac{\text{UGF}}{P_3}}{1 - \dfrac{\text{UGF}^2}{P_2 P_3}}\right] \leq 90^\circ - \text{PM}_{\text{SPEC}}$$

or

$$\frac{1}{\tan(90^\circ - \text{PM}_{\text{SPEC}})}\left[\frac{\text{UGF}}{P_2} + \frac{\text{UGF}}{P_3}\right] + \frac{\text{UGF}^2}{P_2 P_3} \leq 1$$

or

$$\frac{1}{\tan(90^\circ - \text{PM}_{\text{SPEC}})}\left[\frac{g_{m1} C_L}{g_{m1\bar{5}} C_C} + \frac{g_{m1} C_{gs1\bar{5}}}{g_{m1\bar{5}} C_C}\right] + \frac{g_{m1}^2 C_L C_{gs1\bar{5}}}{g_{m1\bar{5}}^2 C_C^2} \leq 1. \tag{53}$$

The low-frequency CMRR of an op-amp is

$$\text{CMRR}_0 = \frac{A1_0}{\text{CMG1}_0}$$

where

$$\text{CMG1}_0 = \frac{g_{d9}}{2}\left(\frac{a_2}{g_{m\bar{5}}} + \frac{1}{g_{m7}}\right)$$

and

$$A1_0 = \frac{g_{m1}}{g_{o1}}$$

i.e.

$$\frac{1}{\text{CMRR}_0} = \frac{g_{d9}}{2g_{m1}}\left(\frac{a_2}{g_{m\bar{5}}} + \frac{1}{g_{m7}}\right) \times \left(g_{d1}\left[\frac{g_{d3}}{g_{m3}}\right]^{a_1} + g_{d\bar{7}}\left[\frac{g_{d\bar{5}}}{g_{m\bar{5}}}\right]^{a_2}\right). \tag{54}$$

To meet its specification, we require

$$\frac{\text{CMRR}_{\text{SPEC}}}{\text{CMRR}_0} \leq 1$$

i.e.

$$t_{\bar{5}} + t_6 + t_7 + t_8 \leq 1 \tag{55}$$

where

$$t_{\bar{5}} = \frac{\text{CMRR}_{\text{SPEC}} a_2 g_{d1} g_{d9}}{2 g_{m1} g_{m\bar{5}}}\left[\frac{g_{d3}}{g_{m3}}\right]^{a_1}$$

$$t_6 = \frac{\text{CMRR}_{\text{SPEC}} g_{d1} g_{d9}}{2 g_{m1} g_{m7}}\left[\frac{g_{d3}}{g_{m3}}\right]^{a_1}$$

$$t_7 = \frac{\text{CMRR}_{\text{SPEC}} a_2 g_{d7} g_{d9}}{2 g_{m1} g_{m\bar{5}}}\left[\frac{g_{d\bar{5}}}{g_{m\bar{5}}}\right]^{a_2}$$

$$t_8 = \frac{\text{CMRR}_{\text{SPEC}} g_{d7} g_{d9}}{2 g_{m1} g_{m7}}\left[\frac{g_{d\bar{5}}}{g_{m\bar{5}}}\right]^{a_2}.$$

Note that the above equation for CMRR—which, as our results in Section VII show, matches SPICE simulation well—does not include the effect of mismatch. Therefore, for a more realistic estimate of CMRR, the effect of statistical mismatch variations should be routinely incorporated in SPICE simulation and in analytical techniques such as ours.

3) Objective Function: In the optimization formulation, a weighted sum of the total gate area of the transistors and the power dissipation is taken as the objective function to be minimized.

The total power dissipation is

$$\begin{aligned}\text{PD} &= 2V_{\text{dd}}(I_b + I_{D9} + I_{D12} + a_2 I_{D17}) \\ &= 2V_{\text{dd}}(I_b + I_{D9}(1 + a_2/2) + I_{D12})\end{aligned}$$

and the total effective gate area is

$$\begin{aligned}\text{EFF\_AREA} &= \sum_{i=0,1,2,\bar{7},8,9,12,\bar{15}}\left(\frac{W_i}{L_i}\right)L_i^2 + a_1 2\left(\frac{W_3}{L_3}\right)L_3^2 \\ &\quad + a_2\left[2\left(\frac{W_{\bar{5}}}{L_{\bar{5}}}\right)L_{\bar{5}}^2 + \left(\frac{W_{16}}{L_{16}}\right)L_{16}^2 + \left(\frac{W_{17}}{L_{17}}\right)L_{17}^2\right] \\ &\quad + a_3\left(\frac{W_{14}}{L_{14}}\right)L_{14}^2 + a_4\left(\frac{W_{13}}{L_{13}}\right)L_{13}^2.\end{aligned}$$

Note that all the performance constraint functions are posynomials in the device parameters V_{GTi}'s, I_{Di}'s, g_{mi}'s, and g_{di}'s. These device parameters are PoP function of design variables. Therefore, the performance constraint functions are posynomials in the design variables. Further, the power dissipation and the total effective gate area are posynomials of the design variables. As with the design space constraints, the performance constraints are expressed in terms of device parameters, which are in turn functions of the actual design variables.

VI. Simulation Results

The CMOS op-amp sizing technique that has been described so far has been implemented in MATLAB [30] for the class of two-stage CMOS op-amps. In the implementation, each convex programming problem is solved by using the sequential quadratic programming method, which is available in the optimization toolbox [30]. This was done in the interest of quick prototyping in order to provide a proof of concept. As our experimental results show, this implementation is very fast in spite of the fact that we did not use any of the far more efficient special purpose techniques that exist for convex optimization [25].

TABLE III
SPECIFICATIONS, S–H MODEL BASED PREDICTED PERFORMANCES AND SPICE SIMULATED PERFORMANCES AT THE FINAL DESIGN POINT OF THE OP-AMP IN THE FIRST EXAMPLE

Perf.(unit)	Spec.	S-H model pred.	Spice simul.
A(0) (dB)	70	70	59
UGF (MHz)	10	10	9.6
PM (deg.)	60	65	71
BW (kHz)	–	3.3	10.6
CMRR (dB)	60	79	71
CMR (V)	-2.5,2.5	-2.8,5.0	-3.8,4.7
OS (V)	-3,3	-4.7,4.7	-4.8,4.8
SR (V/μsec)	20	20	19.3
PD (mW)	–	0.27	0.39
Total gate area ($\mu \times \mu$)	–	87.6	–

TABLE IV
OPTIMIZATION STATISTICS FOR THE FIRST EXAMPLE BY STARTING FROM SMALL SIZES OF ALL THE TRANSISTORS

It.no.	CPU time(sec)	MaxErr in L_i (μ)	MaxErr in W_i (μ)	MaxErr in V_n (V)
1	9.5100	–	–	–
2	7.8800	0.2339	1.6499	0.2005
3	7.1000	0.0166	0.2664	0.0305
4	5.2300	0.0008	0.0048	0.0048
5	5.1600	0.0001	0.0006	0.0008
	(total 34.8900)			

A number of two-stage op-amps were sized for a 1.6-μm technology. The experimental results are given in this section. These results include CPU time required to find the final design point, the convergence behavior of the method, and a comparison of predicted performances with SPICE simulations based on the level two MOS model.

Example 1: The topology for the first example was the simplest op-amp (all subcircuits were simple). The set of performance specification of this example is given in the second column of Table III.

The circuit was sized by starting from four different initial design points. The four initial design points are: 1) small size (minimum feature size $L = 1.6$ μm, $W = 2.4$ μm) of all the transistors; 2) small size of the input-stage transistors and large size ($L = 5.6$ μm, $W = 50$ μm) of second-stage transistors; 3) large size of the input stage transistors and small size of the second-stage transistors; and 4) large size of all the transistors. The optimization statistics in case 1), i.e., starting with small size of all of the transistors, is shown in Table IV. In the table, the first column indicates the iteration number of the sequential convex optimization algorithm. The second column gives the CPU time (IBM RS/6000, running AIX) required for solving the convex programming problem using sequential quadratic programming. The last three columns of the table provide the maximum difference in transistor lengths and widths, and node voltages at the two-solution design points in two consecutive iterations. The iteration process was stopped when the maximum difference in the lengths and the widths was less than 0.02 μm and the maximum difference in the node voltages was less than 1 mV. With the other three initial guesses, the convergence behavior was essentially the same. The sequence of convex programs converged in less than 50 s of CPU time with at most five iterations.

TABLE V
OPTIMAL DESIGN POINT OF THE FIRST EXAMPLE

Variables	values	unit
W_0/L_0	2.4000/2.0441	μ/μ
W_1/L_1	2.4000/3.5361	μ/μ
W_2/L_2	2.4000/3.5361	μ/μ
W_7/L_7	3.8246/1.9246	μ/μ
W_8/L_8	3.8246/1.9246	μ/μ
W_9/L_9	2.4000/2.0889	μ/μ
W_{12}/L_{12}	7.1581/1.6000	μ/μ
W_{15}/L_{15}	21.5914/1.6000	μ/μ
Cc	0.1830	pF
Ib	3.3468	μA

TABLE VI
COMPARISON OF THE FOUR FINAL DESIGN POINTS OBTAINED BY STARTING FROM FOUR INITIAL GUESSES IN THE FIRST EXAMPLE

Comparison points	MaxDiff in L_i (μ)	MaxDiff in W_i (μ)	MaxDiff in V_n (V)
FDa vs. FDb	0.0001	0.0005	0.0000
FDa vs. FDc	0.0001	0.0005	0.0000
FDa vs. FDd	0.0001	0.0009	0.0000

The four solution design points, which were obtained by starting from the four initial design points 1)–4), are denoted by $FD1$, $FD2$, $FD3$, and $FD4$, respectively. The solution point $FD1$ is given in Table V, while Table VI reports the differences among the four solution design points. Note that all four solution design points are essentially the same. It should also be noted in this context that in this and the other examples in this paper, we have reported the details of the iterative process with precision in the millivolt range for node voltages and in the nanometer range for device sizes. This has been done in order to demonstrate that our technique converges to a consistent dc operating point and that the final solution is the same even with very different initial guesses. However, this does not mean that the transistors have to be sized this precisely to get the reported performance.

An expert designer would pick a nonminimum channel length for the input stage transistors to get high gain. On the other hand, for the second-stage transistors, in order to achieve high slew rate, he/she would choose minimum channel length with large channel width. The optimal design point given in Table V is qualitatively similar to such a choice.

The various performances at the optimal design point as predicted by the program using the S–H model and the corresponding SPICE simulations (using the level two MOS model) are given in the last two columns of Table III.

Note that while the S–H model based predictions satisfy the specifications, many of the performances as actually measured in SPICE (using level two model) do not. The inaccuracies are

TABLE VII
OPTIMAL DESIGN POINT IN THE SECOND EXAMPLE

Variables	values	unit
W_0/L_0	2.4000/3.4674	μ/μ
W_1/L_1	2.4000/2.6598	μ/μ
W_2/L_2	2.4000/2.6598	μ/μ
W_7/L_7	2.5782/2.6109	μ/μ
W_8/L_8	2.5782/2.6109	μ/μ
W_9/L_9	2.4000/4.0244	μ/μ
W_{12}/L_{12}	8.1017/2.6000	μ/μ
W_{15}/L_{15}	27.9688/2.6000	μ/μ
Cc	0.2943	pF
Ib	6.7618	μA

TABLE VIII
SPECIFICATIONS, S–H MODEL BASED PREDICTED PERFORMANCES, AND SPICE SIMULATED PERFORMANCES AT THE OPTIMAL DESIGN POINT OF THE SECOND EXAMPLE

Perf.(unit)	Spec.	S-H model pred.	Spice simul.
A(0) (dB)	70	70	67
UGF (MHz)	10	10	9
PM (deg.)	60	66	61
BW (kHz)	–	3.2	4.3
CMRR (dB)	60	81	82
CMR (V)	-2.5,2.5	-2.5,4.8	-3.4,4.8
OS (V)	-3,3	-4.4,4.5	-4.4,4.7
SR (V/μsec)	20	20	18.4
PD (mW)	–	0.51	0.54
Total gate area ($\mu \times \mu$)	–	137.99	–

due to the inadequacy of the S–H MOS model in the short channel regime. However, for long channels the model is quite accurate. To demonstrate this, in the following example we restrict the channel length of the transistors to be more than 2.6 μm and do the design optimization.

Example 2: This example is the same as the first example except that the lengths of the transistors were restricted to be more than or equal to 2.6 μm (though the technology limit is 1.6 μm). The solution design point is given in Table VII. The predicted performance of the op-amp at the final design point is given in Table VIII. Note that at the final design point, the predicted performances using S–H MOS model is close to those of SPICE simulation. However, the total gate area and the power dissipation at the solution design point obtained in this example are, respectively, 57% and 38% more than those at the solution design point obtained in the first example. It is therefore clear that the S–H model is inadequate for synthesis tools targeted at modern short-channel processes. So, we need a MOS model that is accurate in the short-channel regime. To increase the design accuracy in the prototype, the S–H model is replaced by a newly proposed MOS model. The corresponding simulation results are provided in the following section.

VII. IMPROVING DESIGN ACCURACY USING ACCURATE MOS MODEL

In the previous section it has been demonstrated that the proposed sizing method is robust and it has the capability of picking the optimal design point. However, in the implementation of the technique, the S–H MOS model was used, which is not an accurate model for small-channel length transistor. Therefore, at the design points, which were obtained by the design optimization, the op-amps could not meet some required performances. To improve the design accuracy, in the second version of the prototype, the S–H model was replaced by a newly proposed MOS model α–A (alpha–Analog). While the details of the model are available in [20], in the following section the key equations are provided. Section VII-B provides some modified equations compared to those in Section V, which are required for replacing the S–H model by the α–A model. Simulation examples are provided in Section VII-C.

A. The Model Description

The α–A MOS model is an extension of the simple α-power law MOS transistor model proposed by Sakurai and Newton [31]. It may be noted that, since in the α-power model the drain conductance over the saturation region is taken as zero, it can not be directly used for analog circuit analysis.

A summary description of the α–A model is given below. As the model is intended for use in analog circuit design, the model description is only for the saturation region of operation. A transistor is in the saturation region when

$$V_{\rm GS} > V_T \quad \text{and} \quad V_{\rm DS} \geq V_{D\rm SAT}. \tag{56}$$

The threshold voltage and the drain saturation voltage are, respectively,

$$\left.\begin{aligned} V_T &= V_{To} + V_{\rm TB} \cdot V_{\rm TL} \\ \text{and } V_{D\rm SAT} &= K_V {V_{GT}}^{\beta} \end{aligned}\right\}. \tag{57}$$

In the saturation region, drain current, transconductance, and drain-to-source conductance are, respectively,

$$\left.\begin{aligned} I_D &= K_C\left(\frac{W}{L}\right) V_{GT}^{\alpha}(1+\lambda V_{\rm DS}) \\ g_m &= \alpha {I_D}^{(\alpha-1)/\alpha}\left[K_C\left(\frac{W}{L}\right)(1+\lambda V_{\rm DS})\right]^{1/\alpha} \\ g_d &= I_D \cdot B_{f1} L^{\gamma_{f1}} \cdot f_2 \cdot f_3 \end{aligned}\right\}. \tag{58}$$

In (57), V_{To} is the zero back bias threshold voltage, while $V_{\rm TB}$ and $V_{\rm TL}$ represent dependencies of threshold voltage on the back bias $V_{\rm SB}$ and the channel length L, respectively. The dependency functions are

$$\left.\begin{aligned} V_{\rm TB} &= a_{V\rm TB1} V_{\rm SB} + a_{V\rm TB2} \cdot V_{\rm SB}^2 \\ \text{and } V_{\rm TL} &= 1 + B_{V\rm TL} L^{\gamma_{V\rm TL}} \end{aligned}\right\}. \tag{59}$$

In the same equation (57), the parameters K_V and β are given as

$$\left.\begin{aligned} K_V &= A_{KV} + B_{KV} L^{\gamma_{KV}} \\ \beta &= A_\beta + B_\beta L^{\gamma_\beta} \end{aligned}\right\}. \tag{60}$$

For a long-channel transistor drain current, I_D is proportional to V_{GT}^2, i.e., in (58) α is two. However, in [31] it is shown that the value of α is dependent on the channel length and in the small channel length regime, the value is quite close to one. We found that the parameter K_C in the equation for I_D also depends on the channel length. For the two parameters the following models are taken:

$$\left.\begin{aligned} K_C &= A_{KC} + B_{KC} L^{\gamma_{KC}} \\ \alpha &= A_\alpha + B_\alpha L^{\gamma_\alpha} \end{aligned}\right\}. \tag{61}$$

The transconductance g_m [in (58)] is obtained by taking the derivative of I_D with respect to V_{GT}.

Now we consider the drain conductance g_d. In the S–H model, $g_d \propto I_D/L$. Using this information, we have taken the model of g_d which is given in (58). In this model, B_{f1} and γ_{f1} are constant while the term f_2 represents the dependencies of g_d on V_{GT} and $V_{\rm SB}$. The model of f_2 is as follows:

$$\left.\begin{aligned} f_2 &= \frac{g_d}{I_D \cdot f_1} = A_{f2} + B_{f2} V_{GT}^{\gamma_{f2}} \\ \text{where } f_1 &= \frac{g_d}{I_D} = B_{f1} L^{\gamma_{f1}} \\ B_{f2} &\text{ and } \gamma_{f2} \text{ are second-order polynomials of } V_{\rm SB} \end{aligned}\right\}. \tag{62}$$

The term f_3 in g_d (58) represents the dependency of g_d on $V_{\rm DS}$. Over the saturation region of operation, the function can be well approximated by an exponential function. The model we have taken for f_3 is

$$\left.\begin{aligned} f_3 &= \frac{g_d}{I_D \cdot f_1 \cdot f_2} \\ &= 1 + B_{f3} \exp\left[-C_{f3} V_{GT}^{\gamma_{f31}} (V_{\rm DS} - V_{D\rm SAT})^{\gamma_{f32}}\right] \\ &\quad \text{where } C_{f3},\ \gamma_{f31} \text{ and } \gamma_{f32} \text{ are second-order} \\ &\quad \text{polynomials of } V_{\rm SB} \text{ and } L \end{aligned}\right\}. \tag{63}$$

The model parameter extraction procedure is given in [20].

B. Application of the New Model

The new MOS model, which is described in the last section, is used for op-amp design automation. In the new model, (58) provides the first-order models of I_D, g_m, and g_d as function of biases and transistor sizes. These first-order models are PoP functions. This property of the new model helps to use it in design optimization through sequential convex programming. In the sequential design optimization, (58), which provides first-order models, is used within the main convex programming (CP) optimization. For the iterative to update from one CP to the next CP in the optimization, the higher order effects are captured by using (57) and (59)–(63). With the new model, all the design equations provided in Section V remain the same. However, the expressions of the drain currents and effective gate-to-source voltages of the transistors are different. The current through the input and the second stages are, respectively,

$$\left.\begin{aligned} I_{D9} &= K'_{C9}\left(\frac{W_9}{L_9}\right)(1+\lambda_9 V_{\rm DS9}) \\ &\quad \left[\frac{I_b}{K'_{C0}(1+\lambda_0 V_{\rm DS0})}\left(\frac{L_0}{W_0}\right)\right]^{\alpha_9/\alpha_0} \\ \text{and } I_{D12} &= K'_{C12}\left(\frac{W_{12}}{L_{12}}\right)(1+\lambda_{12} V_{\rm DS12}) \\ &\quad \left[\frac{I_b}{K'_{C0}(1+\lambda_0 V_{\rm DS0})}\left(\frac{L_0}{W_0}\right)\right]^{\alpha_{12}/\alpha_0} \end{aligned}\right\} \tag{64}$$

and the effective gate to source voltages of the transistors are

$$V_{GTi} = \left[\frac{I_{Di}}{K'_{Ci}(1+\lambda_i V_{{\rm DS}i})}\left(\frac{L_i}{W_i}\right)\right]^{1/\alpha_i}. \tag{65}$$

TABLE IX
SPECIFICATIONS, NEW MODEL BASED, AND S–H MODEL BASED PREDICTED PERFORMANCES AND SPICE SIMULATED PERFORMANCES AT THE FINAL DESIGN POINT OF THE OP-AMP IN EXAMPLE 3

Perf.(unit)	Spec.	New model pred.	Spice simul.	S-H model pred.
A(0) (dB)	60	60	58	68
UGF (MHz)	10	10	9	9
PM (deg.)	60	65.4	62.3	56.7
BW (KHz)	–	10	12	3.7
CMRR (dB)	52	68	66	75
CMR (V)	-2.5,2.5	-3.0,4.9	-3.8,4.85	-2.8.5.0
OS (V)	-3,3	-4.8,4.7	-4.8,4.85	-4.7, 4.7
SR (V/μsec.)	20	20	17.3	12.4
PD (mW)	–	0.28	0.28	0.198
Total gate area ($\mu \times \mu$)	–	55.32	–	55.32

C. Simulation Results

Example 3: In this example, the topology is the same as that in Example 1 (Section VI-A), i.e., simple op-amp. The specification set of this example is shown in Table IX. Like the other example, here the op-amp was designed by starting from the following four initial design points:

1) small size of all the transistors;
2) small size of the input-stage transistors and large size of second-stage transistors;
3) large size of the input-stage transistors and small size of the second-stage transistors;
4) large size of all the transistors. Further, the op-amp was designed by starting from 25 random initial guesses that are uniformly distributed in the space of design variables.

The optimization statistics for case 1) is given in Table X. In the table, the maximum constraint violation reported for each iteration is reported by the SQP in MATLAB at the start of the solution process corresponding to that iteration. Note that all the specs have been normalized to have a value of one. Thus, a constraint violation of two is a 200% violation, while a constraint violation of 0.002 is a 0.2% violation. The final solution design point is given in the Table XI.

TABLE X
OPTIMIZATION STATISTICS FOR EXAMPLE 3 BY STARTING FROM SMALL SIZE OF ALL THE TRANSISTORS IN THE OP-AMP

It.no.	CPU time (sec)	MaxErr in L_i (μ)	MaxErr in W_i (μ)	MaxErr in V_n (V)	MaxConst violation
1	10.1400	–	–	–	2.0742
2	7.0300	0.5704	4.2527	0.4248	0.1804
3	7.0200	0.3932	1.0228	0.0692	0.1572
4	5.2300	0.0629	0.6738	0.0594	0.0376
5	5.0100	0.0775	0.8558	0.0128	0.0104
6	4.4400	0.0365	0.2491	0.0112	0.0103
7	4.3600	0.0123	0.1183	0.0021	0.0053
8	2.8400	0.0066	0.0466	0.0011	0.0032
9	3.7900	0.0041	0.0058	0.0004	0.0019
	(total 49.8900)				

TABLE XI
THE OPTIMAL DESIGN POINT IN EXAMPLE 3

Variables	values	unit
W_0/L_0	2.4000/2.9025	μ/μ
W_1/L_1	2.4000/3.1213	μ/μ
W_2/L_2	2.4000/3.1213	μ/μ
W_7/L_7	2.4000/1.6000	μ/μ
W_8/L_8	2.4000/1.6000	μ/μ
W_9/L_9	2.4000/1.8484	μ/μ
W_{12}/L_{12}	4.8551/1.6000	μ/μ
W_{15}/L_{15}	11.8890/1.6000	μ/μ
Cc	0.2408	pF
Ib	3.2032	μA

For each of the other 28 cases, the required CPU time is less than 60 s and the required number of iterations are 8 to 11. Note that in this example, the required CPU time is more than that in Example 1. All the solution points, which are obtained by starting from the four deterministic and 25 random initial guesses, were compared. It is found that in these solution points, the maximum deviation in W_i's and L_i's are, respectively, 0.0077 μm and 0.0032 μm. The performance of the circuit at the final design point was predicted using the new model, SPICE simulation, and the S–H model, which are given in Table IX. Note that the predictions based on the new model and SPICE simulation are quite close to the specifications, while the S–H model based prediction overestimates some of the performance metrics.

Example 4: In this example, the topology is the most complex op-amp among the considered class of op-amps. Both the first and second stages of the op-amp are cascoded. The design specifications are shown in the second column of Table XII. Like the other examples, here the op-amp was designed by starting from the four widely varying deterministic initial design points and 25 uniformly distributed random initial design points. The optimization statistics for the starting point with all transistors at minimum sizes are provided in Table XIII. Note that because of the circuit complexity, the required iteration number is high. However, it converges steadily to the final solution point and the required CPU time is less than three minutes. The final solution design point is given in the Table XIV. The design optimizations, by starting with the other initial design points, also converge to the final design point within five minutes CPU time and with less than 30 iterations. All the solution points, which are obtained by starting from the four deterministic and 25 random initial guesses, are very close to each other. It is found that in these solution points, the maximum deviation in W_i's and L_i's are, respectively, 0.0135 μm and 0.0032 μm. The performance of the circuit at the final design point was predicted using the new model, SPICE simulation, and the S–H model, which are given in the Table XII. In the same table, the second column provides the specifications. Like the previous example, here the predictions based on the new model and SPICE simulation are

TABLE XII
SPECIFICATIONS, NEW MODEL BASED, AND S–H MODEL BASED PREDICTED PERFORMANCES AND SPICE SIMULATED PERFORMANCES AT THE FINAL DESIGN POINT OF THE OP-AMP IN EXAMPLE 4

Perf.(unit)	Spec.	New model pred.	Spice simul.	S-H model pred.
A(0) (dB)	120	120	117	137
UGF (MHz)	10	10	9.5	9.7
PM (deg)	60	66	68	55
BW (Hz)	–	10	14	1.4
CMRR (dB)	80	99	103	109
CMR (V)	-2.5,1.0	-2.9,1.0	-3.0,1.5	-2.7. 0.24
OS (V)	-2,2	-2,2	-2.1,2.3	-2.3, 2.2
SR (V/μsec)	20	20	19	12.6
PD (mW)	–	0.33	0.32	0.25
Total gate area ($\mu \times \mu$)	–	110.59	–	110.59

TABLE XIII
OPTIMIZATION STATISTICS IN EXAMPLE 4 BY STARTING FROM SMALL SIZES OF ALL THE TRANSISTORS IN THE OP-AMP

It.no.	CPU time(sec)	MaxErr in L_i (μ)	MaxErr in W_i (μ)	MaxErr in V_n (V)
1	39.0100	–	–	–
2	15.3200	0.3908	4.3495	0.3503
3	10.6200	0.7167	1.6212	0.1707
4	9.1900	0.3055	0.8367	0.1422
5	7.8300	0.1308	0.6928	0.0797
6	7.9100	0.0691	0.6532	0.0510
7	7.9400	0.0388	0.5257	0.0279
8	9.3000	0.0207	0.4044	0.0147
9	7.8200	0.0122	0.2852	0.0077
10	11.5400	0.0078	0.1970	0.0038
11	8.0200	0.0052	0.1327	0.0017
12	7.8100	0.0035	0.0895	0.0011
13	7.9700	0.0025	0.0604	0.0008
14	5.7000	0.0018	0.0412	0.0006
15	5.9300	0.0013	0.0285	0.0004
16	5.8300	0.0010	0.0200	0.0004
17	6.6800	0.0009	0.0143	0.0004
	(total 174.4300)			

TABLE XIV
THE OPTIMAL DESIGN POINT IN EXAMPLE 4

Variables	Values	Units
W_0/L_0	2.40000 / 2.3731	μ/μ
W_1/L_1	2.40000 / 2.6992	μ/μ
W_2/L_2	2.40000 / 2.6992	μ/μ
W_3/L_3	2.40000 / 2.2437	μ/μ
W_4/L_4	2.40000 / 2.2437	μ/μ
W_5/L_5	4.80000 / 2.0728	μ/μ
W_6/L_6	4.80000 / 2.0728	μ/μ
W_7/L_7	2.6335 / 1.60000	μ/μ
W_8/L_8	2.6335 / 1.60000	μ/μ
W_9/L_9	2.40000 / 2.1607	μ/μ
W_{12}/L_{12}	3.4874 / 1.60000	μ/μ
W_{13}/L_{13}	2.4000 / 1.60000	μ/μ
W_{14}/L_{14}	2.6260 / 1.60000	μ/μ
W_{15}/L_{15}	15.2867 / 1.60000	μ/μ
W_{16}/L_{16}	2.5870 / 2.0728	μ/μ
W_{17}/L_{17}	2.6335 / 1.60000	μ/μ
Cc	0.2891	pF
Ib	4.2128	μA
Vb0	1.3468	V
Vb1	0.2962	V
Vb2	1.1623	V

quite close to the specifications, while the S–H model based prediction overestimates some of the performance metrics.

VIII. DISCUSSION

The results presented so far have left a few questions unanswered, while also opening up new avenues of research. We address these points in this section.

The key concept in our approach is that the CMOS op-amp sizing problem can be formulated as a sequence of (convex) geometric programs. This is achieved by modeling V_{GT}, g_m, and g_d as a PoP function of the transistor sizes and the bias current at a "relaxed" estimate of the dc operating point [see (1) and (58)]. Because of this iterative formulation, as the iterants proceed and approach convergence, the coefficient and powers of the first-order PoP model are made accurate via the use of second order model functions, such as those described in (59)–(63) for the α–A model that we have introduced. While such an approach maintains convexity in each optimization step, the need for a special model (e.g., α–A) that will have to accurately mimic a standard model (e.g., BSIM3) could be viewed as a drawback. Therefore, we now outline an approach to sequential convex programming (SCP) that does not use any special models.

The key observation is that with a log transformation, a PoP function becomes a linear function whose coefficients can be easily determined using linear least-squares curve fitting techniques [33]. Thus, in the inner loop of the flow graph of Fig. 4, we would only use a standard model in the "device model" box and the only technology in the "find constants and powers …" box would be linear-response surface modeling. Specifically, one would first model the I_D of the current source transistors and then model the V_{GT}, g_m, and g_d of the other transistors using the standard device model for function evaluation. While doing this modeling, the quantities $V_{\rm DS}$ and $V_{\rm SB}$ would be kept at the values determined in the previous iteration of the SCP. Note in this context that for a PoP model, if I_D is a PoP function of V_{GT}, W, and L, then so is V_{GT} of I_D, W, and L, i.e., I_D and V_{GT} are completely interchangeable. Further, as the iterants of the SCP proceed to convergence—as indicated by the proximity of the solutions and the reduction of the maximum constraint violation (see Table X)—the region over which the PoP response surface is built can be shrunk, resulting in a better fit with the standard model. Note that this approach is reminiscent of the iterative simulate-approximate-optimize strategy of CENTER [35].

There is one cautionary note, however. As the independent variables are varied in order to build the response surface, it is important that the device model stays in the domain (e.g., saturation) in which a PoP model is a good approximation. Thus, traditional fractional-factorial design [33], [34], which places sample points at the corners of the experiment-design box, should be avoided. Instead, Latin-hypercube sampling [36] should be used since it spreads the samples more uniformly in the box [35]. Further, those samples that do not fall in the required regime of device operation, i.e., saturation, should be discarded.

It should be noted that the approach outlined above, though promising, has not yet been implemented. It remains to be seen whether this approach of using only standard models will converge to the same final solution from widely varying initial guesses, as has been demonstrated in this paper for the α–A model.

The next issue that we will address is that of restricting the devices to operate in the saturation region. We do this because in standard CMOS op-amp design, the mosfets that are used as loads or amplification devices are biased in the saturation region for, among other reasons, the low g_d that is achievable in this region [28]. There are of course specific exceptions to this rule, e.g., when a parallel connection of an NMOS and PMOS device is used to build a resistor, or the common mode feedback transistor in a fully differential two-stage op-amp. At this time, we handle these situations as special cases that we can fit into our paradigm. However, the question remains as to whether the SCP approach can be applied to MOS circuits without any concern about the region of operation of the individual devices, or even for bipolar circuits. The answer depends on how well the iterative model-optimize approach that we have outlined in the context of using a standard model can be extended to these situations. Essentially, what we require is that the derived device parameters, e.g., g_m and g_d, be modeled as PoP functions of the independent design variables and that these PoP models become accurate approximations of the original device models as the iterants converge.

A requirement of our approach is that the performance constraints have to be manipulated into the form of an upper bounded posynomial. While we have shown that this is indeed

possible for a large class of CMOS op-amps and a large number of specifications, this is nevertheless a restriction. Indeed, our investigations with complex folded-cascode op-amps show that it is very simple to cast performance specifications as *signomials* (unlike posynomials, these are sums and *differences* of PoP terms). We would then have a "signomial programming problem" (where the direction of the constraint inequality is not an issue). This can be solved as a sequence of geometric programs [23] and would fit directly into our existing sequential convex programming approach.

The various issues discussed so far in this section have also served to highlight the superiority of our approach over that of GPCAD [16]. Since the approach in [16] is that of approximating and solving the CMOS op-amp sizing problem as a single geometric program, i.e., one pass of the loop shown in Fig. 4, their approach could be viewed as a first-order-approximate globally optimal solution. Further, none of the possible extensions discussed so far in this section are applicable to [16], since these ideas are predicated on our sequential convex programming formulation. Thus, it would be fair to say that our approach is more general than that of [16].

Note also that the SCP provides a sequence of iterants that are globally optimal solutions of convex subproblems. Further, our computational results show that this sequence converges to the same final solution for widely varying initial guesses. While these are very appealing features, it should be borne in mind that this does not guarantee that the final solution is the global optimum of the original sizing problem.

Not withstanding the extensions described above, there may remain certain performance metrics like settling time, which cannot be modeled as a suitable analytic function. Nor is it likely that symbolic analysis [14], which is excellent for generating expressions for ac performance metrics, would provide a solution. While it is possible to meet a given settling time specification by suitably constraining slew rate, unity gain frequency, and phase margin (which are modeled as posynomials), this nevertheless requires user intervention. Indeed, as has been observed in [8], even the one-time user effort required to derive analytic expressions of performance metrics (for a new op-amp) is a barrier to the widespread use of techniques such as ours. This is a barrier worth breaching in order to bring the major advantages of convex optimization into a truly automated circuit-sizing tool.

The best bet for achieving this is response surface modeling. The idea is essentially the same as the one previously outlined for accommodating standard device models. The difference is that instead of building PoP functions based on model evaluations, we would need to build posynomial or signomial models of the circuit performance metrics (e.g., settling time) in terms of the independent design variables using standard circuit simulation for the "experimental measurements." Unfortunately, the rich history of response surface modeling [34] has been focused on modeling polynomials rather than the more complex posynomials or signomials. In particular, determining the optimal number of PoP terms for the posynomial model and picking the correct powers for these terms without requiring prior analytical knowledge of the physical quantity being modeled are difficult problems for which only limited results are available [37]. Extending these results and applying them to the analog circuit sizing problem within the framework of sequential convex programming would be a worthwhile area of research.

Finally, since we have formulated the performance metrics for the class of two-stage op-amps in the notation of [29], we trust that it is obvious that their results can be enhanced by replacing their integer-nonlinear programming by the more powerful integer-convex programming.

IX. Summary

An efficient technique for sizing CMOS op-amps has been proposed. In this method, the op-amp sizing problem is formulated as a sequential convex programming problem. Such a formulation has two major advantages as enumerated below.

1) Since the convex programming problem is very well understood, it is very straight forward to solve it in a robust and computationally efficient manner.
2) The sequence of solutions generated is a sequence of global optimal of convex programming subproblems. Intuition therefore suggests that the point to which this sequence converges is the globally optimal solution of the original problem. This belief is supported by experimental results, where it is shown that the method converges to the same final design point for widely varying initial guesses.

The method has been prototyped in MATLAB and applied to a number of two stage CMOS op-amps. The experimental results highlight the robustness and computational efficiency of the technique. Further, the optimal design point is qualitatively similar to one that would be picked by an expert designer.

Although in the initial version of the implementation the S–H MOS model was used, in the short channel length regime the accuracy of performance prediction was not very good. To address this problem, in the second version of the prototype a new MOS model, called the α–A model, has been used to replace the S–H model. A number of op-amps were sized using the new model. The experimental results show the accuracy of performance prediction at the final design point. Further, the results demonstrate that with the accurate model the sizing technique maintains its robustness.

Acknowledgment

The authors would like to thank J. Fishburn for bringing the work of Hershenson *et al.* to their attention and K. Singhal for sharing her insights into response surface modeling. The author would also like to thank the anonymous reviewers for their comments that have helped improve this paper.

References

[1] L. R. Carley, G. G. E. Gielen, R. A. Rutenbar, and W. M. C. Sansen, "Synthesis tools for mixed-signal ICs: Progress on frontend and backend strategies," in *Proc. DAC*, June 1996, pp. 298–303.

[2] M. J. S. Smith, C. Portmann, C. Anagnostopoulos, P. S. Tschang, R. Rao, P. Valdenaire, and H. Ching, "Cell libraries and assembly tools for analog/digital CMOS and BiCMOS application-specific integrated circuit design," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1419–1432, Oct. 1989.

[3] F. El-Turky and E. E. Perry, "BLADES: An artificial intelligence approach to analog circuit design," *IEEE Trans. Computer-Aided Design*, vol. 8, pp. 680–692, June 1989.
[4] R. Harjani, R. A. Rutenbar, and L. R. Carley, "OASYS: A framework for analog circuit synthesis," *IEEE Trans. Computer-Aided Design*, vol. 8, pp. 1247–1265, Dec. 1989.
[5] M. G. R. Degrauwe *et al.*, "IDAC: An interactive design tool for analog CMOS circuits," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 1106–1115, Dec. 1987.
[6] ——, "Toward an analog system design environment," *IEEE J. Solid-state Circuits*, vol. 24, pp. 659–671, June 1989.
[7] W. Nye *et al.*, "DELIGHT.SPICE: An optimization-based system for the design of integrated circuits," *IEEE Trans. Computer-Aided Design*, vol. 7, pp. 501–519, Apr. 1988.
[8] E. S. Ochotta, R. Rutenbar, and L. R. Carley, "Synthesis of high-performance analog circuits in ASTRX/OBLX," *IEEE Trans. Computer-Aided Design*, vol. 15, pp. 273–294, Mar. 1996.
[9] F. Medeiro *et al.*, "A statistical optimization based approach for automated sizing of analog cell," in *Proc. Int. Conf. Computer-Aided Design*, 1994, pp. 594–597.
[10] M. Krasnicki, R. Phelps, R. Rutenbar, and L. R. Carley, "Maelstrom: Efficient simulation-based synthesis for custom analog cells," in *Proc. 36th DAC*, 1999, pp. 945–950.
[11] R. Phelps, M. Krasnicki, R. Rutenbar, L. R. Carley, and J. R. Hellums, "ANACONDA: Robust synthesis of analog circuits via stochastic pattern search," in *Proc. IEEE Custom Integrated Circuits Conf.*, 1999, pp. 567–570.
[12] H. Y. Koh, C. H. Sequin, and P. R. Gray, "OPASYN: A complier for CMOS operational amplifiers," *IEEE Trans. Computer-Aided Design*, vol. 9, pp. 113–125, Feb. 1990.
[13] G. G. E. Gielen, H. C. C. Walscharts, and W. M. C. Sansen, "Analog circuit design optimization based on symbolic simulation and simulated annealing," *IEEE J. Solid-State Circuits*, vol. 25, pp. 707–713, June 1990.
[14] ——, "ISAAC: A symbolic simulator for analog integrated circuits," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1587–1597, Dec. 1989.
[15] P. C. Maulik, L. R. Carley, and D. J. Allstot, "Sizing of cell level analog circuits using constrained optimization techniques," *IEEE J. Solid-State Circuits*, vol. SC-28, pp. 233–241, Mar. 1993.
[16] M. M. Hershenson, S. P. Boyd, and T. H. Lee, "GPCAD: A tool for CMOS op-amp synthesis," in *Proc. Int. Conf. Computer-Aided Design*, 1998, pp. 296–303.
[17] ——, "Automated design of folded-cascode opamps with sensitivity analysis," in *Proc. 1998 IEEE Int. Conf. Electronics, Circuits, Syst.*, vol. 1, pp. 121–124.
[18] M. M. Hershenson, S. S. Mohan, S. P. Boyd, and T. H. Lee, "Optimization of inductor circuits via geometric programming," in *Proc. 36th DAC*, 1999, pp. 994–998.
[19] M. M. Hershenson, A. Hajimiri, S. S. Mohan, S. P. Boyd, and T. H. Lee, "Design and optimization of LC oscillators," in *Proc. Int. Conf. Computer-Aided Design*, 1999, pp. 65–69.
[20] P. Mandal, "Synthesis and biasing of CMOS op-amps," Ph.D. dissertation, ECE Dept., Indian Inst. Sci., Bangalore, India, 1997.
[21] P. Mandal and V. Visvanathan, "A new approach for CMOS op-amp synthesis," in *Proc. 12th Int. Conf. VLSI Design*, 1999, pp. 189–194.
[22] C. Taumazou and C. A. Markis, "Analog IC design automation: Part I—Automated circuit generation: New concept and methods," *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 218–238, Feb. 1995.
[23] R. J. Duffin, E. L. Peterson, and C. Zener, *Geometric Programming—Theory and Application*. New York: Wiley, 1967.
[24] J. P. Fishburn and A. E. Dunlop, "TILOS: A posynomial programming approach to transistor sizing," in *Proc. Int. Conf. Computer-Aided Design*, 1985, pp. 326–328.
[25] S. S. Sapatnekar, V. B. Rao, P. M. Vaidya, and S. M. Kang, "An exact solution to the transistor sizing problem for CMOS circuits using convex optimization," *IEEE Trans. Computer-Aided Design*, vol. 12, pp. 1621–1634, Nov. 1993.
[26] H. Onodera, H. Kanbara, and K. Tamaru, "Operational-amplifier compilation with performance optimization," *IEEE J. Solid-State Circuits*, vol. 25, pp. 466–473, Apr. 1990.
[27] R. Gregorian and G. C. Temes, *Analog MOS Integrated Circuits for Signal Processing*. New York: Wiley, 1986.
[28] P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design*. New York: Holt, Rinehart and Winston, 1987.
[29] P. C. Maulik, L. R. Carley, and R. A. Rutenber, "Integer programming based topology selection of cell-level analog circuits," *IEEE Trans. Computer-Aided Design*, vol. 14, pp. 401–412, Apr. 1995.
[30] A. Grace, "Optimization Toolbox User's Guide," The MathWorks Inc., Nov. 1992.
[31] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, pp. 584–594, Apr. 1990.
[32] P. V. Wolf *et al.*, *An Introduction to the NELSIS IC Design System*. Delft, The Netherlands: Delft Univ. Technol., 1990.
[33] P. Mandal and V. Visvanathan, "Macromodeling of the a.c. characteristics of CMOS op-amps," in *Proc. Int. Conf. Computer-Aided Design*, 1993, pp. 334–340.
[34] G. E. P. Box and N. R. Draper, *Emperical Model Building and Response Surfaces*. New York: Wiley, 1987.
[35] K. Singhal, C. C. McAndrew, S. Nassif, and V. Visvanathan, "The CENTER design optimization system," *AT&T Tech. J.*, vol. 68, pp. 77–92, May/June 1989.
[36] R. L. Iman, J. C. Helton, and J. E. Campbell, "An approach to sensitivity analysis of computer models, Part I: Introduction, input variable selection and priliminary variable assesment," *J. Qual. Technol.*, vol. 13, pp. 174–183, 1981.
[37] N. R. Draper and H. Smith, *Applied Reression Analysis*. New York: Wiley, 1966.

Pradip Mandal received the B.E. degree in electronics and telecommunication engineering from the Bengal Engineering College, Shibpur, in 1989. He received the M.E. degree and the Ph.D. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1991 and 1999 respectively.

Dr. Mandal started his career in 1997 with Motorola India Electronics, Bangalore, India, where he has worked on development of design automation tools for VLSI circuit and SAW filter. In 1998 he joined Philips Semiconductors, Bangalore, where he is a Senior Technical Specialist in the area of I/O circuit design. His research interests are analog circuit design and design automation for analog and mixed-signal integrated circuits.

V. Visvanathan (S'77–M'82) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, the M.S.E.E. degree from the University of Notre Dame, IN, and the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley.

He was with AT&T Bell Laboratories, Murray Hill, NJ, and has been on the faculty of the University of Maryland, College Park, and the Indian Institute of Science, Bangalore. He is currently with Cadence Design Systems, Allentown, PA, where he is an Architect in the area of core algorithms for the simulation of analog and mixed-signal integrated circuits. He has two pending U.S. patents and has authored or coauthored over 50 papers in archival journals and conference proceedings in various areas of VLSI design and electronics design automation (EDA).

Dr. Visvanathan received the AT&T Bell Laboratories Distinguished Technical Staff Award in 1989 and the Honorable Mention Award at the 6th. Int. Conf. on VLSI Design, Bombay, India, in 1993.