Accurate metamodels of device parameters and their applications in performance modeling and optimization of analog integrated circuits

Liang Tao(梁涛)†, Jia Xinzhang(贾新章) and Chen Junfeng(陈军峰)

(Key Laboratory of Ministry of Education for Wide Band-Gap Semiconductor Materials and Devices, School of Microelectronics, Xidian University, Xi'an 710071, China)

Abstract: Techniques for constructing metamodels of device parameters at BSIM3v3 level accuracy are presented to improve knowledge-based circuit sizing optimization. Based on the analysis of the prediction error of analytical performance expressions, operating point driven (OPD) metamodels of MOSFETs are introduced to capture the circuit's characteristics precisely. In the algorithm of metamodel construction, radial basis functions are adopted to interpolate the scattered multivariate data obtained from a well tailored data sampling scheme designed for MOSFETs. The OPD metamodels can be used to automatically bias the circuit at a specific DC operating point. Analytical-based performance expressions composed by the OPD metamodels show obvious improvement for most small-signal performances compared with simulation-based models. Both operating-point variables and transistor dimensions can be optimized in our nesting-loop optimization formulation to maximize design flexibility. The method is successfully applied to a low-voltage low-power amplifier.

Key words: CMOS analog integrated circuits; optimization; metamodels of device parameters; RBF interpolation
DOI: 10.1088/1674-4926/30/11/115008 **EEACC:** 2570A; 2570D; 1130B

1. Introduction

Automatic sizing of analog circuits continues to be a research focus for the EDA industry. Although the sizing accuracy can be ensured by commercial transistor models, simulation-based sizing is extremely slow due to the large number of iterations and circuit evaluations. Therefore, many researchers tend to speed up the optimization process by adopting a model-based technique. During model-based circuit sizing, the circuit's behavior is captured in a set of models describing the specifications as functions of circuit parameters. These models can be derived using handcrafted analysis or an automated simulation-based method[1−4]. Model-based sizing has the advantage of speed over simulation-based sizing but the accuracy is inferior.

In Refs. [2, 4], simulation-based posynomial and least squares support vector machine (LS-SVM) models are generated through parametric regression and non-parametric regression techniques respectively. Circuit simulation results are used to train the models. Both models are formulated in terms of device voltages or currents, which in turn are expressed in terms of device dimensions by fixing the channel length (L) of MOSFETs. This formulation obstructs the straightforward interpretation of these models especially when small-signal performances are concerned[2]. In addition, not all the transistors should be designed with fixed L when specifications cannot be met in this case. Fixing L is equivalent to cutting away a part of the design space. When device dimensions have to be considered simultaneously, the number of optimization variables increases rapidly in large circuits and the model construction cost is high since an increase of dimensionality would need more training points for accurate models[5]. Performance models based on LS-SVM fitting are more accurate than posynomial models which have to be cast in posynomial format[4]. For some circuit characteristics, posynomial performance models bring about large approximation errors[6].

In this paper, a novel knowledge-based circuit sizing optimization method is developed. The symbolic expressions for circuit characteristics can be expressed in terms of operating-point variables or device parameters consistent with designers' habits without simulations. Device parameters are represented by the operating point driven (OPD) metamodels at BSIM3v3 level accuracy through radial basis functions (RBF) interpolation. Both the models of device parameters and circuit performances could be reused multiple times on a given silicon technology, which greatly reduces the model construction cost. Optimization variables consist of operating-point variables and transistor dimensions as well to maximize design flexibility.

2. Overview of the proposed methodology

The major sources of discrepancies between analytical performance expressions and simulation results are two-fold. First, since most expressions are derived based on correct biasing of all the transistors in the circuit, the error in estimating the circuit biasing can result in a significant prediction discrepancy[6]. Second, the discrepancy is inherent because accurate device parameters are not included in the expressions. To reduce the prediction discrepancies, the OPD metamodels for device parameters are applied in the optimization

† Corresponding author. Email: t_liang@yahoo.cn
Received 2 June 2009, revised manuscript received 10 July 2009

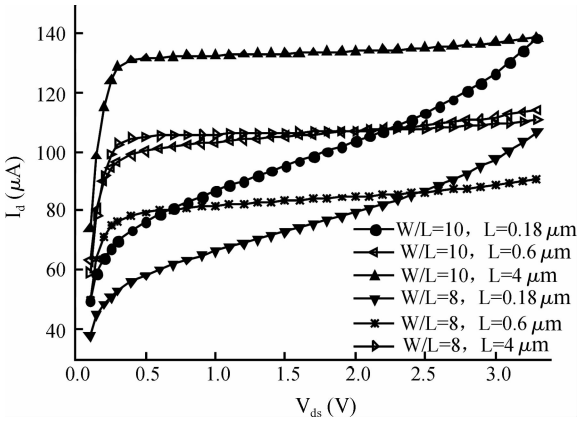

Fig. 1. Output characteristic curves of NMOS obtained from BSIM3v3 device models (V_{gs} = 0.7 V, V_{sb} = 0).

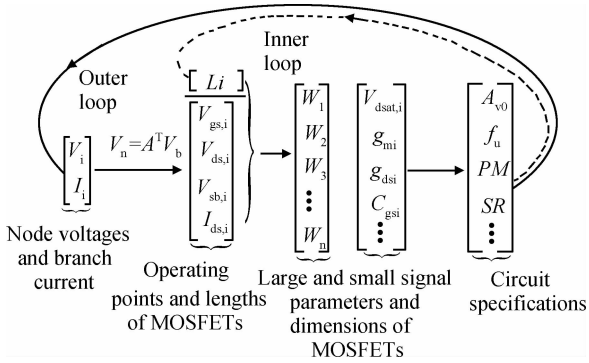

Fig. 2. Nesting-loop optimization formulation.

loop. Instead of estimating the circuit biasing by solving it iteratively[7] or hierarchically[8], the operating point is specified directly[9], which brings about no iterations and no convergence problems suffered by other methods. Metamodels for various large or small signal parameters including the ratio of width to length (W/L) of the transistor are built in terms of operating points and the channel lengths of the transistor by biasing the BSIM3v3 model at nearly all possible combinations of port voltages with different values of the transistor dimensions, which is called operating point driven metamodels. Including the channel lengths of MOSFETs into device modeling, circuit performance modeling and circuit optimization is significant because MOSFETs with the same DC operating point, the same W/L ratio and different channel lengths may differ greatly in large or small signal behavior as shown in Fig. 1. The output characteristic curves intersect with each other due to short-channel phenomena such as velocity saturation and drain-induced barrier lowering under deep-submicron technology[10], which means the same DC operating point is shared by them. Obviously, small signal parameters such as drain–source conductance g_{ds} differ a lot in the intersection points.

The optimization loop is depicted in Fig. 2. In the outer loop, operating points are cast as optimization variables and in the inner loop, the operating points are then invariable and the L's serve as optimization variables. In the inner loop, analytical-based performance expressions are formulated by the accurate OPD metamodels. Since the two major sources of prediction errors are eliminated, the analytical-based performance models can be at SPICE level accuracy. This nesting-loop optimization formulation can handle a larger number of optimization variables and is suitable for circuit problems of high dimensionality. From the above, we see that the OPD metamodels play a critical role in this methodology.

Table 1. Design variable mapping using a Latin hypercube.

Variable	value
V_{gs}	$d_1(V_{gs,max} - \hat{V}_{th}) + \hat{V}_{th}$
V_{ds}	$d_2(V_{ds,max} - \hat{V}_{dsat} + \hat{V}_{dsat}$
V_{sb}	$d_3(V_{sb,max} - V_{sb,min}) + V_{sb,min}$
L	$d_4(L_{max} - L_{min}) + L_{min}$
r	$d_5(r_{max} - r_{min}) + r_{min}$

3. OPD metamodel construction

A metamodel is a "model of a model" that provides an approximation to a physics-based model that is much faster to execute[11]. In addition, metamodels can also discover the unknown relationship between variables that physics-based models are unable to give directly. For example, BSIM3v3 models accept device dimensions and port voltages as the inputs and output small and large signal parameters such as I_d; the dependence of W upon operating points can then be captured by a metamodel. The creation of metamodels often uses design of experiments (DOE) to select a limited but optimal set of sample points in the design space[12]. A well tailored DOE plan and metamodel choice are the two key points during metamodeling. In this section, these two points are introduced in detail.

3.1. Data sample strategy for the metamodeling of BSIM3v3 device parameters

As stated in section 2, a vast number of combinations of operating points and channel lengths of MOSFETs are needed for metamodel construction. Although the implementation of all these combinations by BSIM3v3 device models is in some sense computationally inexpensive, strategies from DOE ensure a uniform and unbiased representation of the sample space. Under most cases, MOSFETs of analog circuits should be biased in saturation. This constraint can be expressed as $V_{gs} \geqslant V_{th}$ and $V_{ds} \geqslant V_{dsat}$. V_{th} and V_{dsat} are threshold voltage and saturation voltage respectively. Both of them are the outputs of BSIM3v3 models. We find that V_{th} can be approximated by linear regression of L and V_{sb} very accurately and V_{dsat} can also be approximately expressed as the function of L, V_{gs} and V_{sb} via response surface modeling. The ratio variable $r = W/L$ is used as a substitute for W in our problem. Instead of using port voltages and device dimensions directly in DOE, a five-factor Latin hypercube sampling (LHS) is carried out beforehand with values uniformly distributed on [0, 1]. Then, a set of sample points of port voltages and device dimensions are generated according to Table 1. In Table 1, d_1, d_2, d_3, d_4 and d_5 form the original five-factor Latin hypercube, $\hat{V}_{th}$ and $\hat{V}_{dsat}$

are given by the fitted functions other than BSIM3v3 model equations.

According to Table 1, V_{sb}, L and r are independent variables and the sample value of V_{gs} depends on them. Further, V_{ds} can be decided by the other 4 variables. The sample dataset can not only represent the design space but also ensure transistors biased in saturation during training data generation. L_{max} need not be very large because the transistor characteristics of large L will be identical and close to the behavior of the long-channel model described in Refs. [1, 13] when a threshold value of L is reached. Transistors of very large dimensions can be divided via multiplier (M) as multiple parallel MOSFETs. For this reason, r_{max} need not be very large either. After the sample dataset of device dimensions and port voltages are imported into BSIM3v3 models, various large and small signal device parameters are obtained. Here the following 7 MOS parameters are modeled in terms of operating points and channel lengths of the devices: V_{dsat}, g_m, g_{ds}, C_{gs}, C_{gd}, C_{db} and r.

Unlike the time-consuming circuit simulations during performance metamodel construction, this process will take seconds even though a large number of sample points is involved.

3.2. Scaling of the generated training data

Data scaling is an essential step to improve the training quality during metamodeling. When the data are properly scaled, the relationship between variables becomes weakly nonlinear. Most equations encountered in analog circuit design have logarithmic behavior[9, 10]. For example, $\log I_d$ depends in an asymptotically linear way on $\lg V_{gs}$ and $\lg V_{ds}$. The same is valid for the small-signal parameters. Thus variables $v_i \in [lb_i,\ ub_i]$ is logarithmically scaled onto $x_i \in [0, 1]$ as below before training the dataset[2]:

$$x_i = \lg \frac{v_i}{lb_i} \Big/ \lg \frac{ub_i}{lb_i}. \tag{1}$$

3.3. Metamodeling through RBF interpolation

RBF is a powerful tool for interpolating/approximating multidimensional scattered data. The method tries to approximate any amount of data samples by a linear combination of a radially symmetric function based on Euclidean distance or another such metric. For a set of n d-dimensional data points x_j and corresponding data values $f(x_j)$ (scattered data), an RBF method creates an interpolant function:

$$s(x) = c_0 + c_1 x + \sum_{j=1}^{n} \lambda_j \phi(\|x - x_j\|_2),\ \ x \in R^d, \tag{2}$$

where $\phi(\cdot)$ is the scalar radial basis function and c_0, c_1, λ_j are coefficients. The interpolation should fulfill the condition $s(x_j) = f(x_j)$ for $j = 1, \ldots, n$. Hence, c_0, c_1 and λ_j should satisfy the following linear equation:

$$C_0 + [x_1, x_2, \ldots, x_n]^T c_1^T + \Phi\lambda = f, \tag{3}$$

where $f, C_0, \lambda \in R^n$ have the components $f(x_j)$, c_0, λ_j, $j = 1, \ldots, n$, respectively. Φ is the $n \times n$ symmetric matrix with elements $\Phi_{ij} = \phi(\|x_i - x_j\|)$, $1 \leqslant i, j \leqslant n$. Here, we pick the function $\phi(r) = r^3$, $r \geqslant 0$. Once the coefficients are found, $s(x)$ can be used to estimate the value of the original function $f(x)$ at any point. A whole interpolation procedure needs to compute the distance between multidimensional data points and solve an $n \times n$ system. Some efficient numerical methods have been developed to make the procedure easy to implement even if huge amounts of sample points are involved[14], whereas metamodeling by other methods such as kriging[11] and LS-SVM[4] always needs to tune the parameters to obtain good performances by solving a multidimensional non-linear optimization problem, which always makes the computational cost intolerable with a large dataset.

During the construction of the metamodels, some device characteristics can be modeled by operating points and L accurately, and others cannot. They could be modeled by adding more parameters as explanatory variables. For example, the variation of C_{gs}, C_{gd} and C_{db} could be interpreted by operating points, L, V_{dsat}, g_m and g_{ds} precisely, where V_{dsat}, g_m and g_{ds} are certainly in the form of metamodels.

Metamodels are to be validated before being used as a "surrogate" of the physics-based model. When additional points are used for validation, there are a number of different measures of model accuracy. Two of them are used here to assess the fit quality of the generated RBF models, the quality-of-fit parameter (q), also known as normalized root mean-square (RSM) error, and the average relative error (ARE):

$$q = \frac{\sqrt{\sum_{k=1}^{num} (\hat{y}_k - y_k)^2}}{\sqrt{num}\left(\max\limits_{k=1}^{num} y_k - \min\limits_{k=1}^{num} y_k\right)}, \tag{4}$$

$$\mathrm{ARE} = \frac{1}{num} \sum_{k=1}^{num} \frac{|\hat{y}_k - y_k|}{y_k} \times 100\%, \tag{5}$$

where num is the number of validation points. $\hat{y}_k$ is the corresponding predicted value for the observed value y_k. To demonstrate the predicting ability of the OPD metamodels, both $\hat{y}_k$ and y_k are data values of additional confirmation samples before scaling. The smaller the values of q and ARE, the more accurate the models.

3.4. Applications of the OPD metamodels in performance modeling

Aided by accurate OPD metamodels of MOSFETs, both simulation-based and analytical-based performance models can be built easily and accurately. Seemingly rude handcrafted performance expressions can achieve relatively high accuracy, especially for small-signal circuit characteristics. Meanwhile, simulation-based performance models can also be improved by the metamodels in the generalization ability and flexibility. Therefore, different kinds of performance models can be integrated into one optimization cycle through the OPD metamodels. The OPD metamodels can also serve as an automatic biasing tool during manual or automated circuit design by using the metamodels of the ratio r.

Table 2. Validation measures of the OPD metamodels (range of data sample: $V_{gs} \leqslant 1$ V, $V_{ds} \leqslant 1.8$ V, $0.18\ \mu m \leqslant L \leqslant 2.5\ \mu m$, $1.4 \leqslant r \leqslant 40$).

		V_{dsat}	g_m	g_{ds}	r	C_{gs}	C_{gd}	C_{db}
NMOS	ARE	0.16896	0.24082	0.37707	0.26451	0.36229	0.38245	0.29342
	q	0.00139	0.00105	0.00150	0.00172	0.00073	0.00119	0.00088
PMOS	ARE	0.16276	0.21775	0.39753	0.24061	0.29797	0.44898	0.27422
	q	0.00227	0.00099	0.00224	0.00157	0.00070	0.01206	0.00111

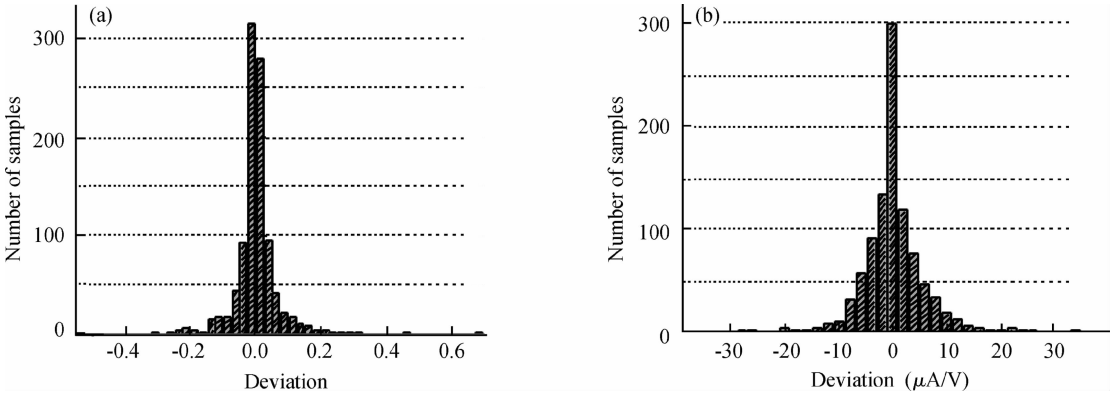

Fig. 3. Model deviation histograms for (a) r and (b) g_m.

4. Experimental results

4.1. Prediction ability of the OPD metamodels

The OPD metamodels are based on the technology of the 1.8 V 0.18 μm CMOS process. According to whether the source and bulk are shorted, 4000 samples and 5000 samples were generated respectively to train the models. Thus the obtained metamodels can be categorized in two groups. As mentioned in Section 3, here L_{max} is no more than 2.5 μm and r is between 1.4 and 40. Thus, the range for W is from 0.25 to 100 μm. Another 1000 sample points were generated likewise to assess the fit quality of the metamodels. The validation measures for the 7 parameters of MOS devices when $V_{sb} = 0$ are reported in Table 2. Our modeling approach brings the maximal normalized RMS error under 1.2%. In Fig. 3, the actual error distributions of ratio and g_m are given using error histograms. Clearly, the opportunity to produce large errors is extremely small for RBF metamodels.

4.2. Knowledge-based circuit optimization aided by OPD metamodels

Our methodology was applied to the low-voltage low-power amplifier shown in Fig. 4. The specifications are given in Table 3. First, to show the usage of the OPD metamodels as in the inner loop optimization, the circuit was successfully biased at nearly the same DC operating point with all MOSFETs taking five different lengths by using the metamodels of the ratio r. The simulated operating points for these five cases are listed in Table 4 for comparison with the assigned nominal operating points values (represented by $V_{norm,i}$ and $I_{norm,i}$). The voltages all are transistor port voltages. The currents have been indicated on the schematic. We impose equal DC port voltages for matching transistors, input and output node voltages are fixed by the circuits of the test

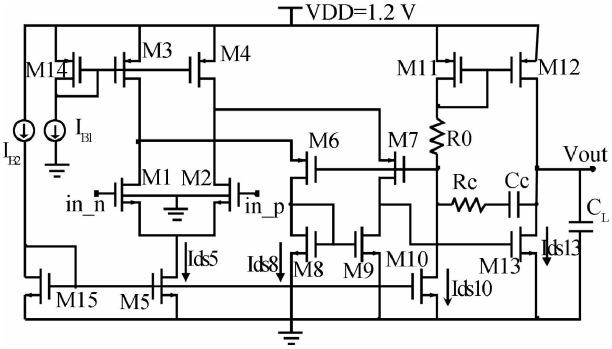

Fig. 4. Schematic of a low-voltage low-power amplifier.

Table 3. Specifications and obtained performances of the amplifier ($V_{dd} = 1.2$ V, $C_L = 10$ pF).

Performance	Required	Obtained	Predicted
A_{v0} (dB)	Max	94.63	94.82
f_{-3dB} (Hz)		261.67	263.68
f_u (MHz)	$\geqslant 10$	13.53	14.53
PM (°)	$\geqslant 60$	67.81	$\geqslant 60$
Gain_{CM} (dB)	$\leqslant -3$	−3.44	−3.10
Power (μW)	$\leqslant 150$	135.50	135.66
Output swing	[0.2, 1]	[0.197, 1.010]	[0.184, 1.005]
SR^+ (V/μs)	$\text{Min}(SR^+, SR^-) \geqslant 2$	2.73	3.04
SR^- (V/μs)	$\text{Min}(SR^+, SR^-) \geqslant 2$	11.37	9.76

bench. Several circuit performances are also listed in Table 4 for comparison. Although biased at nearly the same DC operating point, the circuit performance varies greatly with channel length, which demonstrates the necessity to consider operating points and channel lengths simultaneously in circuit optimization and performance modeling. The dimensions of the transistors in these five cases are shown in Table 5.

As stated in section 2, all transistors should be biased

Table 4. Comparison of simulated operating points for different channel lengths.

	Nominal value		Simulated value for different L				
	$V_{\text{norm, i}}$	$I_{\text{norm, i}}$	0.35 μm	0.5 μm	0.8 μm	1 μm	2 μm
$V_{\text{sd,3}}$ (V)	0.4	-	0.4010	0.4127	0.4026	0.4008	0.3996
$V_{\text{ds, 5}}$ (V)	0.4	-	0.4082	0.3996	0.4014	0.4003	0.3996
$V_{\text{gs, 8}}$ (V)	0.65	-	0.6456	0.6508	0.6505	0.6498	0.6501
$V_{\text{ds, 9}}$ (V)	0.65	-	0.6478	0.6527	0.6511	0.6501	0.6497
$V_{\text{ds, 10}}$ (V)	0.3	-	0.3015	0.2887	0.2979	0.2997	0.3012
$V_{\text{sg, 11}}$ (V)	0.65	-	0.6491	0.6539	0.6510	0.6500	0.6496
$V_{\text{sg, 14}}$ (V)	0.65	-	0.6493	0.6506	0.6505	0.6499	0.6500
$V_{\text{gs, 15}}$ (V)	0.6	-	0.6009	0.6046	0.6000	0.5998	0.6002
$I_{\text{d, 5}}$ (μA)	-	20	20.34	20.99	20.04	20.05	20.03
$I_{\text{d, 8}}$ (μA)	-	15	14.74	14.70	15.06	14.94	15.00
$I_{\text{d, 10}}$ (μA)	-	10	9.98	10.30	10.04	10.01	9.97
$I_{\text{d, 13}}$ (μA)	-	40	39.63	41.10	40.32	40.12	39.92
A_{v0} (dB)			68.48	76.01	83.13	86.38	95.50
$f_{-\text{3dB}}$ (Hz)			4847	2099	869.93	594.05	198.64
PM (°)			88.06	86.43	82.42	79.01	55.94

Table 5. Ratios of transistors of different channel lengths with identical circuit biasing point.

		$r_1 = r_2$	$r_3 = r_4$	r_5	$r_6 = r_7$	$r_8 = r_9$	r_{10}	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}
L (μm)	0.35	8.894	19.471	6.221	11.617	2.751	3.482	7.799	15.117	6.914	7.799	3.202
	0.5	7.093	19.563	5.728	10.598	2.684	3.155	7.727	15.219	6.505	7.727	2.966
	0.8	6.008	19.146	5.052	9.777	2.336	2.700	7.484	15.022	5.819	7.484	2.620
	1	5.614	19.031	4.854	9.441	2.210	2.547	7.457	15.039	5.629	7.457	2.488
	2	4.160	17.468	3.975	8.122	1.862	2.036	6.890	13.873	4.842	6.890	2.013
M		2	1	1	8	1	1	1	2	1	1	1

properly by which the derived symbolic expressions for circuit performances are workable. For all the transistors of the amplifier,

$$V_{\text{gs},i} - V_{\text{th},i} \geqslant V_{\text{od, min}}, \tag{6}$$

$$V_{\text{ds},i} \geqslant V_{\text{dsat},i}, \tag{7}$$

$$r_{\min} \leqslant r_i \leqslant r_{\max}, \tag{8}$$

$$L_{\min} \leqslant L_i \leqslant L_{\max}. \tag{9}$$

Instead of deriving the expressions for phase margin to guarantee the stability of the amplifier with unity-gain feedback, we set $C_{\text{c}} = 0.22C_{\text{L}}$, and $R_{\text{c}} = \frac{(C_{\text{c}}+C_{\text{L}})}{C_{\text{c}}}\frac{1}{g_{\text{m13}}}$ to achieve the same goal by canceling the output pole[13]. In addition, the high frequency pole should be placed high enough to guarantee the stability of the amplifier:

$$f_{\text{h}} = \frac{g_{\text{m6}}}{2\pi C_{\text{gs6}}} \geqslant 10 f_{\text{u}}. \tag{10}$$

The mirror pole of M8 and M9 also can be neglected provided the device dimensions are small. This can be fulfilled by limiting the channel lengths of M8 and M9 smaller than 1 μm.

Symbolic expressions for some circuit performances are given for illustration. The open-loop voltage gain is:

$$\begin{aligned} A_{\text{v0}} &= A_{\text{v1}}A_{\text{v2}} = g_{\text{m1}}R_{\text{o1}}g_{\text{m13}}R_{\text{o2}} \\ &= g_{\text{m1}}\left(\frac{1}{g_{\text{ds8}}} /\!/ \left(\frac{g_{\text{m6}}}{g_{\text{ds6}}}\frac{1}{g_{\text{ds1}}+g_{\text{ds3}}}\right)\right)\frac{g_{\text{m13}}}{g_{\text{ds12}}+g_{\text{ds13}}}, \end{aligned} \tag{11}$$

and the 3-dB bandwidth is given by:

$$f_{-\text{3dB}} = \frac{1}{2\pi R_{\text{o1}}C_{\text{t}}} = \frac{1}{2\pi R_{\text{o1}}(A_{\text{v2}}+1)C_{\text{c}}}. \tag{12}$$

The unity-gain bandwidth can be expressed as $f_{\text{u}} = A_{\text{v0}}f_{-\text{3dB}}$. The positive and negative slew rates can be approximately expressed as:

$$\text{SR}^+ = \frac{I_{\text{d,12}}}{C_{\text{L}}+C_{\text{c}}},\ \text{SR}^- = \frac{I_{\text{d,5}}}{C_{\text{c}}}. \tag{13}$$

In order to meet the requirement for output swing, the following two constraints should be imposed:

$$V_{\text{out, min}} \geqslant V_{\text{dsat, 13}},\ V_{\text{dd}} - V_{\text{out, max}} \geqslant V_{\text{dsat, 12}}. \tag{14}$$

In the inner loop of the optimization problem, the constraint functions are added as penalty terms to the goal function which expresses the design objectives. The L's serve as variables for the specific operating point passing from the outer loop. In the inner loop, the problem is solved using a pattern search algorithm which will be called by the outer loop iteratively. To reduce the number of iterations of the outer loop, a nearly global optimization approach is adopted by partitioning the variable space into several sections via DOE. For the amplifier in Fig. 4, the supply voltage and power constraints are relatively harsh. Consequently, operating points are sampled around the nominal value of Table 4. Transistor port

Table 6. Validation measures of analytical-based models and simulation-based models.

		A_{v0}	f_{-3dB}	f_u	CMRR	Power	SR^+	SR^-
Proposed method	ARE	4.1196	2.0362	6.0822	3.3796	0.2514	12.1041	8.8140
	q	0.0132	0.0083	0.0508	0.0125	0.0063	0.1371	0.0708
LS-SVM[4]	ARE	4.5303	4.3673	2.6564	4.1307	0.2894	0.8729	4.4602
	q	0.0137	0.0151	0.0206	0.0166	0.0075	0.0119	0.0323

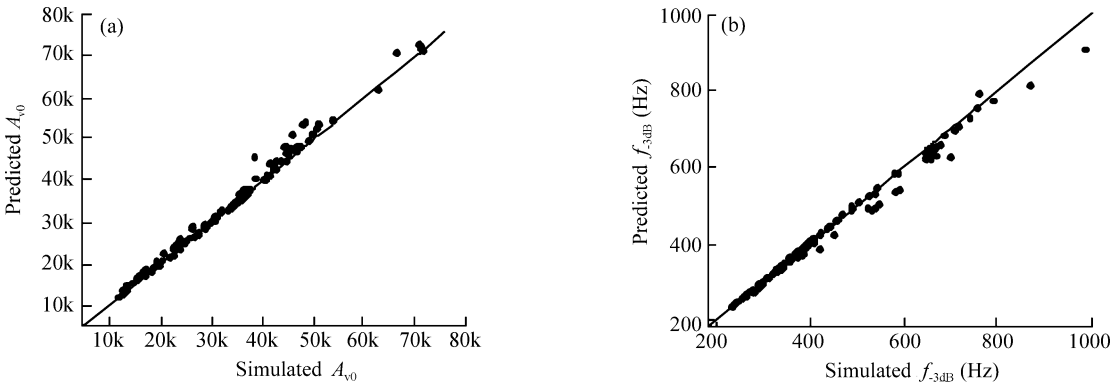

Fig. 5. Scatter plots of (a) A_{v0} and (b) f_{-3dB} for the validation data.

voltages are sampled between $V_{\text{norm},i} \pm 0.05$ V (let $V_{\text{gs},8} = V_{\text{ds},9}$) and drain currents are sampled between $(1\pm20\%)I_{\text{norm},i}$. Two current biases I_{B1}, I_{B2} are sampled between 8 and 12 μA. Twenty groups of operating points are generated by LHS. The optimum gives the highest value of the objective function as shown in Table 3. As seen from Table 3, the optimal design meets the specifications, simulated and predicted performances are close to each other.

The quality of analytical-based circuit performance models should be checked. From the twenty groups of operating points, two of them violate one of the constraint functions of Ineq. (6)–(10) till the last iteration of the L's. For each of the remaining eighteen operating points, validation points are chosen every five iterations from all the iterations satisfying the constraints Ineq. (6)–(10). Hence, a total of 166 data points are used to validate the models. The validation measures for some circuit performances are given in Table 6. The scatter plots of SPICE-simulated and model estimated A_{v0} and f_{-3dB} are shown in Fig.5. The data points locate near a straight line along the diagonal axis, which demonstrates the accuracy of the analytical-based models composed by the OPD metamodels of device parameters.

Finally, simulation-based performance models were constructed by the OPD metamodels using LS-SVM regression. The training dataset of operating points together with channel lengths of circuit devices are taken as follows: transistor port voltages are sampled between $V_{\text{norm},i} \pm 0.05$ V and drain currents are sampled between $(1\pm20\%)I_{\text{norm},i}$. I_{B1}, I_{B2} are sampled between 8 and 12 μA. Due to the matching requirement, channel lengths of circuit devices can be simplified into seven independent variables. The range for L is from 0.18 to 2.5 μm. The design space is therefore $X \subset R^{20}$. 600 groups of data points are generated by LHS. Aided by the metamodels of the ratio r, 600 SPICE netlists are generated and simulated. When the simulation results are collected, LS-SVM performance models can be built[4, 15]. Quality measures of LS-SVM models are listed in Table 6 using the same validation points above. As seen from Table 6, analytical-based models are superior in describing DC and small-signal circuit behavior except for f_u, while simulation-based models show advantages in capturing nonlinear large-signal circuit behavior[2]. Since f_u is expressed as the product of A_{v0} and f_{-3dB}, the approximation error of f_u is bigger than the errors of both A_{v0} and f_{-3dB} in our analytical-based performance models. f_u as well as SR could be expressed more accurately if more parasitic effects were added, which may not be necessary in this specific problem.

4.3. Discussion of different performance modeling methods

Although both analytical-based and simulation-based performance models can be built accurately via OPD metamodels, their generalization abilities differ a lot. The advantage of using analytical-based models in circuit optimization is that the models are adequate under different working conditions, which is not the case for simulation-based models. The simulation-based models have to be rebuilt if any working condition, such as the supply voltage, changes. To demonstrate the superiority of analytical-based models, suppose the constraint cast upon static power can be relaxed to achieve higher f_u. For the amplifier in Fig. 4, the channel length is taken to be 0.5 μm for each MOSFET. While all the transistor port voltages take the value of $V_{\text{norm},i}$ shown in Table 4, all the drain currents as well as current biases I_{B1}, I_{B2} are increased to 1.2 times, 1.5 times, 2 times and 3 times their original values respectively. The simulated and model predicted power–f_u curves are compared in Fig. 6. Although the points predicted by previously built LS-SVM models match the simulation results initially, the predicting accuracy begins to deteriorate drastically when branch currents are increased by 20%. Not surprisingly, the

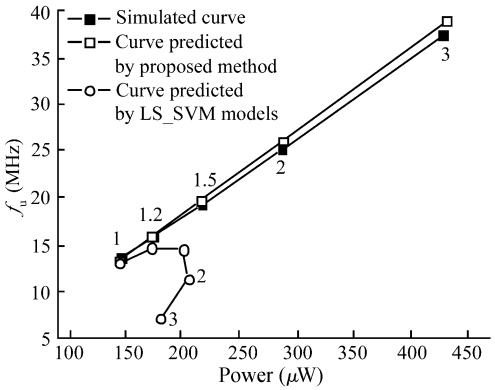

Fig. 6. Comparison of power–f_u curves between simulation and model-based methods.

generalization ability of LS_SVM models is poor when the values of branch currents locate near or beyond the boundary of the dataset scope where the models were built. In contrast, the power–f_u curve given by the analytical-based method coincides with the simulated one over a wide range, revealing the relation between power and f_u (constraints in Eqs. (6)–(10) are proved to be satisfied in all these five cases). Obviously, the analytical-based performance models can be reused in the new optimization problems without any change.

The main limitation of the proposed method is that symbolic expressions for circuit characteristics have to be derived. Despite the progress in computer-automated symbolic analysis techniques[16], these are still mainly limited to small-signal characteristics, necessitating the manual derivation of large-signal and transient characteristics. Although the analytical performance models are similar in form for a certain category of circuits, the procedure requires some user expertise in circuit design and optimization. When transient specifications must be described accurately, simulation-based performance models can be a useful alternative. Different kinds of performance models can be integrated into one optimization cycle through the OPD metamodels. Therefore, our nesting-loop optimization formulation is applicable to many other CMOS circuits such as large, complex mixed-signal circuits.

5. Conclusion

In this paper, a knowledge-based circuit sizing optimization method is presented. Analytical-based symbolic expressions of circuit performance are composed by operating point driven metamodels of MOS devices. By eliminating the two major prediction errors of traditional analytical-based performance expressions, the OPD metamodels are accurate at BSIM3v3 level and reusable for multiple circuits on a given technology. The OPD metamodels can also serve as an automatic biasing tool during manual circuit design. The simulated DC operating points are compared to the assigned values to ensure their correctness. The proposed analytical-based performances expressions are more generic than the simulation-based ones and provide more physical insights. In addition, both operating-point variables and transistor dimensions can be optimized in the nesting-loop optimization formulation of our paper. A low-voltage low-power amplifier was successfully biased and optimized using the proposed method. The results indicate that knowledge-based circuit sizing performs well aided by OPD metamodels of transistors. Our future work will be devoted to expressing transient specifications more accurately through reusable models.

References

[1] Hershenson M, Boyd S P, Lee T H. Optimal design of a CMOS op-amp via geometric programming. IEEE Trans Comput-Aided Des Integr Circuits Syst, 2001, 20(1): 1

[2] Daems W, Gielen G, Sansen W. Simulation-based generation of posynomial performance models for the sizing of analog integrated circuits. IEEE Trans Comput-Aided Des Integr Circuits Syst, 2003, 22(5): 517

[3] Gao Xuelian, Shi Yin. Generation of polynomial response surface models for sizing of an analog IC. Chinese Journal of Semiconductors, 2005, 26(11): 2241

[4] Kiely T, Gielen G. Performance modeling of analog integrated circuits using least-squares support vector machines. Proceedings of the Design, Automation, and Test in Europe, 2004: 448

[5] Harjani R, Shao J. Feasibility and performance region modeling of analog and digital circuits. Analog Integrated Circuits and Signal Processing, 1996, 10(1/2): 23

[6] Kim J, Lee J, Vandenberghe L. Techniques for improving the accuracy of geometric-programming based analog circuit design optimization. Proceedings IEEE/ACM International Conference on Computer-Aided Design, 2004: 863

[7] Maulik P C, Carley L R. Automating analog circuit design using constrained optimization techniques. Proceedings IEEE International Conference on Computer-Aided Design, 1991: 390

[8] Iskander R, Louërat M, Kaiser A. Automatic DC operating point computation and design plan generation for analog IPs. Analog Integrated Circuits and Signal Processing, 2008, 56(1/2): 93

[9] Leyn F, Gielen G, Sansen W. An efficient DC root solving algorithm with guaranteed convergence for analog integrated CMOS circuits. Proceedings IEEE/ACM International Conference on Computer-Aided Design, 1998: 304

[10] Liu Weidong, Jin Xiaodong, Xi Xuemei, et al. BSIM3v3.3 MOSFET model users' manual. Berkeley, CA: The Regents of the University of California, 2005

[11] Martin J D, Simpson T W. Use of adaptive metamodeling for design optimization. Proceeding of the 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, 2002

[12] Kleijnen J P C. An overview of the design and analysis of simulation experiments for sensitivity analysis. European Journal of Operational Research, 2005, 164(2): 287

[13] Allen P E, Holberg D R. CMOS analog circuit design. UK: Oxford University Press, 2002

[14] Wendland H. Scattered data approximation. New York: Cambridge University Press, 2005

[15] Suykens J A K, Gestel T V, Brabanter J D, et al. Least squares support vector machines. Singapore: World Scientific, 2002

[16] Fernández-Fernandéz F V, Rodríguez-Vázquez A, Huertas J L, et al. Symbolic analysis techniques—applications to analog design automation. Piscataway, NJ: IEEE Press, 1998