Gradient Optimization of RF Amplifiers for Digital Communications

Vuk Borich, Jack East, George Haddad

EECS Department, the University of Michigan, 1301 Beal Avenue,
2437 EECS, Ann Arbor, Michigan 48109

ABSTRACT: A gradient optimization methodology for use with nonlinear envelope simulation is presented. Emphasis is placed on efficient evaluation of cost function gradients. To this end, an envelope sensitivity equation is derived and methods for its efficient solution are proposed. In the case of circuits of moderate size, the solution of the sensitivity equation is shown to be simple and inexpensive. The more troublesome case of larger circuits is treated by a novel application of a recently developed iterative linear equation solver. The result is a general-purpose, rigorous optimization methodology for the fine tuning of gain, adjacent channel power, power efficiency, and related performance measures in radio frequency (RF) amplifiers for digital communications. As an illustrative example, the optimization of a feedforward-linearized power amplifier is presented.   © 2000 John Wiley & Sons, Inc. Int J RF and Microwave CAE 10: 353–365, 2000.

Keywords: circuit simulation; circuit optimization; envelope simulation

1. INTRODUCTION

There has been a growing interest in simulation technology that is capable of efficient circuit analysis with modulated signal sources. The development of the *nonlinear envelope* (NE) (also known as modulation-oriented harmonic balance, circuit envelope, envelope transient, etc.) method [1–3] was an important contribution, providing an excellent combination of accuracy, memory consumption, and robustness for narrowband applications. However, circuit optimization, in conjunction with the NE technology, has not been treated in the literature. The aim of this article is to bridge that gap by offering a general-purpose, rigorous, and efficient optimization methodology for use in the design of digital communications amplifiers.

Despite the known advantages of the NE method, modulated signal analyses are slow relative to periodic and quasiperiodic analyses by means of harmonic balance. Thus, any optimization algorithm to be used with the NE method must be designed in a way that minimizes the number of necessary iterations and, hence, the number of underlying simulations.

In this regard, gradient optimizers are viewed as primary candidates—they are known for fast convergence to the locally optimal solution and are particularly effective when the initial design is, in some sense, close to the optimal. However, in order to exploit the power of gradient optimizers to the full extent, the computation of the cost (objective) function gradient should be inexpensive. Ideally, the computational overhead (that is, the fraction of the computational effort expended on gradient calculation) should be negligible. This is especially important when designing an optimizer in conjunction with the NE method where individual iterations may be slow.

In principle, it is always possible to calculate the gradient by parameter perturbation using a procedure known as *finite-difference* gradient computation. Here, one evaluates the cost function using the current set of parameters; then, for each variable parameter, the evaluation of the cost function is repeated with a small change in the parameter; finally, the gradient is computed by the finite-difference approximation to the partial differentiation operator. Clearly, this results in as many additional cost function evaluations (and, hence, NE simulations) as there are variable parameters—a costly process that is in some cases impractical.

Instead, this article offers a considerably more efficient approach, based on direct sensitivity analysis of circuit envelope waveforms. Specifically, equations are derived for partial derivatives (sensitivities) of envelope waveforms with respect to variable parameters. As shown in the article, the gradients follow in a simple way from the knowledge of envelope sensitivities.

Most importantly, the article deals with efficient methods for the solution of the envelope sensitivity system. Here, it is necessary to treat two separate cases depending on the underlying solution of the NE equations. Ultimately, the solution of the NE system reduces to repeated solutions of linear algebraic systems similar to those that arise from the application of the Newton–Raphson algorithm to the method of harmonic balance. The conventional approach to the solution of such linear systems is based on the application of Gaussian elimination (direct) solvers. As circuit size grows, however, direct solvers become slow and memory consuming, presenting a problem that had plagued the method of harmonic balance until recently. In the last few years, the advent of iterative linear solvers [4] and their use in circuit analysis [5–9] extended the applicability of the method to much larger circuits. It will be shown that the underlying solution method for the NE equations (direct or iterative) has a significant impact on the efficiency of gradient computation. For smaller circuits where direct methods may be used effectively, gradient computation is simple and relatively inexpensive. For larger circuits which necessitate the use of iterative solvers, gradient computation carries more overhead. In the latter case, the numerical expense is mitigated by a novel application of a recently proposed iterative algorithm, known as the *hybrid generalized minimal residual method for multiple right-hand sides* [10] and abbreviated as MHGMRES.

As remarked earlier, gradient optimization works especially well when the initial set of variable parameters is, in some way, close to the optimal. Such an initial set is readily obtained by the more conventional analysis and optimization methods. For example, a modulated input signal may be approximated by a two-tone input of equal available power, and the optimization may be carried out by well-established means [11, 12]. Design optimization may also be carried out by means of memoryless models extracted from single-tone distortion characteristics. While both of the above provide a good initial guess, neither is rigorous for general nonlinear circuits with digitally modulated carrier excitations—two-tone signals are a poor approximation to some modulation formats, and the memoryless model may give inaccurate results by ignoring envelope dynamics and out-of-band nonlinear effects. Thus, the use of conventional methods is appropriate in early design stages. The proposed technique provides an accurate optimization tool for the fine tuning of the final design performance.

Several of the principles presented in this article have already been reported [13]. This article extends the authors' previous work to allow optimization of larger circuits in an efficient way. Specifically, "large-scale" sensitivity methods of Sections 4.3 and 4.4 are presented for the first time.

The article is organized as follows. Section 2 reviews the nonlinear envelope method and lays the foundation for the development of optimization and sensitivity analysis methods in Sections 3 and 4. A design example is presented in Section 5. Conclusions are reserved for Section 6.

2. NONLINEAR ENVELOPE METHOD

All performance measures, to be optimized by methods of Sections 3 and 4, are evaluated by the NE method, considered as the most appropriate method for nonlinear analysis of radio frequency (RF) circuits subject to narrowband modulated excitations [14]. The implication of the narrowband assumption will be apparent from the discussion. This section gives a review of the principles of NE simulation, providing the necessary background for subsequent optimization and sensitivity analysis; a detailed account of the method is necessary because of its intimate connection with the principles of sensitivity analysis

in Section 3. The discussion is based, with several modifications, on the theory presented in ref. [1].

Consider an amplifier consisting of sources, nonlinear *RC*, and linear *RLC* elements; linear distributed elements are included at the end of this section. The modified nodal formulation [15] leads to a system of *K* nonlinear equations in *K* circuit variables of the form

$$\frac{d}{dt}\mathbf{q}(\mathbf{x}(t)) + \mathbf{i}(\mathbf{x}(t)) = \mathbf{b}(t), \qquad (1)$$

where $\mathbf{x}(t) \in \mathbf{R}^K$ is the vector of node voltages and certain branch currents, $\mathbf{b}(t) \in \mathbf{R}^K$ is the vector of source contributions, and $\mathbf{q}, \mathbf{i} : \mathbf{R}^K \to \mathbf{R}^K$ are differentiable functions that describe the contributions of reactive and conductive elements, respectively.

The driving source is a modulated carrier of the form

$$v_s(t) = 2\,\mathrm{Re}\{\tilde{v}_s(t)e^{j\omega_c t}\}, \qquad (2)$$

where, by convention, the tilde denotes a complex envelope. Under the usual assumption of stable and bandlimited behavior, all circuit waveforms appear as

$$y(t) = \sum_{n=-N}^{N} \tilde{y}_n(t)e^{jn\omega_c t} = \tilde{\mathbf{y}}^T(t)\mathbf{w}(t), \qquad (3)$$

where $\tilde{\mathbf{y}}(t)$ is the $(2N+1)$-vector of real and imaginary envelope components and $\mathbf{w}(t)$ is the Fourier basis vector:

$$\begin{aligned}
\tilde{\mathbf{y}}(t) &= \begin{bmatrix} \tilde{y}_0(t) & \tilde{y}_1^R(t) & \tilde{y}_1^I(t) & \cdots & \tilde{y}_N^I(t) \end{bmatrix}^T, \\
\mathbf{w}(t) &= [1 \quad 2\cos(\omega_c t) \quad -2\sin(\omega_c t) \quad \cdots \\
&\quad -2\sin(N\omega_c t)]^T.
\end{aligned} \qquad (4)$$

When the assumed form of circuit waveforms is substituted in eq. (1), following some manipulation, each equation in the system may be expressed as

$$\begin{aligned}
\frac{d}{dt}\{\tilde{\mathbf{q}}_k^T(\widetilde{\mathbf{X}})\}\mathbf{w} + \tilde{\mathbf{q}}_k^T(\widetilde{\mathbf{X}})\Omega_0\mathbf{w} + \tilde{\mathbf{i}}_k^T(\widetilde{\mathbf{X}})\mathbf{w} \\
= \tilde{\mathbf{b}}_k^T\mathbf{w}, \qquad k = 1, \dots, K.
\end{aligned} \qquad (5)$$

Here, $\widetilde{\mathbf{X}}(t)$ is the block *K*-vector of real and imaginary components of circuit envelopes,

$$\widetilde{\mathbf{X}}(t) = \begin{bmatrix} \tilde{\mathbf{x}}_1^T(t) \cdots \tilde{\mathbf{x}}_K^T(t) \end{bmatrix}^T; \qquad (6)$$

Ω_0 is the matrix of time-differentiation of the Fourier basis,

$$\frac{d\mathbf{w}}{dt} = \Omega_0\mathbf{w}; \qquad (7)$$

and $\tilde{\mathbf{q}}_k, \tilde{\mathbf{i}}_k, \tilde{\mathbf{b}}_k$ are vectors of envelope components defined by analogy to eq. (4). Time dependencies have been omitted for clarity.

If the envelopes were constant in time (hence, if the source were a pure sinusoid), each of the *K* equations (5) could be converted to a system of $2N+1$ equations in the unknown envelope coefficients, by an application of the familiar orthogonality relationships for the Fourier basis. The critical assumption in the NE method is that similar orthogonality relationships hold, even in the time-varying envelope case, due to the narrowband nature of the signals and the implied near constancy of the envelope components over the carrier period $2\pi/\omega_c$. With this assumption, eq. (5) implies

$$\begin{aligned}
\frac{d}{dt}\tilde{\mathbf{q}}_k(\widetilde{\mathbf{X}}) + \Omega_0^T\tilde{\mathbf{q}}_k(\widetilde{\mathbf{X}}) + \tilde{\mathbf{i}}_k(\widetilde{\mathbf{X}}) = \tilde{\mathbf{b}}_k, \\
k = 1, \dots, K.
\end{aligned} \qquad (8)$$

Using eq. (8), one may recast the governing equation (1) in a system of $(2N+1)K$ equations in the $(2N+1)K$ unknown envelope components by defining

$$\begin{aligned}
\Omega &= \mathrm{diag}(\overbrace{\Omega_0 \dots \Omega_0}^{K\,\mathrm{times}}), \\
\widetilde{\mathbf{B}}(t) &= \begin{bmatrix} \tilde{\mathbf{b}}_1^T(t) \cdots \tilde{\mathbf{b}}_K^T(t) \end{bmatrix}^T, \\
\widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}(t)) &= \begin{bmatrix} \tilde{\mathbf{q}}_1^T(\widetilde{\mathbf{X}}(t)) \cdots \tilde{\mathbf{q}}_K^T(\widetilde{\mathbf{X}}(t)) \end{bmatrix}^T, \\
\widetilde{\mathbf{I}}(\widetilde{\mathbf{X}}(t)) &= \begin{bmatrix} \tilde{\mathbf{i}}_1^T(\widetilde{\mathbf{X}}(t)) \cdots \tilde{\mathbf{i}}_K^T(\widetilde{\mathbf{X}}(t)) \end{bmatrix}^T
\end{aligned} \qquad (9)$$

to obtain the NE equation

$$\frac{d}{dt}\widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}) + \Omega^T\widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}) + \widetilde{\mathbf{I}}(\widetilde{\mathbf{X}}) = \widetilde{\mathbf{B}}. \qquad (10)$$

The NE equation may be solved by any of the available methods for numerical solution of nonlinear differential equations (DEs). All DE solvers discretize the time axis in a set of time points $\{t_i\}$ and solve for the unknown samples $\widetilde{\mathbf{X}}(t_i)$ one at a time over the interval of interest; specific techniques differ in the method of approximation of the derivative operation, and circuit simulators traditionally employ the trapezoidal or the backward-difference approximation [15]. With no

loss of generality, this article makes use of the backward-difference formula, given as [15]

$$\frac{d}{dt}\widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}_i) \approx \alpha_0 \widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}_i) + \sum_{j=1}^{m} \alpha_j \widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}_{i-j}), \quad (11)$$

where α_j are known constants, m is the order of approximation, and $\widetilde{\mathbf{X}}_i$ is short for $\widetilde{\mathbf{X}}(t_i)$; for simplicity eq. (11) assumes a constant time-step approximation, but all of the following may be applied to adaptive time-stepping algorithms as well.

Upon substitution of the derivative approximation in the NE system, one obtains a set of non-linear algebraic equations, one for each envelope sample, of the form

$$\mathbf{F}(\widetilde{\mathbf{X}}_i) = \alpha_0 \widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}_i) + \Omega^T \widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}_i) + \widetilde{\mathbf{I}}(\widetilde{\mathbf{X}}_i)$$
$$+ \widetilde{\mathbf{A}}_i = \mathbf{0}, \quad i = 0, \ldots, M-1, \quad (12)$$

where $\widetilde{\mathbf{A}}_i$ is a vector of known quantities. At each step i, the solution is found by the Newton–Raphson iteration,

$$\mathbf{J_F}(\widetilde{\mathbf{X}}_i^{\text{old}})\widetilde{\mathbf{X}}_i^{\text{new}} = \mathbf{J_F}(\widetilde{\mathbf{X}}_i^{\text{old}})\widetilde{\mathbf{X}}_i^{\text{old}} - \mathbf{F}(\widetilde{\mathbf{X}}_i^{\text{old}}), \quad (13)$$

where, as usual, $\mathbf{J_F}$ is the Jacobian matrix of $\mathbf{F}$,

$$\mathbf{J_F} = \alpha_0 \mathbf{J}_{\widetilde{\mathbf{Q}}} + \Omega^T \mathbf{J}_{\widetilde{\mathbf{Q}}} + \mathbf{J}_{\widetilde{\mathbf{I}}}. \quad (14)$$

The evaluation of $\widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}_i)$ [and $\widetilde{\mathbf{I}}(\widetilde{\mathbf{X}}_i)$] proceeds in the following way. Recall that $\widetilde{\mathbf{X}}_i = [\widetilde{\mathbf{x}}_1^T(t_i), \ldots, \widetilde{\mathbf{x}}_K^T(t_i)]^T$, and recall the definition (3) of $\widetilde{\mathbf{x}}_k(t_i)$. By arguments that were used to arrive at eq. (8), $\widetilde{\mathbf{x}}_k(t_i)$ may be viewed as a set of Fourier coefficients that is related, by the inverse discrete fourier transformation (DFT), to a set of samples of $\mathbf{x}_k(t)$; the samples are taken at a set of $2N + 1$ time instants, spaced uniformly by $2\pi/(2N+1)\omega_c$ and denoted by ζ. This suggests that $\widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}_i)$ may be calculated in a way that is similar to the procedure, well known in the method of harmonic balance, for finding the phasor components of a time-domain nonlinear function, given the phasor components of its independent variables [16]. Let $\bar{\mathbf{x}} = [\bar{\mathbf{x}}_1^T, \ldots, \bar{\mathbf{x}}_K^T]^T$ be the vector of sampled nodal waveforms, where $\bar{\mathbf{x}}_k$ stores the samples of $\mathbf{x}_k(t)$ obtained from $\widetilde{\mathbf{x}}_k(t_i)$ by an application of the inverse DFT; for each k, the nonlinear relationship $\mathbf{q}_k(\mathbf{x}(t))$ is evaluated at time instants ζ to obtain its samples $\bar{\mathbf{q}}_k(\bar{\mathbf{x}})$; and the DFT is applied to $\bar{\mathbf{q}}_k(\bar{\mathbf{x}})$ to obtain $\widetilde{\mathbf{q}}_k(\widetilde{\mathbf{X}}_i)$. Formally,

$$\widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}_i) = \Gamma\bar{\mathbf{q}}(\bar{\mathbf{x}}) = \Gamma\bar{\mathbf{q}}(\Gamma^{-1}\widetilde{\mathbf{X}}_i), \quad (15)$$

and an analogous expression holds for $\widetilde{\mathbf{I}}(\widetilde{\mathbf{X}}_i)$; Γ is the matrix that transforms the nodal envelope components to samples of nodal waveforms—formally, a block-($K \times K$) diagonal matrix, each block being a $(2N+1) \times (2N+1)$ DFT matrix.

From eq. (15) one also obtains the formula for calculating $\mathbf{J}_{\widetilde{\mathbf{Q}}}(\widetilde{\mathbf{X}}_i)$:

$$\mathbf{J}_{\widetilde{\mathbf{Q}}}(\widetilde{\mathbf{X}}_i) = \Gamma\frac{\partial\bar{\mathbf{q}}}{\partial\mathbf{x}}\Gamma^{-1}, \quad (16)$$

and an analogous expression holds for $\mathbf{J}_{\widetilde{\mathbf{I}}}(\widetilde{\mathbf{X}}_i)$. The term $\partial\bar{\mathbf{q}}/\partial\mathbf{x}$ is a block-($K \times K$) matrix with diagonal blocks, where the (j, k)th diagonal block contains the samples of the (j, k)th element of $\partial\mathbf{q}/\partial\mathbf{x}$ evaluated at ζ.

Distributed circuit elements are included in the NE equation as follows. Each element contributes a term of the form

$$\int_{-\infty}^{\infty} \mathbf{h}^0(t - \tau)\mathbf{x}(\tau)\, d\tau \quad (17)$$

to the left-hand side of the governing circuit equation (1). Here, $\mathbf{h}^0(t)$ is the inverse Fourier transformation of $\mathbf{H}^0(\omega)$, the frequency-domain modified nodal description of the distributed element. With the assumed form (3) of $\mathbf{x}(t)$, the elements of eq. (17) may be expressed as

$$\sum_{i=1}^{K} \int_{-\infty}^{\infty} h_{k,i}^0(t - \tau)\left\{\sum_{n=-N}^{N} \tilde{x}_{i,n}(\tau)e^{jn\omega_c\tau}\right\} d\tau,$$
$$k = 1 \ldots K. \quad (18)$$

The impulse response of practical distributed elements decays to zero over an interval that is small compared to the inverse of the bandwidth of complex envelopes $\tilde{x}_{i,n}(t)$. Hence, a low-order Taylor series expansion of $\tilde{x}_{i,n}(\tau)$, about t, may be used to obtain an accurate approximation to the convolution integrals; the first-order expansion reduces eq. (18) to

$$\sum_{i=1}^{K}\left[\mathbf{H}_{k,i}\tilde{\mathbf{x}}_i(t) + \widehat{\mathbf{H}}_{k,i}\frac{d}{dt}\tilde{\mathbf{x}}_i(t)\right]^T \mathbf{w}(t),$$
$$k = 1, \ldots, K. \quad (19)$$

As before, $\tilde{\mathbf{x}}_i(t)$ is the $(2N+1)$-vector of real and imaginary components of the ith unknown variable and $\mathbf{H}_{k,i}$, $\widehat{\mathbf{H}}_{k,i}$ are tridiagonal matrices constructed from real and imaginary parts of the (k, i)th element of $\mathbf{H}^0(\omega)$ and $(d/d\omega)\mathbf{H}^0(\omega)$, respectively. Equation (19) may be written in

block-matrix form and incorporated in a simple way in the NE equation (10) by appending the contributions of distributed elements to $(d/dt)\widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}})$ and $\widetilde{\mathbf{I}}(\widetilde{\mathbf{X}})$.

A common application of the NE method—and the one that is of primary interest in this article—is the analysis with digitally modulated signals for an accurate calculation of adjacent-channel power and various measures of efficiency and gain. Now, digital modulation envelopes are random signals that must be approximated by finite-length waveforms for the purposes of NE analysis. To enhance the accuracy and repeatability of the simulations, two conditions on the input signal are imposed in this article: it should be (1) "long" and (2) periodic. A long sequence reduces the run-to-run variations of the results, 512 data pulses being commonly adopted as the minimum for acceptable repeatability; the condition of periodicity eliminates the potential loss of accuracy in the spectral analysis of simulation results.

Thus, the steady-state response of the NE equation to a periodic waveform of a long period is sought. Since the envelope period is long, and since practical amplifiers exhibit startup transients that last no more than a few envelope pulses, the following approach is used to find the steady-state response: if the envelope of the source signal is periodic on $[0, T]$, the circuit is solved over $[0, T + \Delta T]$, where ΔT is a predefined interval that is small in comparison to T. The solution $\widetilde{\mathbf{X}}(t)$ is assumed to have reached the steady state on $[\Delta T, T + \Delta T]$ and may be accurately Fourier analyzed on that interval. This approach, although heuristic, is inexpensive and adequate in practice.

Once the solution $\widetilde{\mathbf{X}}(t)$ is found, network functions are computed by simple postprocessing steps. The following presents a means of adjusting circuit parameters in a way that performance indicators match their specifications.

3. OPTIMIZATION

The formulation of the optimization problem is based on the principles established in ref. [17]. Let $\mathbf{p}$ denote the vector of P arbitrary variable parameters and let the desired performance be defined as a set of S_L lower bound and S_U upper bound specifications of the form

$$
\begin{aligned}
r_{l, j}(\mathbf{p}) &\geq s_{l, j}, \qquad j = 1, \ldots, S_L, \\
r_{u, j}(\mathbf{p}) &\leq s_{u, j}, \qquad j = 1, \ldots, S_U,
\end{aligned}
\tag{20}
$$

where $\mathbf{s}_l = [s_{l, j}]$ is the vector of lower bound specifications, $\mathbf{s}_u = [s_{u, j}]$ is the vector of upper bound specifications, and $\mathbf{r}_l(\mathbf{p})$, $\mathbf{r}_u(\mathbf{p})$ are the vectors of response functions that correspond to $\mathbf{s}_l$ and $\mathbf{s}_u$. Note that equality requirements $r_j(\mathbf{p}) = s_j$ may be imposed by simultaneous upper and lower bound specifications of the form (20).

To quantify the discrepancy between the desired and the actual response, vector $\mathbf{e}(\mathbf{p})$ of $S = S_L + S_U$ error elements is defined as

$$
\begin{aligned}
e_j(\mathbf{p}) &= s_{l, j} - r_{l, j}(\mathbf{p}), \quad j = 1, \ldots, S_L, \\
e_{S_L+j}(\mathbf{p}) &= r_{u, j} - s_{u, j}(\mathbf{p}), \quad j = 1, \ldots, S_U,
\end{aligned}
\tag{21}
$$

so that a nonpositive error indicates that the corresponding specification is satisfied.

Design optimization is based on searching for the minimum of an objective (cost) function that encompasses the goals defined by eq. (20). This article makes use of the generalized least pth objective function [18] given by

$$
E(\mathbf{p}) = \begin{cases} \left[\sum_{j \in J} |e_j|^p\right]^{1/p}, & J \text{ nonempty}, \\ -\left[\sum_{j=1}^{S} (-e_j)^{-p}\right]^{-1/p}, & \text{otherwise}, \end{cases}
\tag{22}
$$

where J denotes the index set $J = \{j | e_j \geq 0\}$; p is a parameter that allows for design flexibility [17, 18] and should not be confused with the notation for variable parameters.

An appealing feature of the generalized least pth objective is that it allows the optimization to proceed once the specifications are satisfied. This property is useful in the design of power amplifiers where it is common to attempt to exceed the linearity and efficiency requirements by as much as possible.

As mentioned in the introduction, gradient optimization is very effective in finding the set of variable parameters that minimize cost functions such as eq. (22). There are several mature gradient optimization algorithms, but central to their effectiveness is efficient and accurate computation of the parameter-space gradient

$$
\nabla_{\mathbf{p}} E(\mathbf{p}) = \left[\frac{\partial E}{\partial p_1} \cdots \frac{\partial E}{\partial p_P}\right]^T.
\tag{23}
$$

The next section deals with methods for efficient gradient computation, designed specifically for the NE method.

4. GRADIENT CALCULATION

4.1. Preliminaries

Assume that all performance indicators are expressible as functions of the vector $\mathbf{P}_\Omega = [P_{\Omega 1} \cdots P_{\Omega Q}]^T$, where a generic P_Ω represents the power dissipated in a branch of the circuit over some frequency band Ω. The following methods are not limited to such performance indicators, but it is readily verified that those of present interest—such as various gain, efficiency, and distortion definitions—may all be expressed in such form.

Based on signal definition (3), the power dissipated in any circuit branch, over a frequency band Ω located near the mth multiple of the carrier frequency, is given as

$$P_\Omega = 2\,\mathrm{Re}\Big\{ \sum_{n:\ m\omega_c + n\omega_0 \in \Omega} \widetilde{V}_{m,n}\widetilde{I}^*_{m,n} \Big\}, \quad (24)$$

where $\widetilde{V}_{m,n}$ and $\widetilde{I}_{m,n}$ are the complex Fourier coefficients of the mth envelope component of the branch voltage and current and $\omega_0 = 2\pi/MT$ is the envelope frequency resolution. In the sequel, the carrier index is dropped for simplicity.

By the chain rule of differentiation, the cost function gradient becomes

$$\nabla_{\mathbf{p}} E = \mathbf{J}^T_{\mathbf{P}_\Omega}(\mathbf{p})\mathbf{J}^T_{\mathbf{e}}(\mathbf{P}_\Omega)\nabla_{\mathbf{e}} E. \quad (25)$$

The last two factors on the right-hand side are known from explicit expressions for the cost function and the performance indicators. What remains to be found, therefore, are the elements of the Jacobian $\mathbf{J}_{\mathbf{P}_\Omega}(\mathbf{p})$,

$$\frac{\partial P_\Omega}{\partial p}, \quad (26)$$

for all $P_\Omega \in \mathbf{P}_\Omega$ and $p \in \mathbf{p}$. From eq. (24), one immediately obtains

$$\frac{\partial P_\Omega}{\partial p} = 2\,\mathrm{Re} \sum_{n:\ m\omega_c + n\omega_0 \in \Omega} \left[\frac{\partial \widetilde{V}_n}{\partial p}\widetilde{I}^*_n + \frac{\partial \widetilde{I}_n}{\partial p}\widetilde{V}^*_n \right], \quad (27)$$

where

$$\frac{\partial \widetilde{V}_n}{\partial p} = \frac{1}{T}\int_{\Delta T}^{\Delta T + T} \frac{\partial \widetilde{v}}{\partial p}e^{-jn\omega_0 t}\,dt, \quad (28)$$

and a similar expression holds for the current derivative. Note, $\partial\widetilde{V}_n/\partial p$ is just the nth Fourier coefficient of the partial derivative $\partial\widetilde{v}/\partial p$. Thus, the problem of derivative calculation has been reduced to finding $\partial\widetilde{v}/\partial p$ and $\partial\widetilde{i}/\partial p$, the quantities that are in this article referred to as *envelope sensitivities*.

4.2. Envelope Sensitivity Equation

In this section, an equation for parameter sensitivities of the nodal envelope waveforms

$$\widetilde{\mathbf{Z}}_p(t) = \frac{\partial \widetilde{\mathbf{X}}(t)}{\partial p} \quad (29)$$

is derived. Any voltage or current that is needed in formulating the elements of the power vector (24) follows from $\widetilde{\mathbf{Z}}_p(t)$, either directly or as a linear combination of its elements.

To derive the envelope sensitivity equation, the NE system is differentiated at $\widetilde{\mathbf{X}}(t)$ with respect to p, keeping in mind that $\widetilde{\mathbf{X}}(t) \equiv \widetilde{\mathbf{X}}(\mathbf{p}, t)$, $\widetilde{\mathbf{Q}}(\widetilde{\mathbf{X}}(t)) \equiv \widetilde{\mathbf{Q}}(\mathbf{p}, \widetilde{\mathbf{X}}(t))$ and that similar dependencies on the parameter vector hold for vectors $\widetilde{\mathbf{I}}$ and $\widetilde{\mathbf{B}}$. Thus,

$$\frac{d}{dt}\{\mathbf{J}_{\widetilde{\mathbf{Q}}}(\widetilde{\mathbf{X}})\widetilde{\mathbf{Z}}_p\} + \Omega^T \mathbf{J}_{\widetilde{\mathbf{Q}}}(\widetilde{\mathbf{X}})\widetilde{\mathbf{Z}}_p$$
$$+ \mathbf{J}_{\widetilde{\mathbf{I}}}(\widetilde{\mathbf{X}})\widetilde{\mathbf{Z}}_p = \widetilde{\mathbf{C}}(p), \quad (30)$$

where $\widetilde{\mathbf{C}}(p)$ is the vector of known elements,

$$\widetilde{\mathbf{C}}(p) = \frac{\partial}{\partial p}\{\widetilde{\mathbf{B}} - \widetilde{\mathbf{I}} - \Omega^T\widetilde{\mathbf{Q}}\} - \frac{d}{dt}\frac{\partial \widetilde{\mathbf{Q}}}{\partial p}; \quad (31)$$

observe that $\widetilde{\mathbf{C}}(p)$ may be computed explicitly, using rules that are very similar to those used to formulate the modified nodal circuit equations.

The envelope sensitivity equation is a linear differential equation that may be solved in a manner similar to numerical integration of the NE equation. Thus, by applying the backward-difference formula to eq. (30), one obtains a set of M independent linear algebraic systems

$$\alpha_0\mathbf{J}_{\widetilde{\mathbf{Q}}}(\widetilde{\mathbf{X}}_i)\widetilde{\mathbf{Z}}_{pi} + \Omega^T \mathbf{J}_{\widetilde{\mathbf{Q}}}(\widetilde{\mathbf{X}}_i)\widetilde{\mathbf{Z}}_{pi}$$
$$+ \mathbf{J}_{\widetilde{\mathbf{I}}}(\widetilde{\mathbf{X}}_i)\widetilde{\mathbf{Z}}_{pi} = \widetilde{\mathbf{D}}_i(p), \quad (32)$$

where $i = 0, \ldots, M - 1$ and the known quantities have been collected in $\widetilde{\mathbf{D}}_i(p)$. By eq. (14), eq. (32) may be expressed as

$$\mathbf{J}_{\mathbf{F}}(\widetilde{\mathbf{X}}_i)\widetilde{\mathbf{Z}}_{pi} = \widetilde{\mathbf{D}}_i(p). \quad (33)$$

The envelope sensitivity waveforms are assumed to reach the periodic steady state on $[\Delta T, \Delta T + T]$, as was the case with the NE waveforms in eq. (10).

Note that the computation of envelope sensitivities involves solutions, one for each parameter, of linear systems similar to those that arise in the solution of the Newton–Raphson iteration (13).

On the other hand, the size of problem (13) is the deciding factor in the approach to its solution; by "problem size," one usually refers to the product of the number of analysis frequencies, $2N + 1$, and the number of nonlinear elements in the circuit, because these quantities influence the number and the size of dense blocks in the Jacobian $\mathbf{J_F}$ [16]. For small problems, the solution of eq. (13) is handled efficiently by Gaussian elimination (direct) methods; one resorts to iterative means when the problem size renders direct methods ineffective. In either case, the solution method for the NE system has an important impact on the solution of the sensitivity equation, as discussed in the following.

4.3. Numerical Solution of the Sensitivity Equation

If problem size permits an efficient solution of the Newton–Raphson iteration by direct methods, it is here referred to as a *small-scale* problem. Given that eq. (13) is solved by direct means, triangular factors of $\mathbf{J_F}(\widetilde{\mathbf{X}}_i^{\mathrm{old}})$ are available upon convergence; $\widetilde{\mathbf{X}}_i^{\mathrm{old}}$ is the next-to-last guess at the solution $\widetilde{\mathbf{X}}_i$. If one assumes that $\mathbf{J_F}(\widetilde{\mathbf{X}}_i)$ is well approximated by $\mathbf{J_F}(\widetilde{\mathbf{X}}_i^{\mathrm{old}})$, the triangular factors may be reused in the calculation of eq. (33). Hence, the cost of sensitivity calculation is one back substitution and one forward elimination per variable parameter. The substitution and elimination steps are much more efficient than the process of factoring the Jacobian, so the overhead of gradient calculation tends to be small.

The assumption that $\mathbf{J_F}(\widetilde{\mathbf{X}}_i^{\mathrm{old}})$ represents a good approximation to the Jacobian $\mathbf{J_F}(\widetilde{\mathbf{X}}_i)$ is an important one, but the authors' experience is that it does not appear as an obstacle in accurate computation of cost function gradients. It is, of course, possible to factor the Jacobian at the computed solution and to reuse its factors in the calculation of sensitivities; this comes at an additional expense, roughly equal to the reciprocal of the average number of Newton–Raphson iterations in the course of the NE simulation—a number in the neighborhood of 20% in common applications. However, in all of the authors' experiments, one of which will be described in Section 5, the first approach worked just as well, with gradients that stood in excellent agreement to those that were computed by the finite-difference procedure.

A *large-scale* problem is one that necessitates the application of iterative methods to the solution of the Newton–Raphson iteration (13).

Iterative solution methods, most notably the generalized minimal residual (GMRES) method [4], have been found very effective in overcoming the deficiencies of direct solvers in applications to large problems [5–9]. Large problems result in large and dense (more precisely, block-sparse with many large and dense blocks) Jacobian matrices [16]; since direct solvers require explicit computation and storage of the Jacobian (14), and since the factoring of large and dense matrices is an expensive computational task, their effectiveness diminishes with the growth of problem size. Although iterative methods were introduced as a means to handle extremely large simulation tasks [5, 6], their advantages over direct methods have been observed even in the case of problems of moderate size [7, 8].

The "secret" of the success of iterative solution methods, such as GMRES, is that the only required manipulation of the system matrix—in this case the Jacobian—is the matrix-vector product, and the particular structure of the Jacobian allows an efficient exploitation of this fact. Recall,

$$\mathbf{J_F} = \alpha_0 \Gamma \frac{\partial \bar{\mathbf{q}}}{\partial \mathbf{x}} \Gamma^{-1} + \Omega^T \Gamma \frac{\partial \bar{\mathbf{q}}}{\partial \mathbf{x}} \Gamma^{-1} + \Gamma \frac{\partial \bar{\mathbf{i}}}{\partial \mathbf{x}} \Gamma^{-1}, \quad (34)$$

where $\partial \bar{\mathbf{q}}/\partial \mathbf{x}$ and $\partial \bar{\mathbf{i}}/\partial \mathbf{x}$ are block-sparse matrices with diagonal blocks (in the presence of distributed elements, the first and the third terms on the right-hand side will contain some tridiagonal blocks, too, as discussed in Section 2); thus, vector multiplication by $\mathbf{J_F}$ requires DFTs and sparse multiplications only. The storage of $\partial \bar{\mathbf{q}}/\partial \mathbf{x}$ and $\partial \bar{\mathbf{i}}/\partial \mathbf{x}$ is inexpensive relative to explicit storage of $\mathbf{J_F}$, because of the absence of dense and large blocks.

Because of the similar structure of eqs. (13) and (33), GMRES is directly applicable to gradient computation in large-scale problems. However, recall that small-scale methods featured one highly desirable property—once the NE iteration converged, the computation of sensitivities was a matter of triangular eliminations and substitutions, a relatively inexpensive computational task. This, in effect, sharing of information between the NE solution and the sensitivity solution allowed for efficient gradient computation.

Unfortunately, the GMRES solution of the NE system does not provide information that can be utilized for the solution of envelope sensitivities—one NE iteration costs as much as the solution of one sensitivity system. The number of variable parameters, in contrast to small-scale

methods, strongly influences the efficiency of large-scale gradient computation. A straightforward application of GMRES to the solution of eq. (33)—by repeating the procedure for each variable parameter—is more efficient than finite-difference computation, with the speed-up ratio roughly equal to the average number of Newton–Raphson iterations in the course of the NE simulation; however, one is clearly motivated to seek a better approach.

The authors achieved a noticeable reduction in the computational overhead, as compared to repeated application of GMRES to eq. (33)—by means of the MHGMRES algorithm [10], a GMRES-like method suitable for multiple right-hand sides. The method reduces the computational effort by a simultaneous consideration of all the right-hand sides and a type of "information sharing" that will be apparent in the following. In order to introduce the algorithm, some background on GMRES is needed [4].

The GMRES method belongs to the class of *Krylov subspace* solvers, meaning that the solution of the general linear system $\mathbf{Ax} = \mathbf{b}$ is sought in the form

$$\mathbf{x} \in \mathbf{x}_0 + K_m(\mathbf{A}, \mathbf{r}_0), \qquad (35)$$

where $\mathbf{x}_0$ is an initial guess, $\mathbf{r}_0 = \mathbf{b} - \mathbf{Ax}_0$ is the corresponding residual, and $K_m(\mathbf{A}, \mathbf{r}_0)$ is the vector space defined by $\mathrm{span}\{\mathbf{r}_0, \mathbf{Ar}_0, \ldots, \mathbf{A}^{m-1}\mathbf{r}_0\}$ and known as the Krylov subspace.

In GMRES, one constructs an orthonormal basis for the Krylov subspace, so that the solution may be expressed as

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{V}_m \mathbf{y}, \qquad (36)$$

where the columns of $\mathbf{V}_m$ are the Krylov basis vectors and $\mathbf{y}$ is the vector of expansion coefficients that is selected subject to minimizing the 2-norm of the residual

$$\mathbf{r} = \mathbf{b} - \mathbf{Ax}. \qquad (37)$$

An orthonormal basis for the Krylov subspace is constructed by means of the Arnoldi procedure [4], which produces basis vectors that satisfy the recursion $\mathbf{AV}_m = \mathbf{V}_{m+1}\mathbf{H}$, where $\mathbf{H}$ is an $(m+1) \times m$ matrix. From eqs. (36) and (37), one may derive the following expression for the residual norm:

$$\|\mathbf{r}\| = \|\mathbf{V}_{m+1}^T \mathbf{r}_0 - \mathbf{Hy}\|. \qquad (38)$$

Vector $\mathbf{y}$, which minimizes, eq. (38), is the least-squares solution of the rectangular system $\mathbf{Hy} = \mathbf{V}_{m+1}^T \mathbf{r}_0$.

By definition of the Krylov subspace, the GMRES approximation (35) may be written as $\mathbf{x}_0 + q_{m-1}(\mathbf{A})\mathbf{r}_0$, where $q_{m-1}(z)$ is a polynomial of degree $m - 1$. Likewise, the residual (37) may be written as $\mathbf{r} = p_m(\mathbf{A})\mathbf{r}_0$, where $p_m(z) = 1 - zq_{m-1}(z)$. The polynomial $p_m(z)$ is referred to as the GMRES polynomial.

The m-dimensional Krylov approximation is accepted if the norm of the residual, $\|\mathbf{r}\|$, is smaller than a prescribed tolerance. Larger m lead to better approximations at the expense of computational complexity and larger memory requirements. Thus, m is usually fixed, and in the case that the current approximation is not acceptable, the GMRES algorithm is *restarted* with the current approximation as the initial solution.

Besides restarting, *preconditioning* is often used to improve the convergence of GMRES and is crucial to its effectiveness in circuit analysis [5–8]. The preconditioned system is of the form $\widehat{\mathbf{A}}^{-1}\mathbf{Ax} = \widehat{\mathbf{A}}^{-1}\mathbf{b}$, where $\widehat{\mathbf{A}}$, the preconditioner, approximates the system matrix in some sense. A successful preconditioner is not only one that aids convergence but also one that is inexpensive to store and factor.

With these preliminaries, one can easily sketch the operation of the MHGMRES algorithm. Suppose one is to solve the general block linear system $\mathbf{AX} = \mathbf{B}$, where the columns of $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_P]$ are the P right-hand sides and the columns of $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_P]$ are their solutions. Let $\mathbf{R} = \mathbf{B} - \mathbf{AX}$ be the residual associated with $\mathbf{X}$, where $\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_P]$. In applications to the problem at hand, the columns of $\mathbf{B}$ are the vectors $\widetilde{\mathbf{D}}_i(p)$, and the columns of $\mathbf{X}$ are the envelope sensitivity vectors $\widetilde{\mathbf{Z}}_{pi}$, for all $p \in \mathbf{p}$.

Starting from an initial guess $\mathbf{X}_0$, MHGMRES selects the residual $\mathbf{r}_{0,\,\sigma}$, called the *seed*, with the largest initial norm. The seed solution is found in the regular GMRES fashion, by performing the Arnoldi procedure to find an orthonormal basis $\mathbf{V}_{m,\,\sigma}$ for the Krylov subspace $K_m(\mathbf{A}, \mathbf{r}_{0,\,\sigma})$. The seed solution is updated, as usual, according to eq. (36).

In the next phase of the algorithm, one computes the least-squares solutions to $\mathbf{H}_\sigma \mathbf{y}_i = \mathbf{V}_{m+1,\,\sigma}^T \mathbf{r}_{0,\,i}$ for all $i \neq \sigma$. Vectors $\mathbf{y}_i$ are used to approximate the nonseed solutions according to eq. (36), and the nonseed residuals are computed in the usual fashion.

In the final phase of the algorithm, the GMRES polynomial of the seed, $p_{m,\,\sigma}(z)$, is applied to the

current residuals in the form of Richardson's iteration,

$$\text{for } i = 1, m$$
$$\mathbf{X} = \mathbf{X} + \lambda_i^{-1}\mathbf{R}$$
$$\mathbf{R} = \mathbf{B} - \mathbf{A}\mathbf{X}$$
$$\text{end,}$$

where λ_i are the roots of the seed polynomial, constructed as in ref. [10]. The residuals are checked for convergence, a new seed is selected, and the procedure is restarted with the current solution estimate as the initial guess.

The MHGMRES algorithm features a number of advantages over repeated GMRES. For each restart, it eliminates $P - 1$ Arnoldi procedures that would be called by repeated application of GMRES. It is memory conserving, in the sense that Krylov subspace vectors are only generated for the seed solution. Although the Richardson iteration is not guaranteed to improve the quality of approximation, practical safeguards can be implemented [19] that help prevent unsatisfactory behavior; in all of the authors' experiments it behaved favorably, although no precautions were taken to ensure its convergence. Finally, MHGMRES defaults to GMRES in the case of a single right-hand side.

4.4. Numerical Comparison

The performances of the three methods for the numerical solution of the sensitivity system—Gaussian elimination (direct), repeated GMRES, and MHGMRES—are now compared.

The comparison is illustrated on the circuit shown in Figure 1, a power amplifier whose performance is optimized in Section 5. The analysis was performed at the initial design point and under normal operating conditions to be explained in the next section.

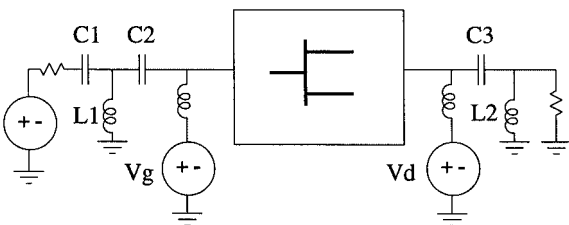The performance of each method was measured by the numerical overhead of gradient calculation, defined as the average value of the ratio

$$\frac{T_S}{T_{\text{NE}}}, \tag{39}$$

where T_{NE} is the time needed to solve for an envelope sample and T_S is the time needed to calculate the sensitivity of that sample with respect to all the variable parameters; the averaging was performed over the entire solution interval.

To conduct the comparison, the sensitivity vectors $\widetilde{\mathbf{Z}}_p(t)$ were computed with the number of parameters varying between 1 and 10. The parameters were arbitrarily chosen from the various sources and linear circuit elements. The gradient of the cost function, also to be described, was computed by each of the three methods, and the results stood in excellent agreement (no less than three significant digits) with each other and with the finite-difference method. The overhead ratios are plotted in Figure 2.

As expected, the direct solver is the least expensive, with the moderate overhead of 22% with as many as ten variable parameters. Among the large-scale methods, MHGMRES clearly outperforms the repeated GMRES method, being nearly twice as efficient when the number of parameters reaches 10.

The comparison between the two large-scale methods was performed under identical conditions—all the solutions were approximated to a relative tolerance of $10^{-10}\|\mathbf{b}\|$, where $\|\mathbf{b}\|$ is the norm of the corresponding right-hand side, and with the maximum Krylov subspace dimension of 15. The preconditioner is the block-diagonal
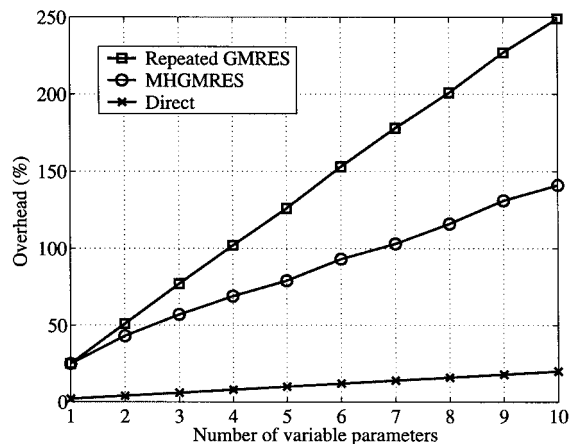

Figure 1. Amplifier circuit.

Figure 2. Performance of repeated GMRES, MHGMRES, and direct methods in the solution of the envelope sensitivity equation

matrix obtained by evaluating the Jacobian at the DC bias point. The algorithms were tested on circuits of different size (as few as 400 variables and as many as 2000) and at different driving conditions, so far with the same general result—a useful improvement in the speed of gradient calculation by means of MHGMRES.

5. EXAMPLE

The concepts of Sections 2–4 were applied to design optimization of the circuit in Figure 1. The aim was to achieve an output power of at least 24 dBm, a drain efficiency of at least 50%, and an upper adjacent channel power ratio (ACPR) of at most -35 dB. The circuit operates at 5 GHz, with a fixed 5-V supply, and with an available source power of 10 dBm. The input is a raised-cosine filtered $\pi/4$-DPQSK-modulated carrier (differential phase shift keying), with a data rate of 10^5 symbols per second. The field effect transistor (FET) is described by the Curtice-cubic nonlinear model that includes terminal parasitics.

The initial design was performed by single-tone load pull simulations to find a gate bias and load conditions that may achieve the specifications. The input matching circuit was designed to match the source resistance at the design frequency and at the nominal gate bias. The initial design provided 25.4 dBm of single-tone output power and a drain efficiency of 53%.

Next, a modulated signal simulation was performed according to principles described in Section 2. The circuit consisted of 23 nodal variables and the analysis included 10 carrier harmonics (including the DC component); this results, at each time step (12), in a nonlinear system of dimension 437. As in most cases where a circuit is operated at the knee of the gain compression curve, the output power of a (varying-envelope) modulated input is lower than that of a single-tone input with equal available power. In this case, the output power decreased to 24.9 dBm, with a corresponding decrease in efficiency to 49.5%.

The circuit was next optimized using the methods of Section 3 and 4 to increase the output power and efficiency to the specified levels; the distortion specification was expected to be unattainable under the operating conditions and it was, for the time being, left out of consideration.

Letting $r_{l,1}$ denote the power measurement and $r_{l,2}$ denote the efficiency measurement, the error

vector was defined as

$$e_1 = 24 - r_{l,1} \qquad e_2 = 50 - r_{l,2}, \qquad (40)$$

and the cost function was assembled according to eq. (22), with $p = 2$. The optimization variables were the load matching elements C_3 and L_2 and the bias level V_g. The optimization algorithm was a quasi-Newton gradient method from the MATLAB suite of optimization utilities. The optimized performance and the corresponding parameter values are shown in Table I. Clearly, the power and efficiency specifications were satisfied, but the ACPR was well below desired. In order to improve the linearity performance, the amplifier was placed inside a feedforward loop, as shown in Figure 3. The loop operates as follows. The distorted signal at the output of the amplifier may be thought of as a linearly scaled version of the input plus a distortion component. The undistorted input and the distorted output are sampled by the first two couplers, and their scaled and phase-shifted versions are added to produce, at the output of the combiner, a scaled version of the distortion component only. The scaled distortion is amplified, phase shifted, and added to the output signal, through the last coupler, to reduce the output distortion. The variable phase shifter, attenuator, and small-signal amplifier are adjusted to provide amplitudes and phases that result in proper signal cancellations.

TABLE I. Amplifier Performance before and after Optimization

Performance indicator	Before optimization	After optimization
Output Power (dBm)	24.9	25.0
Efficiency (%)	49.5	56.0
ACPR (dB)	-27.0	-24.6

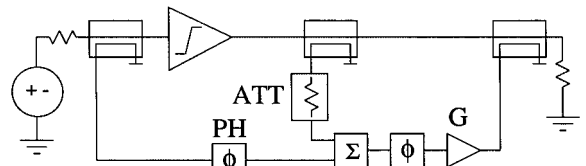| Variable parameter	Before optimization	After optimization
C_3 (pF)	1.27	1.84
L_2 (nH)	1.60	2.28
V_g (V)	-0.80	-0.84

Figure 3.   Feedforward linearization network

TABLE II. Amplifier Performance before Linearization and after Optimization of the Feedforward Loop.

Performance indicator	Before linearization	After loop optimization
Output power (dBm)	25.0	24.1
Efficiency (%)	56.0	51.5
ACPR (dB)	−24.6	−37.0

Variable parameter	Before linearization	After loop optimization
C_2 (pF)	0.45	0.56
L_1 (nH)	2.6	2.5
PH (deg.)	2.8	−5.7
ATT (dB)	−13.2	−12.4
G (dB)	46	46

The improved linearity comes at the expense of some output power loss: unavoidably, small amounts of power flow through the coupled ports, and some power is lost in the direct path of the three couplers. On the other hand, return losses are, in principle, controllable. Although the initial design provided low single-tone return loss, the change in the load conditions and the gate bias level had an adverse effect on power reflection at the gate port. For this reason, in addition to varying the adjustable parameters PH, ATT, and G of the feedforward loop, the components L_1 and C_2 of the input matching circuit were varied too.

The optimization was carried out as before, this time by adding a third specification as

$$e_3 = r_{u,1} + 35, \qquad (41)$$

where $r_{u,1}$ denotes the upper ACPR measurement. The initial values of the loop elements were selected by inspection of the amplifier's single-tone distortion curves and the scattering characteristics of the various distributed components. The performance of the amplifier and a plot of the output spectra, before linearization and after optimization of the feedforward loop, are shown in Table II and Figures 4 and 5.

6. CONCLUSION

A method for optimization of RF amplifiers for digital communications has been proposed. The method is rigorous and efficient, based on the following features: (1) performance indicators are calculated by accurate NE simulations, (2) it employs gradient optimization, and (3) gradients
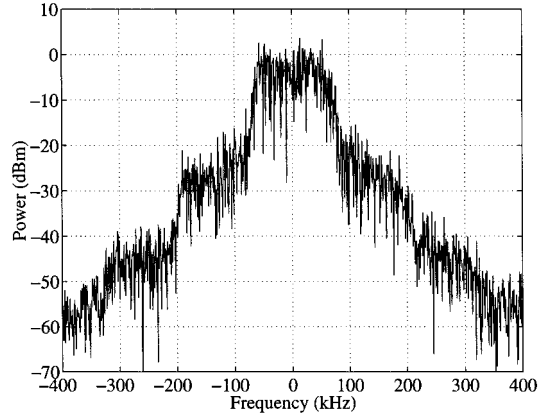
Figure 4. Output spectrum—initial design.
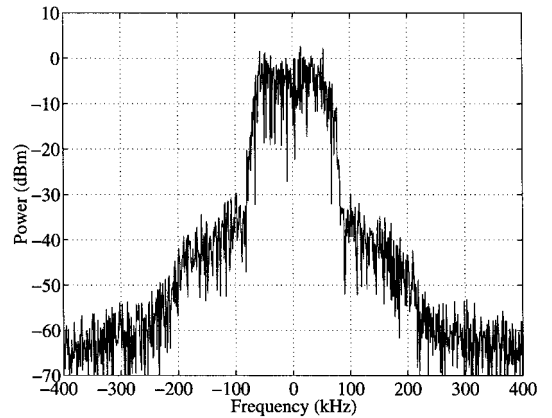
Figure 5. Output spectrum—optimized design.

are calculated by efficient means, starting from sensitivity analysis of envelope waveforms.

The method finds use in many common design tasks—adjustment of bias conditions for maximum power and efficiency, design of linearizing networks, and the design of out-of-band load conditions for optimal nonlinear performance. It is easily extendible to other forced nonlinear circuits, such as RF and microwave mixers. Finally, it is of immediate interest in transistor modeling to improve model accuracy in digitally modulated signal simulations.

ACKNOWLEDGMENTS

REFERENCES

1. P. Feldmann and J. Roychowdhury, Computation of circuit waveform envelopes using an efficient, matrix-decomposed harmonic balance algorithm, Proc Int Conf CAD, 1996, San Jose, CA, pp. 295–300.

2. E. Ngoya and R. Larcheveque, Envelope transient analysis: a new method for the transient and steady state analysis of microwave communication circuits and systems, Proc Int Microwave Symp, San Francisco, CA, 1996, pp. 1365–1368.

3. V. Rizzoli, A. Neri, F. Mastri, and A. Lipparini, A modulation-oriented piecewise harmonic-balance technique suitable for transient analysis and digitally modulated signals, Proc European Microwave Conf, Prague, 1996, pp. 546–549.

4. Y. Saad, Iterative methods for sparse linear systems, PWS, Boston, 1996.

5. R. Melville, P. Feldmann and J. Roychowdhury, Efficient multi-tone distortion analysis of analog integrated circuits, Proc Custom Integrated Circuits Conf, Santa Clara, CA, 1995, pp. 241–244.

6. V. Rizzoli, F. Mastri, C. Cecchetti, and F. Sgallari, Fast and robust inexact Newton approach to the harmonic-balance analysis of nonlinear microwave circuits, IEEE Microwave Guided Wave Lett (Oct. 1997), 359–361.

7. R. Telichevesky, K. Kundert, I. Elfadel, and J. White, Fast simulation algorithms for RF circuits, Proc Custom Integrated Circuits Conf, San Diego, CA, 1996, pp. 437–444.

8. M. Gourary, S. Rusakov, S. Ulyanov, M. Zharov, K. Gullapalli, and B. Mulvaney, Iterative solution of linear systems in harmonic balance analysis, Proc Int Microwave Symp, Denver, CO, 1997, pp. 1507–1510.

9. V. Rizzoli, A. Neri, F. Mastri, and A. Lipparini, Modulation-oriented harmonic balance based on Krylov-subspace methods, Proc Int Microwave Symp, Anaheim, CA, 1999, pp. 771–774.

10. V. Simoncini and E. Gallopoulos, An iterative method for nonsymmetric systems with multiple right-hand sides, SIAM J Sci Comput (July 1995), 917–933.

11. V. Rizzoli and A. Neri, State of the art and present trends in nonlinear microwave CAD techniques, IEEE Trans MTT (Feb. 1988), 343–365.

12. J. Bandler, Q.-J. Zhang, and R. Biernacki, A unified theory for frequency-domain simulation and sensitivity analysis of linear and nonlinear circuits, IEEE Trans MTT (Dec. 1988), 1661–1669.

13. V. Borich, J. East, and G. Haddad, Computer-aided optimization of adjacent channel power in nonlinear communications amplifiers, Proc Int Microwave Symp, Anaheim, CA, 2000.

14. J. Sevic, M. Steer, and A. Pavio, Nonlinear analysis methods for the simulation of digital wireless communication systems, Int J Microwave Millimeter Wave CAE (May 1996), 197–216.

15. J. Vlach and K. Singhal, Computer methods for circuit analysis and design, Van Nostrand Reinhold, New York, 1994.

16. K. Kundert and A. Sangiovanni-Vincentelli, Simulation of nonlinear circuits in the frequency domain, IEEE Trans CAD (Oct. 1986), 521–535.

17. J. Bandler and S. Chen, Circuit optimization: the state of the art, IEEE Trans MTT (1988), 424–443.

18. J. Bandler and C. Charalambous, Practical least pth optimization of networks, IEEE Trans MTT (Dec. 1972), 834–840.

19. N. Nachtigal, L. Reichel, and L. Trefethen, A hybrid GMRES algorithm for nonsymmetric matrix iterations, SIAM J Matrix Anal Appl (July 1992), 796–825.

BIOGRAPHIES

Vuk Borich was born in Serbia in 1973. He earned the B.S.E.E degree (magna cum laude) from the General Motors Institute in 1996. During spring and fall seasons of 1992–1996 he worked at Magnetek, Inc. and Krause Systems as an engineer-in-training, where his duties included industrial control and network programming. He joined the University of Michigan as a research assistant in 1996 and obtained the M.S.E.E. degree in 1997. He is now a Ph.D. candidate at the University of Michigan involved with studies of nonlinear distortion in microwave subsystems.

Jack East received his B.S.E., M.S. and Ph.D. degrees from the University of Michigan. He is presently at the Solid State Electronics Laboratory at the University of Michigan. His research interests include microwave and millimeter-wave device design, fabrication and measurements, and the analysis and design of communications systems.
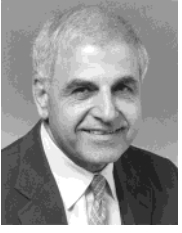
George I. Haddad received the B.S.E., M.S.E., degrees in electrical engineering from the University of Michigan. In 1958 he joined the Electron Physics Laboratory, where he was engaged in research on masers, parametric amplifiers, detectors, and electron-beam devices. From 1960 to 1969 he served successively as instructor, assistant professor, associate professor, and professor in the Electrical Engineering Department. He served as director of the Electron Physics Laboratory from 1968 to 1975. From 1975 to 1986 and 1991 to 1997 he served as chairman of the Department of Electrical Engineering and Computer Science. From 1987 to 1990 he was the director of both the Solid-State Electronics Laboratory and the Center for High-Frequency Microelectronics. He is currently the Robert J. Hiller Professor of Electrical Engineering and Computer Science and the director of the Center for High Frequency Microelectronics. His current research areas are microwave and millimeter-wave solid-state devices and monolithic integrated circuits, microwave-optical interactions and optoelectronic devices and integrated circuits. Dr. Haddad received the 1970 Curtis W. McGraw Research Award of the American Society for Engineering Education for outstanding achievements by an engineering teacher, the College of Engineering Excellence in Research Award (1985), the Distinguished Faculty Achievement Award (1986) of the University of Michigan, and the S. S. Attwood Award of the College of Engineering for Outstanding Contributions to Engineering Education, Research and Administration. He is the recipient of the 1996 IEEE-MTT Distinguished Educator Award and the IEEE Third Millennium Medal. He is a member of Eta Kappa Nu, Sigma Xi, Phi Kappa Phi, Tau Beta Pi, the American Society for Engineering Education, and the American Physical Society. He is a Fellow of the IEEE and a member of the National Academy of Engineering.