Bayesian Optimization for Broadband High-Efficiency Power Amplifier Designs

Peng Chen, *Student Member, IEEE*, Brian M. Merrick, *Member, IEEE*, and Thomas J. Brazil, *Fellow, IEEE*

***Abstract*—This paper proposes a novel, optimization-oriented strategy for the design of broadband, high-efficiency power amplifiers (PAs) using Bayesian optimization (BO). The optimization algorithm optimizes the drain waveforms by maximizing the fundamental output power while minimizing the harmonic and dissipated components. The optimization process is automated using simulation software. Circuit-based BO and electromagnetic-based (EM-based) BO are applied to design 10 W and 30 W PAs. The 10 W PA designed using circuit-based BO achieves a drain efficiency higher than 60% with output power greater than 39.8 dBm from 1.5 GHz to 2.5 GHz, while the 30 W PA designed using EM-based BO offers a drain efficiency higher than 57% with output power greater than 43.8 dBm across the band. Upon comparison of results, it is revealed that the proposed strategy outperforms a commercial electronic design automation (EDA) software's built-in optimizer, thus demonstrating that the EM-based BO is well-suited to the challenge of high power designs.**

***Index Terms*—Bayesian optimization, broadband, drain waveforms, Gaussian process, high efficiency, power amplifier.**

I. Introduction

Modern wireless communication systems, which have high transmission rates and spectral efficiency, are creating a growing demand for broadband high-efficiency power amplifiers (PAs). A series of high-efficiency PA modes, such as Class-F and Class-J, have been developed over the past few decades [1]-[5]. These new PA modes have received widespread interest since they allow designers to analyze PA performance from the point of view of waveform engineering [6]-[7]. The non-overlapping of the drain voltage and drain current waveforms has been recognized as a key criterion in achieving high efficiency in PA designs. Based on this concept, continuous modes have been studied in order to give greater flexibility to the drain waveforms for wideband designs [8]-[12]. In practical PA designs, the desired waveforms are often achieved through tuning the harmonic impedances presented by the matching networks. In order to realize a broadband design, the matching networks require multiple sections of transmission lines, generally greater than 10 in total. This tuning work places a great demand on design experience, especially when the efficiency needs to remain at a high level across the band. The broadband matching problem can be regarded as a high-dimensional optimization problem. In view of this, it is helpful to apply an optimization-oriented technique to tackle this problem.

Although commercial EDA software packages, such as ADS, AWR, etc., are quite useful in microwave circuit optimization, their built-in optimizers may fail to find a satisfactory solution when the dimension of variables is very high or the optimization goals are stringent. To overcome this limitation, it is necessary to develop modifiable, intelligent algorithms tailored to specific optimization tasks. To date, a number of optimization algorithms, such as space mapping [13]-[14], artificial neural networks [15]-[17] and shape-preserving response [18], have been successfully exploited to the applications of filters, couplers, antennas and transistor models.

In the field of intelligent algorithms, Bayesian optimization (BO) is an example of a supervised learning algorithm, used for the global optimization of an expensive objective function in a high-dimensional space [19]-[20]. It is particularly useful when the objective function is non-convex and multi-modal, especially when closed-form expressions for the objective function or its derivatives do not exist. Since the objective function is an unknown 'black-box', BO typically assumes that the function is sampled from Gaussian processes (GP) [21]. Due to its good properties of flexibility and tractability, BO is able to capture the characteristics of the unknown function by building a GP model based on the training data. After that, BO exploits an acquisition function to determine the next point to evaluate [22]. By repeating these steps, BO has the ability to achieve the global optimum value for the unknown objective function.

In [23], we applied Gaussian processes regression (GPR) to optimize the harmonic impedances for a continuous Class-F power amplifier design. Compared to [23], this paper expands GPR into the framework of BO, and applies BO to optimize the drain waveforms of an active circuit rather than of a passive circuit. In addition, the dimension of variables is increased up to 28, which is double that of [23]. Moreover, circuit simulation and electromagnetic (EM) simulation-based optimization are both implemented to design 10 W and 30 W broadband high-efficiency PAs.

In this paper, Section II first discusses the high-efficiency PA theory which serves as the objective function for the PA design optimization. Section III is devoted to describing the

This paper is an expanded version from the IEEE MTT-S International Microwave Symposium, Phoenix, AZ, USA, 17-22 May 2015. This work was supported by a research grant from Science Foundation Ireland (SFI) under its Investigators Programme 2012, grant number 12/IA/1267.

The authors are with the School of Electrical and Electronic Engineering, University College Dublin, Dublin 4, Ireland (e-mail: peng.chen@ucdconnect.ie; brian.merrick@ucdconnect.ie; tom.brazil@ucd.ie).

framework of Bayesian optimization, which is composed of Bayesian inference, modeling, prediction, and acquisition functions. Section IV describes the proposed automatic optimization-oriented strategy for PA design. Section V validates the proposed method through the presentation of 10 W and 30 W broadband high-efficiency PA designs. In this section, the algorithm convergence and measured results are shown to demonstrate that the proposed method is well-suited to broadband high-efficiency PA designs. Comparison designs using the ADS simulated annealing optimizer are also provided. Finally, in Section VI, conclusions are given.

II. High-efficiency PA Theory

The high efficiency power amplifier modes, for instance the Class-F/F^{-1}, achieve high efficiency due to the non-overlapping voltage and current waveforms present at the internal drain. It is of no doubt that waveform engineering has become a critical guiding principle in high-efficiency PA designs. In order to fully understand how the drain voltage, $v_{ds}(t)$, and drain current, $i_d(t)$, significantly affect the efficiency, it is necessary to analyze the power conversion in a PA system. In Fig. 1, the PA receives power both from the input signal and the DC supply. Since the input power is assumed to be totally dissipated in the input network and the transistor, it does not contribute directly to the output power. In (1), the DC power supplied to the device exits the device in three parts: in the form of heat dissipated in the active devices, P_{diss}, as a radio frequency (RF) signal at the fundamental, $P_{out,f}$, and RF signals at the harmonics, $P_{out,nf}$. As a consequence, the drain efficiency, η, is calculated by (2), and P_{diss} and $P_{out,nf}$ are defined in (3)-(4):

$$P_{DC} = P_{diss} + P_{out,f} + \sum_{n=2}^{\infty} P_{out,nf} \tag{1}$$

$$\eta = \frac{P_{out,f}}{P_{diss} + P_{out,f} + \sum_{n=2}^{\infty} P_{out,nf}} \tag{2}$$

$$P_{diss} = \frac{1}{T}\int_0^T v_{ds}(t)\cdot i_d(t)dt \tag{3}$$

$$\sum_{n=2}^{\infty} P_{out,nf} = \frac{1}{2}\sum_{n=2}^{\infty} V_n I_n \cdot \cos(\theta_n), \tag{4}$$

where V_n and I_n represent the drain voltage and drain current at the n^{th} harmonic with intersection angle, θ_n. It was pointed out by Colantonio that a PA with zero dissipated power, i.e., non-overlapping drain waveforms, may only have an efficiency of approximately 80%, depending on the waveforms. The ideal 100% efficiency requires (3)-(4) to satisfy the following condition [2]:

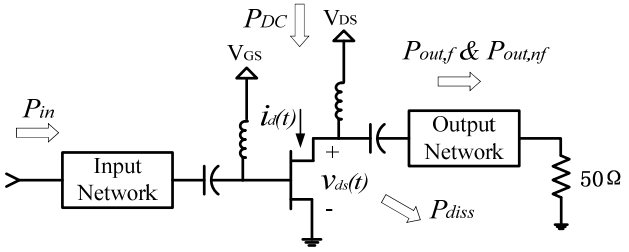


Fig. 1. Power conversion in a PA system.

$$P_{diss} + \sum_{n=2}^{\infty} P_{out,nf} = 0. \tag{5}$$

III. Bayesian Optimization

A. Bayesian Inference

In this paper, the problem of interest is to find the best design parameters for the matching networks such that the PA achieves high efficiency. The optimization problem can be viewed as a probability event (x, y), where $x \in \mathbb{R}^d$ denotes the design parameters and y denotes the observed PA efficiency. Suppose that the 'black-box' function is defined as

$$f(x;\mathrm{w}) = \sum_{i=1}^{n} \mathrm{w}^T \phi(x_i), \tag{6}$$

where w=(w_1,w_2,...,w_m) is a weighting vector, and $\phi(x)$ is a feature space mapping. Given a set of training design parameters, the accumulated observations are $D = \{x_i, y_i\}_{i=1}^n$. Let's define p(w) as the prior probability which represents the belief about $f(x;\mathrm{w})$ prior to the observation of D, and $p(D|\mathrm{w})$ as the likelihood function which expresses how probable high efficiency is given the function $f(x;\mathrm{w})$. According to Bayes' theorem, the posterior probability $p(\mathrm{w}|D)$ for the function after observing D takes the form:

$$p(\mathrm{w} \mid D) = \frac{p(D \mid \mathrm{w})p(\mathrm{w})}{p(D)} \tag{7}$$

$$p(D) = \int p(D \mid \mathrm{w})p(\mathrm{w})d\mathrm{w}, \tag{8}$$

where $p(\mathrm{w}|D)$ indicates the probability to evaluate the weighting vector, w, after the observations. Although the integrated term (8) cannot be expressed in a closed-form, it is a normalization constant. Thus, the posterior probability is proportional to the likelihood of D given w multiplied by the prior probability of w:

$$p(\mathrm{w} \mid D) \propto p(D \mid \mathrm{w})p(\mathrm{w}). \tag{9}$$

In practice, the maximum a posteriori (MAP) metric is a common approximation used to estimate (9), and is written in the following form [19]:

$$\begin{aligned} \mathrm{w}_{MAP} &= \arg\max_{\mathrm{w}} p(D \mid \mathrm{w}) p(\mathrm{w}) \\ &= \arg\max_{\mathrm{w}} \prod_{i=1}^{n} p(y_i \mid f(x_i;\mathrm{w})) p(\mathrm{w}). \end{aligned} \tag{10}$$

B. *Modeling and Prediction*

1) *'Black-box' Modeling*

As discussed before, a 'black-box' function which represents the PA system is needed to be found for the optimization design. The weighting vector w is determined by fitting (6) to the training data. When training the model, an error term is introduced to measure the misfit between the function $f(x;\mathrm{w})$ and the training data. One simple error function, which is widely applied [19], is given by

$$E(\mathrm{w}) = \frac{1}{2}\sum_{i=1}^{n}\{y_i - \mathrm{w}^T\phi(x_i)\}^2 + \frac{\lambda}{2}\|\mathrm{w}\|^2, \tag{11}$$

where $\lambda \|\mathrm{w}\|^2 / 2$ is a penalty term used to control the over-fitting phenomenon, called regularization, and λ governs the balance between the sum-of-squares error term and the regularization term.

As can be seen from (10), the probability distribution plays a critical role in Bayesian optimization. There are many distributions which can be used to form building blocks for complex models, such as the beta distribution, the Dirichlet distribution, and the exponential family distribution [19]. Fig. 2 describes the variation in correlation with the distance between two points for the exponential family distribution. The correlation decreases when two points move far away, while it increases when they are close. In the case of p=2, it is called a Gaussian distribution, which has a smoothness correlation over different points. The Gaussian distribution is written in the following form:

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right). \tag{12}$$

Due to these good properties, a Gaussian distribution is well-suited for the distribution of continuous variables in a high-dimensional space [20]. Assuming that the observed high efficiency, y, has a Gaussian distribution with a mean equal to $f(x;\mathrm{w})$, given the design parameters, x, then the likelihood function of (10) is expressed as

$$\prod_{i=1}^{n} p(y_i \mid f(x_i;\mathrm{w})) = \prod_{i=1}^{n} N(y_i \mid f(x_i;\mathrm{w}), \beta^{-1}), \tag{13}$$

where β is the inverse variance of the Gaussian distribution. The prior probability is commonly selected as a Gaussian distribution with a mean equal to 0 and an inverse variance of α, and can be written as

$$p(\mathrm{w}) = N(\mathrm{w} \mid 0, \alpha^{-1}). \tag{14}$$

Substituting (12) to (13)-(14). the negative log likelihood function for (13)-(14) are derived as

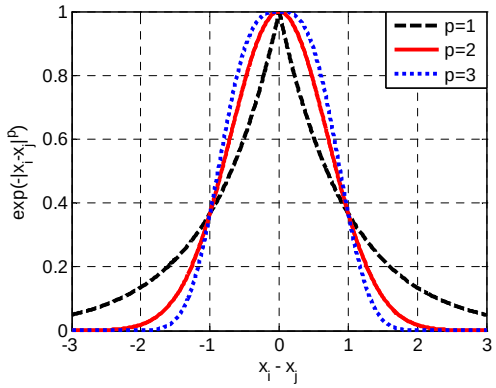


Fig. 2. The exponential family distribution.

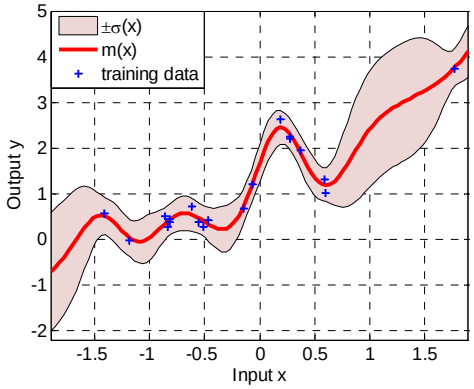


Fig. 3. The GP model with training data.

$$-\ln[\prod_{i=1}^{n} p(y_i \mid f(x_i;\mathrm{w}))] = \frac{\beta}{2}\sum_{i=1}^{n}\{y_i - \mathrm{w}^T\phi(x_i)\}^2 - \frac{n}{2}\ln\beta + \frac{n}{2}\ln(2\pi) \tag{15}$$

$$-\ln\{p(\mathrm{w})\} = \frac{\alpha}{2}\mathrm{w}^T\mathrm{w} - \frac{n}{2}\ln\alpha + \frac{n}{2}\ln(2\pi). \tag{16}$$

Similarly, taking the negative logarithm of (10) and combining with (15)-(16), the maximum of the posterior probability is equal to the minimum of

$$\frac{\beta}{2}\sum_{i=1}^{n}\{y_i - \mathrm{w}^T\phi(x_i)\}^2 + \frac{\alpha}{2}\mathrm{w}^T\mathrm{w}. \tag{17}$$

Comparing (11) with (17), we can see that maximizing the posterior probability is equivalent to minimizing the error function with the regularization parameter $\lambda = \alpha / \beta$. As a consequence, an accurate model for the PA system can be constructed based on the training data. This regression method is known as Gaussian processes regression (GPR), fitting the training data with a mean m(x) and a standard deviation σ(x), as depicted in Fig. 3.

2) *Prediction from GP Model*

A Gaussian process is a probability distribution over the function $f(x;\mathrm{w})$ such that the values of $f(x;\mathrm{w})$ evaluated at different points jointly have a Gaussian distribution. Suppose that x can be separated into two joint Gaussian distributed subsets, x_a and x_b:

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \tag{18}$$

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \tag{19}$$

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}, \tag{20}$$

where (19) and (20) represent the corresponding partitions of the mean vector μ and the covariance matrix Σ, respectively. The conditional distribution $p(x_b|x_a)$ is expressed with the following mean and covariance:

$$\mu_{b|a} = \mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a) \tag{21}$$

$$\Sigma_{b|a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}. \tag{22}$$

When applying the GP model for new predictions, it is necessary to add a noise term which serves as the unmatched error. Here, the function (6) is modified to the following form:

$$y = f(x) + \varepsilon \tag{23}$$

where $f(x)$ is specified by a mean function, m, and a kernel function, k, such that $f(x) \sim N(m(x), k(x, x'))$, and the noise ε also has a Gaussian distribution $\varepsilon \sim N(0, \sigma_n^2)$. According to the GP properties, the values y_i from the observation $D = \{x_i, y_i\}_{i=1}^n$ will have a multivariate Gaussian distribution $N(0, K)$, with the kernel matrix K given by

$$K = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix}, \tag{24}$$

where the kernel function $k(x_i, x_j)$ denotes the covariance matrix between x_i and x_j. The new value, y_{n+1}, which is predicted from the GP model at the new point, x_{n+1}, has a joint Gaussian distribution with $Y=(y_1, y_2, \cdots, y_n)$

$$\begin{bmatrix} y_{1:n} \\ y_{n+1} \end{bmatrix} \sim N\left[0, \begin{pmatrix} K + \sigma_n^2 & \mathrm{k} \\ \mathrm{k}^T & k(x_{n+1}, x_{n+1}) \end{pmatrix}\right] \tag{25}$$

$$\mathrm{k} = [k(x_{n+1}, x_1)\ k(x_{n+1}, x_2)\ \cdots\ k(x_{n+1}, x_n)]. \tag{26}$$

Referring to (21) and (22), the predicted data (x_{n+1}, y_{n+1}) from the GP model is specified as a Gaussian distribution

$$m(x_{n+1}) = \mathrm{k}^T[K + \sigma_n^2 I]^{-1} Y \tag{27}$$

$$\sigma^2(x_{n+1}) = k(x_{n+1}, x_{n+1}) - \mathrm{k}^T[K + \sigma_n^2 I]^{-1}\mathrm{k}. \tag{28}$$

C. Acquisition Functions for Searching

After building the GP model, a non-trivial task is how to search for the next point of interest from the constructed GP model. Acquisition functions are exploited to determine where to next evaluate the GP model. Since our goal is to find a set of design parameters such that high efficiency is obtained, the objective function is defined as

$$x^* = \arg\max_{x \in R^d} y(x). \tag{29}$$

In Bayesian optimization, there are four commonly used acquisition functions: probability of improvement (PI), expected improvement (EI), upper confidence bound, and entropy search [24]-[27]. In this paper, the first two acquisition functions are applied in order to search for the optimal values for the design parameters.

1) Probability of Improvement

PI evaluates $y(x)$ at the point where the improvement will most likely occur [24]. However, PI focuses on local optimization such that the points have high uncertainty. Given the current maximum point x^+, PI is calculated by

$$\begin{aligned} \mathrm{PI}(x) &= p(y(x) \geq y(x^+)) \\ &= \Phi\left(\frac{m(x) - f(x^+)}{\sigma(x)}\right), \end{aligned} \tag{30}$$

where $f(x^+)$ represents the current best value, $m(x)$ and $\sigma(x)$ are calculated from (27)-(28), and $\Phi(\cdot)$ is the normal cumulative distribution function.

2) Expected Improvement

EI has been shown to be an efficient criterion for finding the global optimum in many 'black-box' functions [25]. It allows us to make a trade-off between the local optimization and the global search. This acquisition can be computed analytically as

$$\mathrm{EI}(x) = \begin{cases} [m(x) - f(x^+)]\ \Phi(Z) + \sigma(x)\ \phi(Z) & \text{if } \sigma(x) > 0 \\ 0 & \text{if } \sigma(x) = 0 \end{cases} \tag{31}$$

$$Z = \frac{m(x) - f(x^+)}{\sigma(x)}, \tag{32}$$

where $\phi(\cdot)$ denotes the probability density function. In (31), the first term chooses the points where the mean is high while the second term targets the points where the variance is large.

Considering constraint conditions for the objective function, PI can be used to calculate the probability of being greater than the constraint limit [24]. In this situation, one model is built for the objective function; the other is for the constraint function. Since the two models are independent, the new point in (29) is obtained by maximizing

TABLE I
SUMMARY OF BAYESIAN OPTIMIZATION

Bayesian Optimization Algorithm
Obtain training data $D_0 = \{x_i, y_i\}_{i=1}^n$, select a GP distribution and a kernel function $k(x, x')$.
For $t = 1, 2, \ldots$, do
1. Build the GP model with the training data by maximizing MAP in (10).
2. Choose $x_t = \arg\max_{x \in R^d}[EI(x)PI(x)]$.
3. Obtain y_t at the new predicted x_t by using (27)-(28).
4. Augment $D_t = \{D_{t-1}, (x_t, y_t)\}$, and update the GP model.
End for

$$x^* = \arg\max_{x \in R^d} [EI(x)PI(x)]. \quad (33)$$

Table I summarizes the algorithm steps of Bayesian optimization. It is worth noting that two major choices significantly affect the global search of Bayesian optimization. The first is selecting an appropriate GP distribution and a kernel function for the modeling; the second is choosing an acquisition function to guide the search.

IV. Optimization-oriented Strategy

The automatic optimization strategy for the PA designs is accomplished by combining ADS with MATLAB or R. As is shown in Fig. 4, ADS is used to run the PA simulation so that training data is obtained. MATLAB or R is responsible for the Bayesian optimization since there are many related codes written in these two programming languages, which are available for public use. MATLAB or R starts the Bayesian optimization when the training data is exported from ADS, Similarly, ADS activates the simulation as soon as it receives new design parameters. Finally, the optimization strategy outputs the best design parameters to the users after the iterations finish.

In this paper, the design parameters are the widths and the lengths of the matching network transmission lines. The simplified real frequency technique (SRFT) is exploited to determine the matching network topology and the initial guess for the design parameters [28]-[29]. The S-parameters of the transistor are used as the inputs to the SRFT in order to generate this initial guess. As shown in Fig. 5, there are six stepped-impedance transmission lines and a bias line for both the input and output matching networks, resulting in 28 design parameters in total. The output targets in the optimization are the fundamental output power, and the harmonic and dissipated components described in Section II. It is important to note that it is the voltage and current waveforms at the current-generator plane that we are interested in optimizing, rather than the waveforms seen at the package plane. The objective function is defined as the root mean square (RMS) of the fundamental output power across the band:

$$y_1 = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(P_{out,f})_i^2}, \quad (34)$$

where m denotes the number of the frequency points. The constraint function is the sum of the RMS of the harmonic and dissipated power:

$$y_2 = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\sum_{n=2}^{5}(P_{out,nf})_i^2} + \sqrt{\frac{1}{m}\sum_{i=1}^{m}(P_{diss})_i^2}. \quad (35)$$

Harmonics greater than the fifth are neglected in this paper due to their minor impact upon efficiency improvement. As per (2) and (5), this paper optimizes the drain waveforms by maximizing (34) while minimizing (35), guaranteeing that high efficiency is achieved in a wide band.

Since the 28-dimensional space is very high, it requires a vast number of training data to capture the characteristics of the

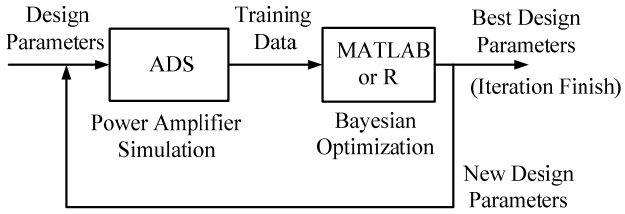


Fig. 4. Automatic optimization strategy.

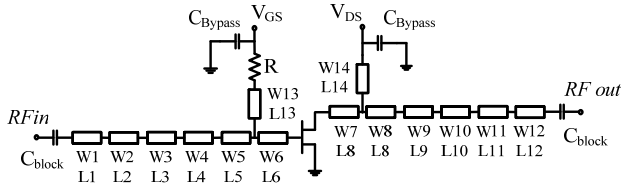


Fig. 5. Matching network topology.

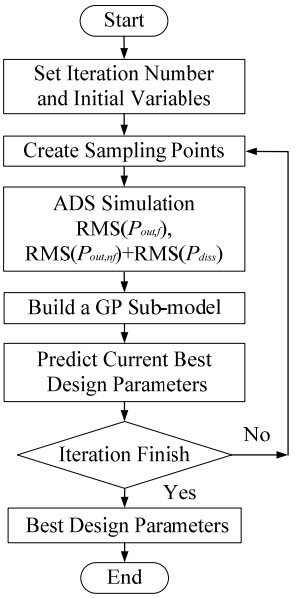
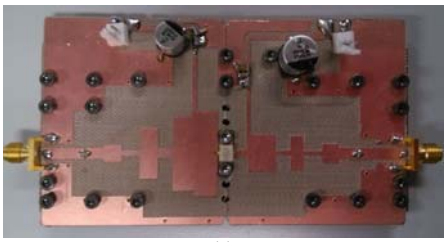


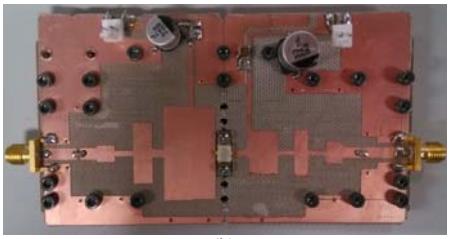
Fig. 6. Optimization process for the PA designs.

PA system, leading to quite expensive computation during the optimization. In order to balance the model accuracy and the computational time, this paper uses sub-models to search for design parameters instead of using the whole model. A sub-model is constructed around the current point, representing a portion of the whole model. After finding a better point, a new sub-model is constructed and is then applied to predict the next point. In this way, the sub-models are able to find the optimum point for the whole model. However, it should be pointed out that there is a risk of the optimization converging on a local minimum, rather than the desired global minimum.

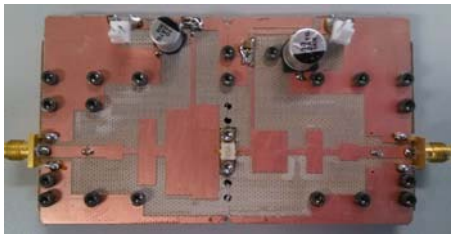
Fig. 6 further describes the optimization process more specifically for the PA designs. First, we set the iteration number and apply the SRFT to obtain the initial guess for the design parameters. The sampling points are created within a ±10% range of the current point using the Latin hypercube sampling technique [30]. By sweeping the sampling design parameters, ADS runs the simulation and exports the RMS of those in (34) and (35). With the available training data, the optimization algorithm builds the GP sub-models and predicts new design parameters by maximizing the acquisition function



(a)



(b)



(c)

Fig. 7. Fabricated 10 W PA designs. (a) EM-based BO with 50 training data points. (b) Circuit-based BO with 50 training data points. (c) Circuit-based BO with 200 training data points.

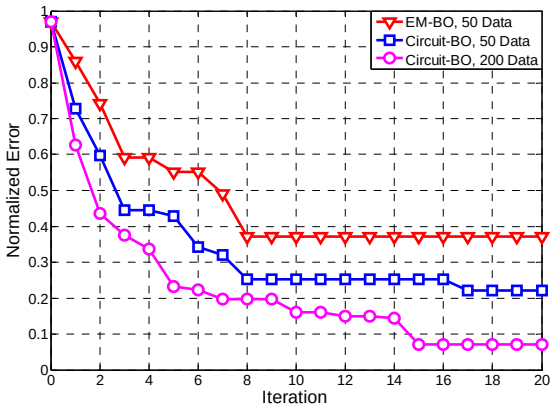


Fig. 8. Convergence for the 10 W PA designs.

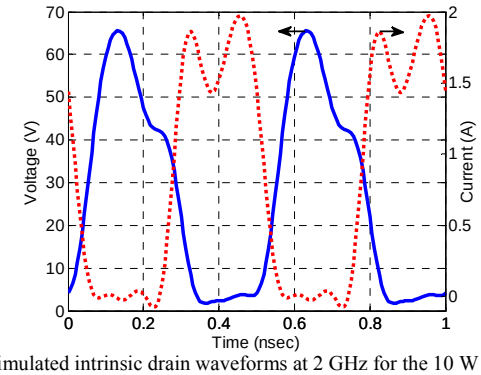


Fig. 9. Simulated intrinsic drain waveforms at 2 GHz for the 10 W PA design using EM-based BO.

in (33) using the DIRECT algorithm [31]. For the sake of good convergence, the optimization algorithm rejects predicted points which worsen performance and keeps the current best one. In order to avoid sub-optimal local minima in the optimization, the random seed is changed such that the sampling points and the sub-models are renewed at each iteration. The optimization algorithm repeats the above steps and finally outputs the optimal design parameters when the iterations are terminated. In this way, we apply Bayesian optimization to optimize a PA design automatically.

V. Practical PA Optimization

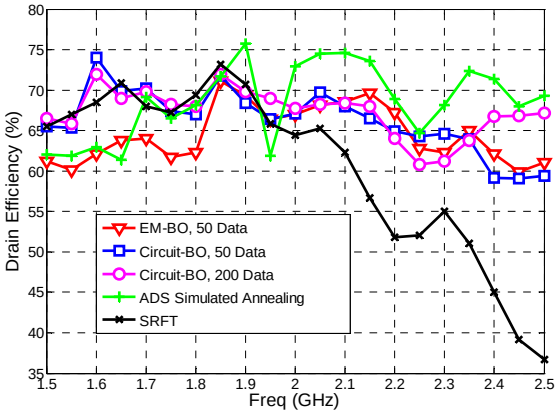
A. Broadband 10 W PA Design

In order to demonstrate the validity of this design approach, we optimized the designs of broadband high-efficiency 10 W PA using circuit-based and EM-based Bayesian optimization, respectively. Cree CGH40010F Gallium Nitride (GaN) high-electron mobility transistors (HEMTs) were used for the designs. In the circuit-based BO, the sub-models were built with the training data from circuit simulation. Similarly, in the EM-based BO, EM simulation supplied the training data for building the sub-models, which are more accurate than those of the circuit simulation. Considering that the EM simulation requires expensive computational time, we used 50 training data points to build the GP sub-models in the EM-based BO. In order to test the sensitivity of the optimization to the number of training data, 50 and 200 training data points were used in the circuit-based BO. For comparison, the ADS simulated annealing optimizer with 1000 iterations was employed to optimize the PAs while keeping the same variable settings and objective functions.

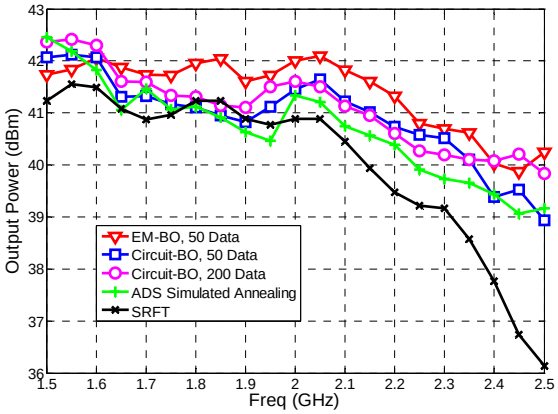
The transistor model is a critical part of the 'black-box' modeling, affecting the deviation between the simulated results and measurements to a great extent. Clearly, if the model is not of sufficient accuracy, then the optimization process will not result in good performance upon measurement. Cree's dynamic load-line model has been shown in a large number of papers to be of good accuracy [10]-[11], and, hence, was used for the designs. Additionally, the model in question gives access to the intrinsic current and voltage waveforms, as desired.

The PA designs were implemented on Taconic RF35 substrate with ε_r=3.5 and a thickness of 1.52 mm. The iteration number was set to t=20 and the initial design parameters of the matching networks of Fig. 5 were obtained using the SRFT, x_0= [W_1, W_2, ..., W_{14}, L_1, L_2, ..., L_{14}], where $W_{1\text{-}14}$= [5.75, 1.63, 15.30, 2.21, 27.91, 26.15, 2.09, 11.34, 1.57, 11.95, 1.83, 5.78, 1, 1] *mm*, and $L_{1\text{-}14}$= [5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 20, 20] *mm*. The bounded range for the width in the matching networks was 0.7 mm < W < 30 mm, corresponding to the characteristic impedance within 9 Ω < Z_c < 105 Ω on the RF35 board. Following the optimization process shown in Fig. 6, the 10 W PAs were optimized from 1.5 GHz to 2.5 GHz, with a step of 50 MHz. The optimization process ran on a computer featuring an Intel Core i7-3770 CPU @ 3.40 GHz with 16.0 GB RAM.

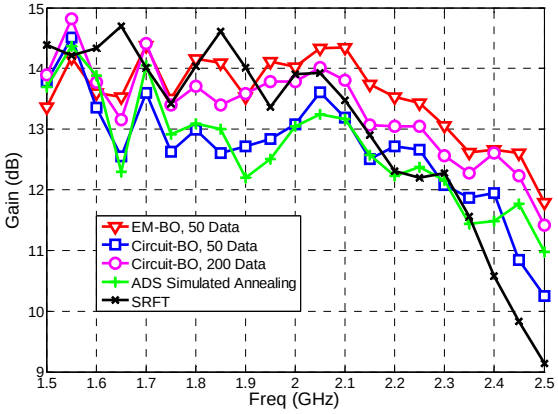
After 20 iterations, the EM-based BO outputted the



(a)



(b)



(c)

Fig. 10. Compared measurements for 10 W PA designs. (a) Drain efficiency. (b) Output power. (c) Gain.

optimized parameters, $W_{1\text{-}14}$ = [5.07, 1.51, 15.38, 2.79, 22.69, 26.12, 2.16, 11.19, 1.60, 10.63, 1.43, 5.96, 1.28, 1.11] *mm*, and $L_{1\text{-}14}$ = [5.20, 5.29, 5.47, 4.42, 5.76, 6.24, 3.54, 7.90, 4.31, 4.03, 5.66, 4.94, 18.54, 15.46] *mm*.

In the circuit-based BO with 50 training data points, the optimized design parameters were, $W_{1\text{-}14}$ = [4.77, 1.57, 18.24, 2.54, 28.58, 27.92, 2.19, 13.08, 1.46, 13.86, 1.42, 6.62, 1.41, 1.23] *mm*, and $L_{1\text{-}14}$ = [5.26, 5.55, 5.68, 3.41, 7.67, 6.33, 4.25, 7.89, 4.98, 4.02, 5.43, 4.67, 18.89, 19.22] *mm*.

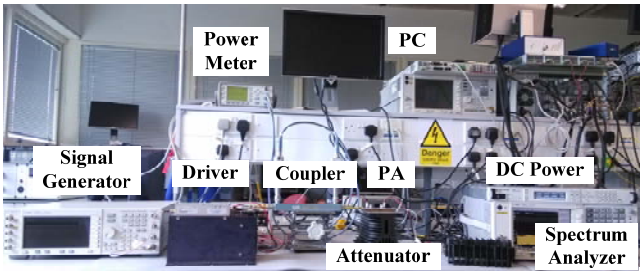
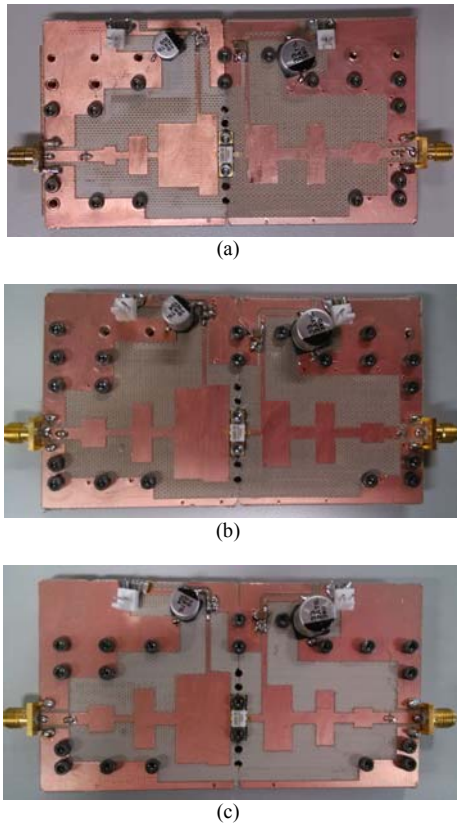


Fig. 11. Measurement setup.

As for the circuit-based BO with 200 training data points, we got the optimized parameters, $W_{1\text{-}14}$ = [5.01, 1.60, 20.14, 3.02, 26.05, 28.62, 2.10, 12.62, 1.57, 12.81, 1.64, 4.97, 1.32, 1.03] *mm*, and $L_{1\text{-}14}$ = [4.82, 4.55, 5.10, 2.96, 7.05, 7.16, 4.07, 9.80, 5.88, 4.46, 5.09, 5.37, 22, 21.04] *mm*.

Fig.7 shows the fabricated PAs which were optimized using the three approaches described above. The PAs were biased at a drain voltage of 28 V with a quiescent drain current of 70 mA. The convergence of the proposed optimization algorithm is shown in terms of a normalized error in Fig. 8. The normalized errors drop considerably over the 20 iterations, indicating that the algorithm converges very quickly. It was observed that the normalized error reduced when the training data number was increased from 50 to 200. Therefore, increasing the training data number is helpful in obtaining better design parameters from accurate models. The simulated drain waveforms of the PA using EM-based BO, depicted in Fig. 9, show only a small amount of overlap, suggesting high performance.

Fig. 10 shows the measurement results of the optimized PAs, together with comparison designs created using the SRFT and the ADS simulated annealing optimizer. These results were obtained with the measurement setup shown in Fig. 11, when an available input power of 28 dBm was provided across the band. The optimized PAs all showed large improvements compared with the initial design using the SRFT. The optimized PA using EM-based BO delivered drain efficiency above 60.0% from 1.5 GHz to 2.5 GHz, with output power greater than 39.9 dBm and gain larger than 11.8 dB. The PA using circuit-based BO with 200 training data points showed better performance than that using 50 training data, corresponding to the lower normalized error of Fig. 8. It achieved greater than 60.8% drain efficiency over the entire band, with output power greater than 39.8 dBm and gain larger than 11.4 dBm. Although the ADS simulated annealing optimizer achieved a drain efficiency of greater than 61.4% over the band, the output power decreased to 39.1 dBm, with gain dropping down to 11.0 dB. It should be noted that the EM-based design provided superior output power and gain when compared with the other designs, particularly at the upper end of the frequency scale.



(a) (b) (c)

Fig. 12. Fabricated 30 W PA designs. (a) EM-based BO with 50 training data points. (b) Circuit-based BO with 50 training data points. (c) Circuit-based BO with 200 training data points.

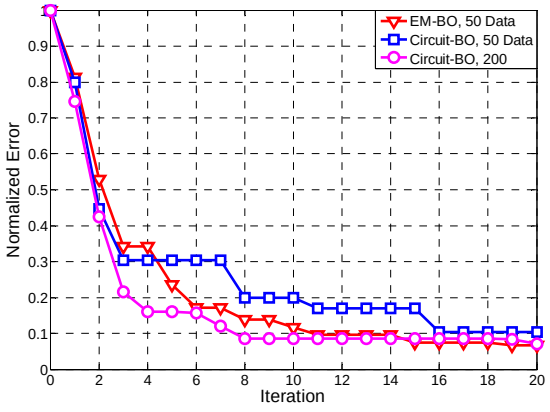
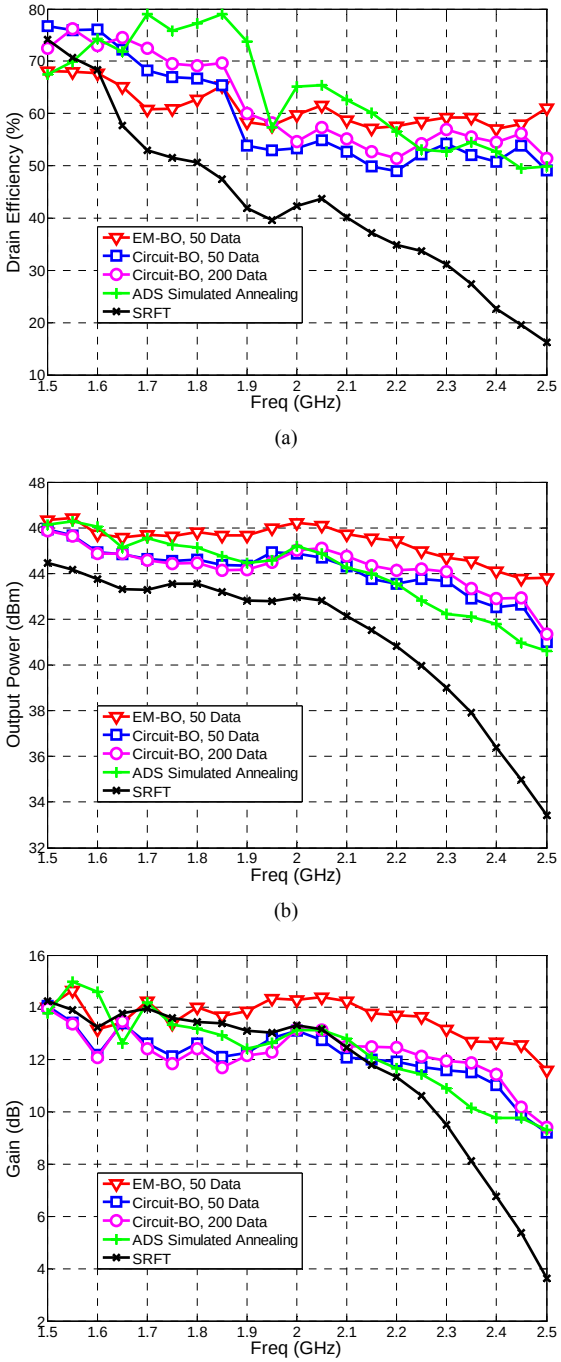


Fig. 13. Convergence for the 30 W PA designs.

B. Broadband 30 W PA Design

We also applied the above approaches to optimize the designs of 30 W PA operating from 1.5 GHz to 2.5 GHz. Cree CGHV40030F GaN HEMTs were used for the designs. Again, the Cree's dynamic load-line model was utilized in the ADS simulation. The SRFT provided the initial design parameters, $W_{1\text{-}14}$ = [7.01, 1.44, 20.79, 3.07, 26.86, 30, 1.66, 20.29, 4.01, 17.61, 3.08, 3.66, 0.8, 0.8] *mm*, and $L_{1\text{-}14}$ = [7, 7, 7, 7, 7, 7, 6.5, 6.5, 6.5, 6.5, 6.5, 6.5, 15, 20] *mm*.

As shown in Fig. 12, the first PA using EM-based BO had the dimensions of $W_{1\text{-}14}$ = [5.51, 1.84, 11.71, 1.50, 19.14, 24.64, 1.67, 17.57, 3.50, 18.40, 2.63, 6.51, 1.03, 1.17] *mm*, and $L_{1\text{-}14}$ = [3.37, 3.92, 5.65, 3.20, 11.79, 5.20, 3.34, 8.59, 8.92, 5.43, 8.28, 8.15, 20.21, 20.49] *mm*.



(a) (b) (c)

Fig. 14. Compared measurements for 30 W PA designs. (a) Drain efficiency. (b) Output power. (c) Gain.

The second PA using circuit-based BO with 50 training data points had the optimized parameters, $W_{1\text{-}14}$ = [8.33, 1.59, 17.11, 2.61, 27.13, 28.98, 1.55, 19.85, 3.62, 18.30, 2.85, 5.90, 1.04, 1.22] *mm*, and $L_{1\text{-}14}$ = [6.82, 7.41, 5.77, 6.60, 8.09, 6.96, 4.14, 7.60, 7.8, 5.77, 7.86, 7.12, 18.68, 15.59] *mm*.

As for the last design using circuit-based BO with 200 training data points, the optimized parameters were, $W_{1\text{-}14}$ =

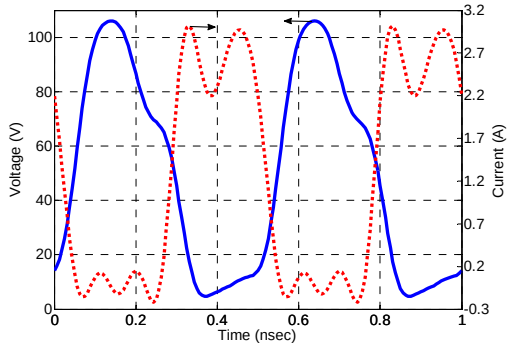


Fig. 15. Simulated intrinsic drain waveforms at 2 GHz for the 30 W PA design using EM-based BO.

TABLE II
COMPARISONS OF THE 30 W PA DESIGNS

Method	DE (%)	Pout (dBm)	Cost Time (min / iteration)
EM-BO, 50 Data	> 57.0	> 43.8	~ 70
Circuit-BO, 50 Data	> 49.2	> 41.0	~ 3
Circuit-BO, 200 Data	> 51.4	> 41.4	~ 4
ADS Simulated Annealing	> 49.5	> 40.6	~ 0.2

The measurements are compared from 1.5 GHz to 2.5 GHz. The cost time is the approximate average time for each optimization iteration.

[7.54, 1.58, 17.76, 3.70, 27.06, 28.52, 1.91, 19.93, 3.56, 16.64, 2.91, 5.75, 0.91, 1.26] *mm*, and $L_{1\text{-}14}$ = [7.22, 6.04, 6.0, 6.56, 9.23, 6.17, 4.58, 7.86, 7.60, 5.39, 6.35, 5.45, 17.33, 17.33] *mm*.

The algorithm convergence of the 30 W PA designs are shown in Fig. 13. The 30 W PAs were biased at a drain voltage of 50 V with a quiescent drain current of 95 mA, and were measured with 32 dBm available input power. As shown in Fig. 14, the optimized PA using EM-based BO obtained drain efficiency greater than 57.0% across the band, with output power greater than 43.8 dBm and gain larger than 11.6 dBm. As high power PAs have a smaller range of ideal harmonic impedances, circuit-based optimization designs show worse performances. The PA using the circuit-based BO with 200 training data gave a minimum drain efficiency of 51.5%, which was still better than that of using the ADS simulated annealing optimizer. The EM-based PA again showed significantly greater output power and gain than the other designs. The non-overlapping drain waveforms for the 30 W PA design using EM-based BO are shown in Fig. 15.

Table II gives the comparison of the 30W PAs using the above different methods. From the compared results, we can see that EM-based BO is more preferable than circuit-based BO in high power designs although it requires expensive computational time.

VI. CONCLUSIONS

A Bayesian optimization-based method is presented in this paper to design broadband high-frequency power amplifiers from 1.5 GHz to 2.5 GHz. This method is implemented automatically by combining ADS with MATLAB or R programming. Its aim is to maximize the fundamental output power, as well as minimize the harmonic and dissipated components, resulting in similar non-overlapping drain waveforms. The optimization algorithm builds the GP sub-models for the PA and utilizes an acquisition function to predict design parameters with which high efficiency is achieved. The 10 W and 30 W PA designs were optimized using circuit-based BO and EM circuit-based BO. Measured results show that the 10 W PA using circuit-based BO with 200 training data points obtained better than 60% drain efficiency and greater than 39.8 dBm output power from 1.5 GHz to 2.5 GHz. In the 30 W PA designs, EM-based BO performed much better than circuit-based BO, due to the higher power designs being more sensitive to simulation accuracy. The 30 W PA using EM-based BO operated at a drain efficiency of greater than 57%, with output power greater than 43.8 dBm across the band. Comparison results show that the proposed strategy greatly surpasses the ADS simulated annealing optimizer in terms of the gain and output power achieved across the band, with comparable efficiency shown with both methods. Moreover, it should be stressed that the performances of the optimized PAs also rely on the accuracy of the transistor models applied in the ADS simulation.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support of Cree, Inc.

REFERENCES

[1] S. C. Cripps, *RF Power Amplifier for Wireless Communications*, 2nd ed. Boston, MA: Artech House, 2006.

[2] P. Colantonio, F. Giannini, and E. Limiti, *High Efficiency RF and Microwave Solid State Power Amplifiers*. New York: Wiley, 2009.

[3] P. Wright, J. Lees, J. Benedikt, P. J. Tasker, and S. C. Cripps, "A methodology for realizing high efficiency Class-J in a linear and broadband PA," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 12, pp. 3196-3204, Dec. 2009.

[4] S. C. Cripps, P. J. Tasker, A. L. Clarke, J. Lees, and J. Benedikt, "On the continuity of high efficiency modes in linear RF power amplifiers," *IEEE Microw. Wireless Compon. Lett.*, vol. 19, no. 10, pp. 665–667, Oct. 2009.

[5] M. Eron, B. Kim, F. Raab, R. Caverly, and J. Staudinger, "The head of the class," *IEEE Microw. Mag.*, vol. 12, no. 7, pp. S16–S33, Dec. 2011.

[6] P. J. Tasker and J. Benedikt, "Waveform inspired models and the harmonic balance emulator," *IEEE Microw. Mag.*, vol. 12, no. 2, pp. 38–54, Apr. 2011.

[7] Y. Y.Woo, Y. Yang, and B. Kim, "Analysis and experiments for high efficiency Class-F and inverse Class-F power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 5, pp. 1969–1974, May 2006.

[8] V. Carrubba, A. L. Clarke, M. Akmal, J. Lees, J. Benedikt, P. J. Tasker, and S. C. Cripps, "On the extension of the continuous class-F mode power amplifier," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 5, pp. 1294–1303, May 2011.

[9] V. Carrubba, A. L. Clarke, M. Akmal, J. Lees, J. Benedikt, S. C. Cripps, and P. J. Tasker, "The continuous inverse class-F mode power amplifier with resistive second-harmonic impedance," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1928–1936, Jun. 2012.

[10] N. Tuffy, L. Guan, A. Zhu, and T. J. Brazil, "A simplified broadband design methodology for linearized high-efficiency continuous Class-F power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1952–1963, Jun. 2012.

[11] K. Chen and D. Peroulis, "Design of broadband highly efficient harmonic-tuned power amplifier using in-band continuous Class-F^{-1}/F mode transferring," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 12, pp. 4107–4116, Dec. 2012.

[12] S. Preis, D. Gruner, and G. Boeck, "Investigation of class-B/J continuous modes in broadband GaN power amplifiers," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2012, pp. 1–3.

[13] J. W. Bandler, Q. S. Cheng, S. A. Dakroury, A. S. Mohamed, M. H. Bakr, K.Madsen, and J. Søndergaard, "Space mapping: The state of the art,"

IEEE Trans. Microw. Theory Techn., vol. 52, no. 1, pp. 337–361, Jan. 2004.
[14] S. Koziel, J. W. Bandler, and K. Madsen, "A space mapping framework for engineering optimization: Theory and implementation," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 10, pp. 3721–3730, Oct. 2006.
[15] Q. J. Zhang, K. C. Gupta, and V. K. Devabhaktuni, "Artificial neural networks for RF and microwave design—From theory to practice," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 4, pp. 1339–1350,Apr. 2003.
[16] J. E. Rayas-Sánchez, "EM-based optimization of microwave circuits using artificial neural networks: The state-of-the-art," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 1, pp. 420–435, Jan. 2004.
[17] L. Zhang, K. Bo, Q.-J. Zhang, and J.Wood, "Statistical space mapping approach for large-signal nonlinear device modeling," in *36th Eur. Microw. Conf. Dig.*, Manchester, U.K., Sep. 2006, pp. 676–679.
[18] S. Koziel, "Shape-preserving response prediction for microwave design optimization," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 11, pp. 2829–2837, Nov. 2010.
[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, Aug. 2006.
[20] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *University of British Columbia, Tech. Rep.*, 2010.
[21] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for Machine Learning*. MIT Press, 2006.
[22] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *Proc. Neural Information and Processing Systems*, 2012.
[23] P. Chen and T. J. Brazil, "Gaussian processes regression for optimizing harmonic impedance trajectories in continuous Class-F power amplifier design," in *IEEE MTT-S Int. Microw. Symp. Dig.*, May 2015, pp. 1–3.
[24] A. I. J. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Progress in Aerospace Sci*, vol. 45, pp. 50–79, 2009.
[25] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black box functions," *Journal of Global Optimization*, vol. 13, pp.455–492, 1998.
[26] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," presented at *theInt. Conf. Machine Learning*, Haifa, Israel, 2010.
[27] P. Hennig and C. J. Schuler, "Entropy search for information-efficient global optimization," *Journal of Machine Learning Research*, 13, 2012.
[28] B. S. Yarman, *Design of Ultra Wideband Power Transfer Networks*. New York: Wiley, 2010.
[29] B. S. Yarman and H. J. Carlin, "A simplified "real frequency" technique applied to broadband multistage microwave amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol.30, no. 12, pp. 2216–2222, Dec. 1982.
[30] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979.
[31] D. E. Finkel, "Direct Optimization Algorithm User Guide," *North Carolina State Univ., Raleigh, CRSC Tech. Rep*. CRSC-TR03-11, Mar. 2003.

Peng Chen (S'15) received the B.E. degree in Communication Engineering and the M.E. degree in Electronic Engineering from Harbin Institute of Technology, Harbin, China, in 2010 and 2012, respectively. He is currently working towards the Ph.D. degree in the RF & Microwave Research Group, University College Dublin, Dublin, Ireland.

His research interests include non-linear device modeling and optimization algorithms applied to power amplifier design.

Brian M. Merrick (S'12-M'15) received the B.E. and Ph.D. degrees from University College Dublin (UCD), Dublin, Ireland, in 2010 and 2015, respectively. He is currently a postdoctoral researcher with the RF & Microwave Research Group at UCD.

His research interests include device characterization and non-linear modeling, and the design of high-efficiency, broadband power amplifiers.

Thomas J. Brazil (F'03) received the PhD degree from the National University of Ireland in 1977. He holds the Chair of Electronic Engineering at University College Dublin's School of Electrical & Electronic Engineering.

His research interests are in the fields of non-linear modelling and characterisation techniques at the device, circuit and system level within high frequency electronics. He is head of the RF & Microwave Research Group in UCD. He was elected a Fellow of the IEEE in 2003, and elected a Member of the Royal Irish Academy (RIA) in 2004. He is a serving member of IEEE MTT-S AdCom.