



Markov chain Monte Carlo methods for statistical analysis of RF photonic devices

Piels, Molly; Zibar, Darko

Published in:
Optics Express

Link to article, DOI:
[10.1364/oe.24.002084](https://doi.org/10.1364/oe.24.002084)

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Piels, M., & Zibar, D. (2016). Markov chain Monte Carlo methods for statistical analysis of RF photonic devices. *Optics Express*, 24(3), 2084-2097. DOI: 10.1364/oe.24.002084

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Markov chain Monte Carlo methods for statistical analysis of RF photonic devices

Molly Piels* and Darko Zibar

DTU Fotonik, Department of Photonics Engineering, Technical University of Denmark, Kgs. Lyngby 2800

*mopi@fotonik.dtu.dk

Abstract: The microwave reflection coefficient is commonly used to characterize the impedance of high-speed optoelectronic devices. Error and uncertainty in equivalent circuit parameters measured using this data are systematically evaluated. The commonly used nonlinear least-squares method for estimating uncertainty is shown to give unsatisfactory and incorrect results due to the nonlinear relationship between the circuit parameters and the measured data. Markov chain Monte Carlo methods are shown to provide superior results, both for individual devices and for assessing within-die variation.

© 2016 Optical Society of America

OCIS codes: (130.0130) Integrated optics; (220.0220) Optical design and fabrication; (250.5300) Photonic integrated circuits.

References and links

1. C. R. Doerr, "Silicon photonic integration in telecommunications," *Frontiers in Physics* **3**, 37 (2015).
2. S. Bowers, B. Abiri, F. Aflatouni, and A. Hajimiri, "A compact optically driven travelling-wave radiating source," in *Optical Fiber Communication Conference*, (OSA, 2014), p. Tu2A.3.
3. C.-M. Chang, J. H. Sinsky, P. Dong, G. de Valicourt, and Y.-K. Chen, "High-power dual-fed traveling wave photodetector circuits in silicon photonics," *Opt. Express* **23**, 22857 (2015).
4. M. S. Hai, M. Ménard, and O. Liboiron-Ladouceur, "A 20 Gb/s SiGe photoreceiver based on optical time sampling," in *European Conference on Optical Communications*, (2015), p. Tu.1.3.5.
5. K. Minoglou, E. D. Kyriakis-Bitzaros, D. Syvridis, and G. Halkias, "A compact nonlinear equivalent circuit model and parameter extraction method for packaged high-speed VCSELs," *J. Lightw. Technol.* **22**, 2823–2827 (2004).
6. P. Verheyen, M. Pantouvaki, J. Van Campenhout, P. P. Absil, H. Chen, P. De Heyn, G. Lepage, J. De Coster, P. Dumon, A. Masood, D. Van Thourhout, R. Baets, and W. Bogaerts, "Highly uniform 25 Gb/s Si photonics platform for high-density, low-power WDM optical interconnects," in *Advanced Photonics for Communications*, (OSA, 2014), p. IW3A.4.
7. S. M. Sze, *Physics of Semiconductor Devices* (John Wiley & Sons, 1981), 2nd ed.
8. G. P. Agrawal, *Optical Receivers* (John Wiley & Sons, 2002), vol. 6, chap. 4, 3rd ed.
9. G. Wang, T. Tokumitsu, I. Hanawa, K. Sato, and M. Kobayashi, "Analysis of high speed p-i-n photodiode S-parameters by a novel small-signal equivalent circuit model," *IEEE Microwave and Wireless Components Lett.* **12**, 378–380 (2002).
10. R. Lewén, S. Imscher, and U. Eriksson, "Microwave cad circuit modeling of a traveling-wave electroabsorption modulator," *IEEE Trans. Microwave Theory Tech.* **51**, 1117–1128 (2003).
11. Cascade Microtech, *Calibration Tools: Consistent Parameter Extraction for Advanced RF Devices* (2012).
12. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman & Hall, 1996).
13. A. Tarantola, *Inverse Problem Theory* (SIAM, 2005).
14. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer Science+Business, 2006), 1st ed.
15. M. Piels and J. E. Bowers, "40 GHz Si/Ge uni-traveling carrier waveguide photodiode," *J. Lightw. Technol.* **32**, 3502–3508 (2014).

16. T.-W. Weng, Z. Zhang, Z. Su, Y. Marzouk, A. Melloni, and L. Daniel, "Uncertainty quantification of silicon photonic devices with correlated and non-Gaussian random parameters," *Opt. Express* **23**, 4242–4254 (2015).
 17. D. Melati, E. Lovati, and A. Melloni, "Statistical process design kits: analysis of fabrication tolerances in integrated photonic circuits," in *Advanced Photonics 2015*, (OSA, 2015), p. IT4A.5.
 18. C. Michael and M. Ismail, "Statistical modeling of device mismatch for analog MOS integrated circuits," *IEEE J. Solid-State Circuits* **27**, 154–166 (1992).
 19. Y. Yang, Y. Ma, H. Guan, Y. Liu, S. Danziger, S. Ocheltree, K. Bergman, T. Baehr-Jones, and M. Hochberg, "Phase coherence length in silicon photonic platform," *Opt. Express* **23**, 16890–16902 (2015).
 20. A. Gelman, X.-L. Meng, and H. Stern, "Posterior predictive assessment of model fitness via realized discrepancies," *Statistica Sinica* **6**, 733–807 (1996).
 21. X. Zhang, A. Hosseini, H. Subbaraman, S. Wang, Q. Zhan, J. Luo, A. K.-Y. Jen, and R. T. Chen, "Integrated photonic electromagnetic field sensor based on broadband bowtie antenna coupled silicon organic hybrid modulator," *J. Lightw. Technol.* **32**, 3774–3784 (2014).
 22. X. S. Yao and L. Maleki, "Dual microwave and optical oscillator," *Opt. Lett.* **22**, 1867–1869 (1997).
 23. W. Loh, S. Yegnanarayanan, J. J. Plant, F. J. O'Donnell, M. E. Grein, J. Klamkin, S. M. Duff, and P. W. Juodawlkis, "Low-noise RF-amplifier-free slab-coupled optical waveguide coupled optoelectronic oscillators: physics and operation," *Opt. Express* **20**, 19420–19430 (2012).
-

1. Introduction

There has recently been a general movement towards a generic foundry approach to photonic integrated circuit design and fabrication. In this approach, a foundry service provides cross-section information and models of building blocks to clients, who design a photonic integrated circuit (PIC) [1] to fit their own specific needs. Recent demonstrations of high-speed PICs fabricated using foundry services [2–4] have shown that it is possible to design PICs with performance better than that of single building blocks. In order for first-time-correct designs to become consistently achievable, accurate microwave impedance models must be included in foundry libraries. Methods for fitting parameter values in an equivalent-circuit model to experimental data (usually the microwave reflection coefficient) are widely used [5, 6]. However, because curve-fitting is nonlinear and because the models are not typically perfect descriptions of the device under test, the fitting process is prone to systematic errors. Despite this, there have been very few studies that critically examine the accuracy of the fit, e.g. by calculating error bars on estimated parameter values. Here, we show that the commonly used nonlinear least squares method can dramatically under- and over-estimate the uncertainty associated with the equivalent circuit model. We further show that Markov chain Monte Carlo (MCMC) methods, specifically the Metropolis-Hastings (MH) algorithm, provide a simple, robust method for statistical descriptions of RF photonic devices both individually and across a wafer.

2. Measurement and Modeling

The end goal of parameter fitting is to find the parameters that best describe the measured data using a model and calculate the uncertainty associated with these values. This requires a systematic examination of both the model and sources of measurement error or noise. A lumped-element RF model of a diode is shown in Fig. 1. The impedance, Z_{PD} , can be modeled as a parallel resistance and capacitance along with a series resistance [7]. Several variants of this model have been used: the parallel resistance R_p is sometimes assumed to be infinite [8], and an additional capacitance, C_{dx} , has been placed in parallel between R_s and ground to distinguish between the intrinsic diode and fringe capacitances [9]. Traveling wave models [10] are more accurate, but not necessary in this case because the device length is far smaller than 1/16 of the microwave wavelength. The relative predictive powers of competing models can be evaluated using a cross-validation procedure, as will be discussed in Section 4.3, though as a general principle, fitting more parameters results in larger confidence intervals and fits that more closely follow the data. The data is the microwave reflection coefficient ($\Gamma(\omega)$, or $S_{11}(\omega)$) measured

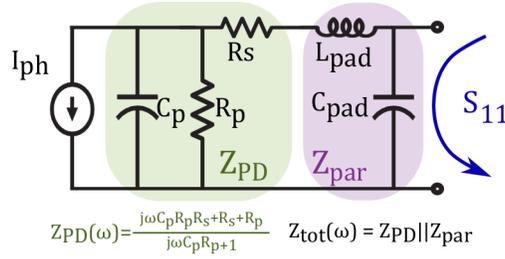


Fig. 1. Small-circuit model of a photodiode. The intrinsic diode impedance is shown at left, while pad parasitics are on the right.

using a network analyzer. Though the diode impedance is also related to the bandwidth of the device, S_{21} , the reflection data is used more often to fit the impedance because it is not affected by the carrier transit time. The microwave reflection coefficient is related to the impedance by

$$S_{11}(\omega) = \frac{Z_{tot}(\omega) - Z_0}{Z_{tot}(\omega) + Z_0} \quad (1)$$

where Z_0 is the characteristic impedance of the network analyzer (50Ω) and ω is the angular frequency. The photodiode impedance can be written as

$$Z_{PD}(\omega) = \frac{j\omega C_p R_s + R_s + R_p}{j\omega C_p R_p} \quad (2)$$

where C_p is the parallel capacitance, R_s is the series resistance, and R_p is the parallel resistance. The total impedance Z_{tot} is the diode impedance in parallel with the parasitic impedance. As Eqs. (1) and (2) show, the relationship between the measurement and the model parameters is nonlinear, and this complicates both the fitting process and confidence interval calculation. Whether it is best to fit to S_{11} , Z_{PD} , or Y_{PD} ($1/Z_{PD}$) depends on: (1) how much these quantities change as parameters are varied; and (2) how easy it is to incorporate known sources of measurement error into the expressions for the measured data.

For on-wafer tests, the primary sources of error are not noise but rather: (1) pad parasitics; (2) calibration error; and (3) an amplitude error due to the finite directivity of the network analyzer. The pad parasitics are also depicted in Fig. 1, and are modeled as an inductance and capacitance and are best dealt with by being included in the fitting process. Calibration error depends on the kind of calibration used. Though short-open-load (SOL) calibrations are less repeatable than other kinds, they are more commonly used for optoelectronic devices because they do not require mating connectors on both network analyzer ports. For SOL calibrations with a high-quality standard, error will arise primarily from probe misplacement during the calibration itself [11] as well as the different probe-to-pad interfaces on the calibration standard and the wafer. This results in an uncertain inductance from the transmission line discontinuity at the probe-to-pad interface as well as a random translation of the microwave reference plane. The inductance can be included as variation in the probe pad inductance, which is in most cases too small to measure. The translation of the reference plane contributes a frequency-dependent phase shift to the measured reflection coefficient:

$$S_{11,meas}(\omega) = e^{2j\omega\beta\Delta\ell} S_{11}(\omega), \quad (3)$$

where β is a microwave propagation constant and the probe placement error is $\Delta\ell$. For placement errors on the order $10 \mu\text{m}$, the magnitude of this error can be significant.

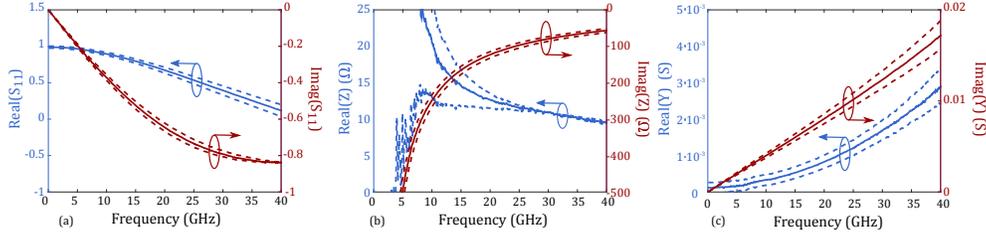


Fig. 2. Microwave reflection data and error margins in the (a) S_{11} , (b) Z , and (c) Y domains. The center curve in each figure is smoothed measured data, while the upper and lower bounds come from the worst case (in the mean-squared sense) error possible for that measurement. The directivity was 37 dB, the microwave propagation constant of the pads was $7.5e-9 \text{ m}^{-1}$, and the misplacement is $25 \mu\text{m}$

The finite directivity of network analyzer results in an unwanted leakage wave at the microwave detector. This leakage wave has a known maximum amplitude relative to the desired reflected wave, but the phase difference between them is unknown. Due the length difference between the leakage wave path and the reflected wave path and frequency-dependent phase response of passive microwave components, the phase difference also has a frequency dependence. Thus we treat the total amplitude error as a uniformly distributed random variable on $[-a, a]$, where a is the maximum amplitude of the wave, and assume that this error is different at all frequencies and for all devices. The effect of the placement error and finite directivity as a function of frequency on $S_{11, meas}$, Z_{PD} , and Y_{PD} is shown in Fig. 2. The center curves shown in Fig. 2 are smoothed measured data, while the upper and lower dashed curves indicate upper and lower error bounds. These were calculated by considering absolute maximum values of both the probe placement error and the amplitude error. As all curves show, the error is frequency-dependent; this is due to the probe misplacement term (Eq. (3)). Fitting the probe misplacement as a separate parameter simplifies the problem by removing this frequency dependence. It also yields smaller confidence intervals for each fit, as will be discussed in Section 3.2.

3. Fitting the response of a single device

In this section, we will introduce the general probabilistic approach to describing uncertainty, describe the Metropolis-Hastings and nonlinear least squares algorithms as they relate to fitting a single curve, and compare the results of different fitting procedures.

3.1. Probabilistic framework

In the probabilistic approach to parameter fitting, all uncertainties are expressed as probability density functions (PDFs), and the goal of the fitting is to find the posterior PDF of model parameters conditioned on the data. A curve fitting result then returns both an estimated parameter value and the degree of confidence we should have in that value based on the experiment performed. A correct result will both attribute a high probability to the true parameter value and accurately describe the efficacy of the experiment.

Generally speaking, we can write the outcome of a measurement \mathbf{d} as

$$\mathbf{d} = g(\mathbf{m}) \quad (4)$$

where \mathbf{m} is a vector of model parameters and g is a function that relates the parameters to the data. All three entities, including g , have associated uncertainties due to noise, mismatch between the model and the true device, and inherent measurement limitations. Mathematically

speaking, the goal is to solve for the posterior probability density function of the model parameters given the measured data and any prior assumptions we have made about the model parameters, $\sigma(\mathbf{m})$. This can be written as

$$\sigma(\mathbf{m}) = k\rho(\mathbf{d}|\mathbf{m})\rho_M(\mathbf{m}) \quad (5)$$

where k is a normalization constant, $\rho(\mathbf{d}|\mathbf{m})$ is the likelihood function, and ρ_M is the prior PDF of the parameters. The likelihood function is the probability of the observed data predicted by the model. The prior distribution represents assumptions about parameter values made prior to collecting data. It is not necessary to use a meaningful prior distribution, but it is often useful. Priors offer a convenient way to include imprecision in calibration quantities in the final result. For example, if the pad capacitance has been measured using independent test structures but there is known variation between different sets of pads or uncertainty in the measurement, fitting for all the parameters in the circuit with this expressed as a prior assumption will yield PDFs of other circuit parameters that take it into account. Bounding search intervals for parameter values can also be viewed as using a prior; in this case, the prior would be the uniform distribution over the search interval. In the absence of uncertainty in the form of g , the likelihood function $\rho(\mathbf{d}|\mathbf{m})$ depends only on the measurement noise. For example, for Gaussian noise with covariance \mathbf{C}_D , this becomes

$$\rho(\mathbf{d}|\mathbf{m}) = k\exp\left(-\frac{1}{2}(\mathbf{d} - g(\mathbf{m}))^T \mathbf{C}_D^{-1} (\mathbf{d} - g(\mathbf{m}))\right). \quad (6)$$

The maximum likelihood solution, \mathbf{m}_{ML} ($= \text{argmax}(\rho(d|m))$), is defined as the model parameter vector that maximizes the likelihood function, and is also often referred to as the best fit.

As discussed in Section 2, the measurement noise distribution, though known, is not Gaussian. This presents no difficulties for Markov chain Monte Carlo fitting, but is not compatible with the least-squares method. Furthermore, the effect of modelization uncertainty on the likelihood function is difficult to quantify. That is, there is almost always some uncertainty associated with g , and it is not usually well-understood. Thus parameter fitting also requires comparing the posterior estimate of the measurement error distribution with the likelihood function used to estimate the model parameters. The posterior estimate of the measurement error can be calculated from the residuals

$$\mathbf{r} = \mathbf{d} - g(\bar{\mathbf{m}}) \quad (7)$$

where $\bar{\mathbf{m}}$ is the mean of the posterior PDF of the parameters. The distribution of these residuals can then be compared (qualitatively) to the measurement error distribution originally assumed. For a perfect model, the residuals would be drawn from the same distribution as the assumed measurement error. In practice, there is often a visible trend in the residuals due to model imperfection. If the magnitude of the variation shown in this trend is similar to the magnitude of the measurement noise, then it can be difficult to remove: if two competing models are being compared, but the predictions they give differ by an amount similar to the noise level, it is not possible to determine which is more accurate.

3.2. Markov Chain Monte Carlo

For large parameter spaces, it is often impractical to evaluate $\sigma(\mathbf{m})$ at all points of interest. In these cases, well-designed random explorations of the parameter space are typically more computationally efficient. Markov chain Monte Carlo methods are a group of algorithms for constructing random walks through the model space (Markov chains) that produce samples from a target PDF ($\sigma(\mathbf{m})$) [12–14]. For fitting a single curve, we use the Metropolis-Hastings algorithm. The MH algorithm factors the PDF from which we are sampling into two parts: a proposal PDF of our choosing, and a weighting function that is proportional to the target PDF but

potentially difficult to generate samples from. In this case, we use the function $k\rho(\mathbf{d}|\mathbf{m})\rho_M(\mathbf{m})$, which is the target PDF according to Eq. (5), as the weighting function. Note that if we do not want to apply a prior, we simply set ρ_M constant. At each step in the chain, a new sample \mathbf{w} is generated from the proposal distribution $q(\mathbf{w}|\mathbf{x}_n)$, where \mathbf{x}_n is the last sample generated by the chain. The proposal sample is then accepted with probability P :

$$P = \left\{ \frac{k\rho(\mathbf{d}|\mathbf{w})\rho_M(\mathbf{w})q(\mathbf{x}_n|\mathbf{w})}{k\rho(\mathbf{d}|\mathbf{x}_n)\rho_M(\mathbf{x}_n)q(\mathbf{w}|\mathbf{x}_n)}, 1 \right\} \quad (8)$$

where k has been included for consistency but clearly need not ever be calculated. If the candidate point is accepted, then $\mathbf{x}_{n+1} = \mathbf{w}$; otherwise, $\mathbf{x}_{n+1} = \mathbf{x}_n$. Common choices of proposal distributions ($q(\mathbf{w}|\mathbf{x}_n)$) include the uniform distribution centered at \mathbf{x}_n and a standard normal distribution centered at \mathbf{x}_n , in which case the q terms in Eq. (8) cancel. Another possibility is to use the prior distribution $\rho_M(\mathbf{m})$, which also results in significant simplification of Eq. (8).

The Markov chain is run until it produces the desired number of independent samples after convergence. Convergence is assessed using Gelman-Rubin statistics, which require that several chains be run in parallel. The method compares the between-sequence variance B and within-sequence variances W [12]:

$$B_\phi = \frac{N}{M-1} \sum_{i=1}^M (\bar{\phi}_i - \bar{\phi})^2 \quad \text{where } \bar{\phi}_i = \text{mean}(\phi_{ij}) \text{ and } \bar{\phi} = \text{mean}(\phi_i) \quad (9)$$

$$W_\phi = \frac{1}{M} \sum_{i=1}^M s_i^2 \quad \text{where } s_i^2 = \text{var}(\phi_{ij}). \quad (10)$$

where ϕ is the parameter of interest, M is the number of threads (indexed by i) and N is the number of iterations (with index j). These two variances can be used to over-estimate the variance of ϕ :

$$\widehat{\sigma}_\phi^2 = \frac{N-1}{N} W_\phi + \frac{1}{N} B_\phi. \quad (11)$$

This is an over-estimate if the local means of the chains differ, as is expected during convergence. Since W_ϕ underestimates the variance of ϕ prior to convergence, the ratio of the two can be used to assess whether or not convergence has been achieved:

$$\beta = \sqrt{\frac{\widehat{\sigma}_\phi^2}{W_\phi}}. \quad (12)$$

A chain is considered to have converged when the value of β associated with each parameter is below 1.1 [12].

Designing the random walk typically involves a trade-off between acceptance rate and convergence time. Example convergence behavior for several different proposal distributions for the diode capacitance are shown in Fig. 3. The output of the MH algorithm is a set of samples, and the curves in the figure are calculated using the kernel smoothing density estimate. The likelihood function was assumed to be uniformly distributed around the measured data and bounded by the directivity and maximum feasible probe misplacement, *i.e.* bounded by the dashed lines in Fig. 2(a). The proposal distribution was a uniform distribution centered at \mathbf{x}_n with varying maximum step size. Other parameters were confined to narrow ranges by centering the priors at the maximum likelihood value and using the priors as the proposal distributions. No prior was used for the device capacitance. As the maximum step size for the capacitance is increased, the convergence time decreases, but the final acceptance rate decreases also. The PDF of the estimated parameter is independent of the proposal distribution, as shown in Fig. 3(b).

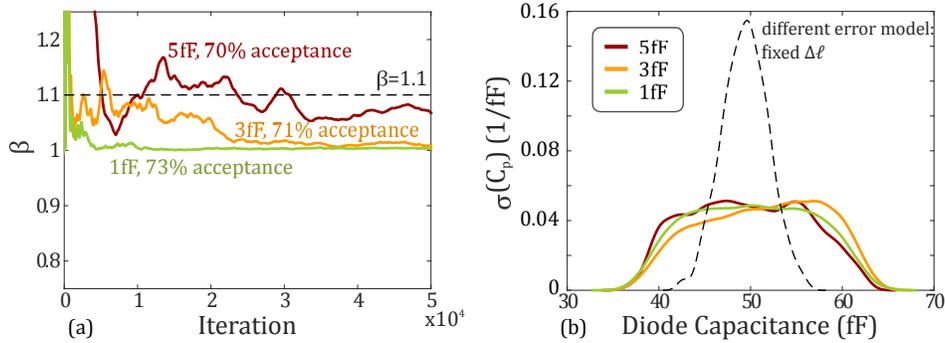


Fig. 3. Influence of Markov chain design parameters on convergence and output. (a) Convergence and final acceptance rates for different proposal distribution widths. (b) Probability density function of the diode capacitance for different proposal distribution widths and problem formulations. Requiring a single probe placement error over the entire fitting range results in a narrower PDF.

As mentioned in Section 2, the error estimate has a strong effect on the shape and size of the PDF calculated for each parameter. Larger error estimates yield wider (more conservative) PDFs; since narrower PDFs are more precise, it is preferable to use the smallest error possible. Under-estimating the error, however, will result in an incorrect estimate of the parameter PDF; incorrect PDFs generated this way usually have a small number of separate peaks because the acceptance rate is very low. Figure 3(b), in addition to the three PDFs calculated using different step sizes, also shows one PDF calculated using a different likelihood function. For this curve, instead of allowing the probe placement error to vary between different frequency points, it was fit as an independent parameter. This results in a smaller error estimate and therefore much more precise PDFs, and is the approach taken in all subsequent sections.

3.3. Nonlinear Least-Squares

Nonlinear least-squares fitting (NLLS) is a common and popular form of curve fitting. The underlying assumption is that all parameters and the measurement noise will have a Gaussian distribution with a single mean and unique covariance matrix. This makes it possible to solve for the mean value of $\sigma(\mathbf{m})$ and then find the associated covariance matrix afterward. There are many algorithms for finding the maximum likelihood solution [13]; we used the gradient descent method. As a general principle, such directed search algorithms are more computationally efficient than Monte Carlo methods. Once the mean has been found, the Jacobian of g , \mathbf{G} , is used to solve for the covariance matrix:

$$\widetilde{\mathbf{C}}_{\mathbf{M}} = \left(\mathbf{G}\mathbf{C}_{\mathbf{D}}^{-1}\mathbf{G}^T + \mathbf{C}_{\mathbf{M},\text{prior}}^{-1} \right)^{-1} \quad (13)$$

where $\widetilde{\mathbf{C}}_{\mathbf{M}}$ is the posterior estimate of the covariance matrix of the model parameters, $\mathbf{C}_{\mathbf{M},\text{prior}}$ is the prior estimate, and $\mathbf{C}_{\mathbf{D}}$ is the covariance of the data uncertainty. Because the noise in the measurement is non-Gaussian, this is not well-defined, but can be approximated as a diagonal matrix with entries a^2 where a is the maximum distance from the measured to the true value. In the case where no prior estimate for the parameter covariance matrix is available, $\mathbf{C}_{\mathbf{M},\text{prior}}^{-1} = \mathbf{0}$.

3.4. Comparison

For nearly-linear problems with nearly-Gaussian noise, both the nonlinear least squares method and Markov chain Monte Carlo give similar results. In these cases, nonlinear least-squares is preferable because it will run faster. However, the Gaussian assumption underlying NLLS can become problematic when a wide range of possible parameter values is considered and the model is nonlinear. Figure 4(a) shows the joint PDF of the series resistance R_s and diode capacitance C_p calculated using MCMC and two NLLS local minima. For all PDFs shown, all five parameters (R_s , C_p , C_{pad} , R_p , and $\Delta\ell$) are fit simultaneously and their values are only constrained to physically reasonable values. The device under test was a $4\ \mu\text{m} \times 70\ \mu\text{m}$ Si/Ge uni-traveling carrier photodiode [15]. Unsurprisingly, there is some ambiguity in how the total capacitance, $C_p + C_{pad}$, is split between the two elements. What is perhaps less obvious is that the splitting ratio affects the best fit for the series resistance. If this were not the case, the PDFs depicted in Fig. 4(a) would be roughly elliptical with major axes parallel to the graph axes. In the MCMC result, shown in blue, this manifests as a curved PDF in the 2-D space and a nearly-uniform or possibly bimodal PDF in C_p . The center of the joint PDF at $C_p \approx 55\ \text{fF}$ and $R_s \approx 10\ \Omega$ is physically reasonable; a parallel plate capacitance model would predict a slightly higher value of $C_p = 65\ \text{fF}$ [15], but this is within the high-confidence range. Most importantly, the shape of the PDF should indicate to the user that the experimental design, *i.e.* fitting all parameters in the model in Fig. 1 without applying any constraints on those parameters, will not yield useful results for the diode and pad capacitances.

The NLLS results, shown in red and green, do not convey this information. Figures 4(b) and 4(c) show the two fits and the raw data, *i.e.* the measured complex reflection coefficient from 1 GHz to 40 GHz, on the Smith chart. Both fits appear to be excellent even though they are contradictory. The first NLLS local minimum may be useful. It gives a low maximum-likelihood value for the diode capacitance, but the PDF assigns a high degree of confidence to more physically reasonable values. The second NLLS local minimum, on the other hand, is misleading. It assigns a very low probability to what is probably the true capacitance, and there are no indicators that the fit is inaccurate. It is worth noting that the mean-squared error (the NLLS cost function) has many local minima along the curve that the MCMC PDF follows. Most of these minima have mean-squared error around -40 dB, which is approximately the minimum expected from the known measurement error. Thus using a more robust optimization algorithm to find the global minimum will not necessarily yield the true parameter values, it will more likely yield the values that are most consistent with the particular error present when the data was taken. In this case, the global minimum, or the least-squares fit, occurs at a pad capacitance around 4 fF, which is inconsistent with the value measured on external test structures (15 fF). Finding the global optimum also does not result in any qualitative conclusions about the problems inherent in this experimental design. Thus though useful results can be obtained with nonlinear least-squares fitting under certain conditions, it is far less robust and informative than MCMC. It should be emphasized that neither curve-fitting method produces the “true” diode capacitance with this experimental setup because the measurement data does not contain any information regarding how the total capacitance is split between the diode and probe pads. The primary benefits of MCMC in this case are that (1) the results make it clear that the confidence interval associated with each capacitance value is very large and (2) the PDFs of other parameters, most notably R_s , are appropriately affected.

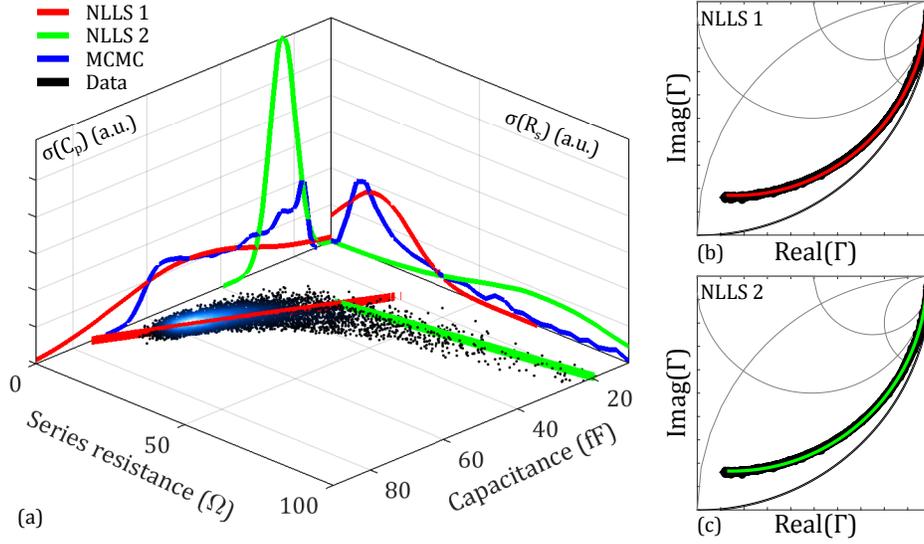


Fig. 4. (a) Joint probability density functions of series resistance and parallel capacitance when all parameters are fit simultaneously to the response of a single device. The joint PDF is shown on the bottom plane, with the MCMC result shown as a scatter plot and the NLLS results shown as contours. The back two planes show 1D PDFs of the parameters separately. (b, c) Original data and both nonlinear least squares maximum likelihood fits on the Smith chart.

4. Describing a group of devices

In order to build a model library for a foundry, circuit parameters should be calculated for both individual devices and groups of devices. Within-die, within-wafer, and between-wafer variation are required for yield estimation, and there are a number of methods of describing this variation and using it for robust circuit design [16, 17]. Within-die variation is of particular interest because it is an important parameter in predicting the performance of analog circuits containing multiple devices (e.g. balanced photodiodes). In both electronics and photonics, much attention has been paid to the spatial correlation between parameters [16, 18, 19] for this reason. Quantifying the spatial correlation necessarily involves small datasets, especially when the device footprint is large relative to the correlation length, and is thus prone to uncertainty. Monte Carlo techniques are useful in this area because they can represent complex probability densities while avoiding the overfitting problems that analytical methods of representing such functions, such as Gaussian mixture models, are prone to [14]. In this section, we will describe using MCMC for multiple fits in order to represent one such PDF.

Fitting multiple devices simultaneously has additional advantages. First, it is possible to resolve the ambiguity in how the total measured capacitance is split between C_p and C_{pad} discussed in Section 3.4 by fitting devices with fixed width and varying length. Since the pad capacitance is the same for all the devices but the diode capacitance is not, this allows the fit to distinguish between the two. We fit twelve $4\ \mu\text{m}$ -wide devices in total: two each of 10, 15, 30, 40, and $70\ \mu\text{m}$ long devices plus one 50 and one $90\ \mu\text{m}$ long device. It should be noted that the ambiguity can also be resolved by measuring the pad capacitance on a separate test structure and using the result of this measurement as a prior on the pad capacitance during fitting. Second, it is possible to assess the results using cross-validation predictive densities [20]. These show that

using the circuit drawn in Fig. 1 in conjunction with the scaling-based approach to resolving the $C_p - C_{pad}$ ambiguity yields reliable results. They also show that the description of within-wafer variation generated by MCMC is more accurate than the Gaussian (least-squares) assumption.

4.1. Gibbs sampling

Gibbs sampling is an implementation of the Metropolis-Hastings algorithm that increases the acceptance rate of proposed samples. Instead of generating a proposal point with new values for each parameter to be fit, only one parameter or a subset of parameters is changed at a time. The proposal point w is drawn from the distribution $\sigma(w|x_n, y_n, \dots)$, where w, x, y now denote distinct parameters and x_n and y_n are the values from the previous iteration. The proposal point is then accepted at all iterations. The Gibbs algorithm is most computationally efficient when the conditional distribution $\sigma(w|x_n, y_n, \dots)$ can be written in analytical form and can be sampled from using standard random number generation algorithms.

To fit a PDF for each circuit parameter value within a die, we separate the parameter set into two groups: the global mean and covariance of all the parameters, and the per-detector parameter values, which each have an associated PDF related to the uncertainty of the measurement. The Gibbs algorithm is used to sequentially sample from these two distributions: the global mean and covariance conditioned on the individual values, followed by the individual values conditioned on the global mean and variance. The global mean is written in vector form as

$$\bar{\mathbf{m}} = \begin{bmatrix} \overline{C_p}/C_{p0} \\ \overline{R_p}/R_{p0} \\ \overline{R_s}/R_{s0} \\ \overline{C_{pad}}/C_{pad0} \\ \overline{\Delta\ell}/\Delta\ell_0 \end{bmatrix} \quad (14)$$

where components with a subscript zero are normalization constants. The vector $\bar{\mathbf{m}}$ is modeled as Gaussian-distributed with a covariance \mathbf{C} :

$$\bar{\mathbf{m}} \sim \mathcal{N}(\overline{\mathbf{m}}_{\text{true}}, \mathbf{C}). \quad (15)$$

To sample from this distribution conditioned on the individual measurements, both the mean of the Gaussian distribution and the covariance need to be updated at each step. The covariance \mathbf{C} is modeled as inverse Wishart distributed. The inverse Wishart distribution is parameterized by an expected covariance matrix Ξ and a number of degrees of freedom ν . It can be shown that for the conditional distribution of the covariance, both of these parameters increase at each step in the Markov chain [12]. The number of degrees of freedom ν is updated according to

$$\nu_{n+1} = \nu_n + N_{PD} \quad (16)$$

where N_{PD} is the number of photodiodes, while the expected covariance matrix is updated according to

$$\Xi_{n+1} = \Xi_n + \frac{1}{2} \sum_{i=1}^{N_{PD}} (\mathbf{m}_i - \bar{\mathbf{m}})^2 \quad (17)$$

where \mathbf{m}_i is a vector of parameters associated with an individual photodiode, and $\bar{\mathbf{m}}$ is the mean of the device-specific parameter vectors $\bar{\mathbf{m}}_i$. The global covariance and mean used in the next iteration are then drawn from

$$\bar{\mathbf{m}}_{n+1} | \mathbf{m}_{i,n} \sim \mathcal{N}(\bar{\mathbf{m}}, \mathbf{C}_{n+1}). \quad (18)$$

and

$$\mathbf{C}_{n+1} | \mathbf{m}_{i,n} \sim \mathcal{W}^{-1}(\Xi_{n+1}, \nu_{n+1}) \quad (19)$$

using standard random number generation algorithms.

The Metropolis-Hastings algorithm is used to generate a sample from the second distribution (values for individual photodiodes conditioned on a global mean and covariance). The target distribution in this case is as written in Eq. (8), where the global mean and variance are incorporated in the prior $\rho_M(\mathbf{m})$ to form the conditional. The algorithm is summarized in Table 4.1.

Table 1. Pseudo-code for Gibbs sampler.

Algorithm 1: Gibbs sampler

Initialize global PDFs for first iteration

$\mathbf{v} = N_{PD} + 1$
 $\Xi = \Xi_{init}$
 $\mathbf{m}0 = \overline{\mathbf{m}}_{init}$

Main loop

for $n=1, 2, \dots, N_{iterations}$

Sample from global PDF

$\mathbf{C} = iwishrnd(\Xi, \mathbf{v})$
 $\mathbf{m}bar[n] = mvnrnd(\mathbf{m}0, \mathbf{C})$

Sample from device-specific PDFs

for $i=1, 2, \dots, N_{PD}$

Initialize - this value will eventually be discarded

$\mathbf{m}temp[1] = mvnrnd(\mathbf{m}0, \mathbf{C})$

Inner Metropolis-Hastings loop

for $j=2, 3, \dots, N_{iterations, innerMH}$

Generate proposal point from q

$\mathbf{w} = random(q, \mathbf{m}temp[j-1])$

Calculate acceptance probability

$$P_{accept} = \frac{\rho(d|\mathbf{w}) * \rho_M(\mathbf{w}) * q(\mathbf{m}temp[j-1]|\mathbf{w})}{\rho(d|\mathbf{m}temp[j-1]) * \rho_M(\mathbf{m}temp[j-1]) * q(\mathbf{w}|\mathbf{m}temp[j-1])}$$

Accept or reject

$r = rand(1)$

if $r \geq P_{accept}$

$\mathbf{m}temp[j] = \mathbf{w}$

else

$\mathbf{m}temp[j] = \mathbf{m}temp[j-1]$

end

end

Final point is a sample from the distribution

$\mathbf{m}i[j] = \mathbf{m}temp[end]$

end

Update global PDF parameters

$\mathbf{v} = \mathbf{v} + N_{PD}$
 $\Xi = \Xi + cov(\mathbf{m}i)$
 $\mathbf{m}0 = mean(\mathbf{m}i)$

end

Analysis performed on mbar

Best Gaussian fit given by $\overline{\mathbf{m}}_0$ and $iwishrnd(\Xi, \mathbf{v})$

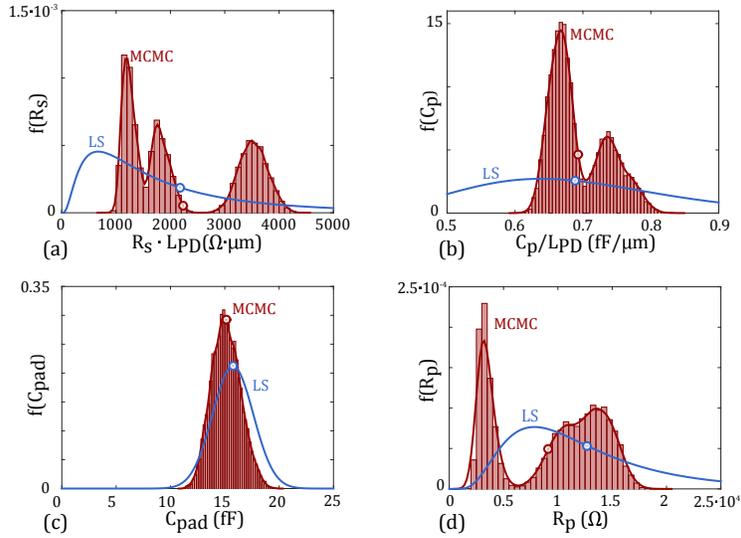


Fig. 5. Probability density functions for (a) series resistance · unit length, (b) diode capacitance per unit length, (c) pad capacitance, and (d) parallel resistance calculated using least-squares fitting (LS) and Markov chain Monte Carlo (MCMC). The means of each distribution are indicated with circular markers.

4.2. Within-die result

Figures 5(a)-5(d) show two estimates of PDFs of equivalent circuit parameters calculated for the group of twelve diodes using least squares regression and Markov chain Monte Carlo. The least-squares fits were generated analogously to the method discussed in Section 3.3, where the parameters were scaled by the device length as appropriate and the mean-squared error was minimized over the entire dataset. All parameters shown in Fig. 1 except L_{pad} (which was fixed at 50 pH) were fit, so these PDFs are defined in a five-dimensional space; the one-dimensional curves shown are projections onto the relevant axes. In order to enforce a positivity constraint for nonlinear least-squares, we calculate the variances of $\ln(R)$ and $\ln(C)$, so that the resulting PDF is log-normal. The output of the Gibbs sampler, used in the MCMC fit, is a set of samples from which the PDF is calculated numerically: both the normalized histogram and kernel smoothing density estimate are shown in the figure. Both NLLS and MCMC agree in mean parameter values as expected, but differ in shape. The tri-modal behavior evident in the MCMC fit could be an accurate description of the process variation, but it could also be due to a missing effect in the equivalent circuit, nonlinear scaling of parameters with diode length, or an underestimate of the measurement error that can result in overly narrow PDFs. Cross-validation assessment can shed some light on whether this is the result of error underestimation, but more data across a wafer would be necessary to conclusively determine the most likely cause of this behavior.

4.3. Model assessment and comparison

As Fig. 5 indicates, very different PDFs can be extracted from the same data set. It is then necessary to determine which description is more accurate, and whether or not it is useful. Cross-validation predictive densities can be used to answer both of these questions. One part of the data set, in this case, the microwave reflection data from a single device, is designated as the validation data (y_i), while the rest of the data set (Y_i) is used as a predictor of this data.

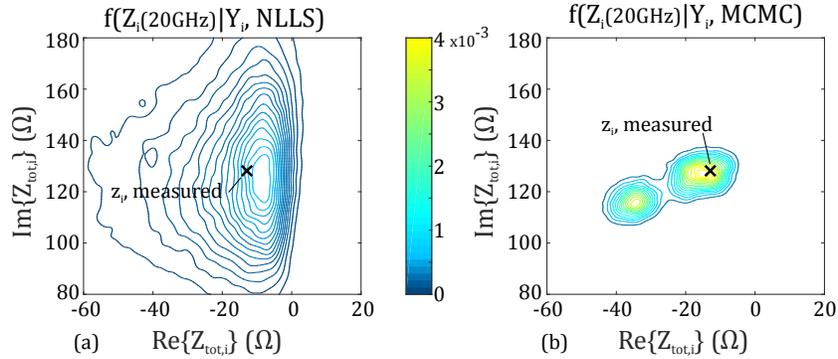


Fig. 6. Cross-validation predictive densities of complex photodiode impedance at 20 GHz calculated by (a) least squares and (b) Markov chain Monte Carlo. The measured value (which is the same in both figures) is also shown.

Using the data from all diodes except diode i , PDFs for the parameters R_s , R_p , C_p , C_{pad} , and $\Delta\ell$ are constructed. A new, predictive, PDF of the complex impedance of diode i as a function of frequency ($f(Z_{tot}(\omega)|Y_i)$) is then calculated from the parameter PDFs. Two examples of this cross-predictive PDF, calculated using NLLS and MCMC, respectively, are shown as contour plots in Figs. 6(a) and 6(b). Calculating the PDF of the complex impedance, rather than reflection coefficient, allows easier interpretation of distances in these plots. As in Fig. 5, both methods predict similar maximum-likelihood values for the complex impedance, but differ in shape.

Model assessment can be performed by defining an application-specific accuracy requirement and analyzing the predictive power of the model with respect to this metric [20]. For example, if the end goal of modeling is to design an impedance matching circuit at a particular frequency, the model could be assessed by counting the number of times the measured impedance (shown as a black ‘x’ in Fig. 6) fell within a certain confidence interval assigned by the cross-validation predictive PDF. In this particular case, the measured impedance nearly always falls within the 95% confidence interval given by both the MCMC and NLLS cross-validation PDFs. This implies that the circuit model shown in Fig. 1 is sufficiently accurate.

The two methods of PDF extraction can be compared by calculating the pseudo-Bayes factor [12]. First, the cross-validation predictive PDFs are evaluated at the observed data points. Then the product of all such validation probabilities is calculated; this product will be proportional to the probability the model assigns to the data. The ratio of two such products, calculated using two different models is the pseudo-Bayes factor, and it compares the relative accuracies. Evaluating $f(Z_{tot}(\omega)|Y_i)$ at an evenly spaced set of frequencies from 5 GHz to 40 GHz, we find that the model generated using MCMC has greater predictive power. This indicates that the long tails for the NLLS fit in Fig. 5 are the result of the Gaussian assumption rather than an accurate description of the process variation (or, alternatively, that the tri-modal behavior is not due to underestimating the error). Cross-predictive likelihoods can also be used to compare different circuit models and to perform standard tests of statistical significance.

5. Conclusion

The nonlinear relationship between circuit-equivalent parameters and the measured microwave reflection coefficient can cause least-squares fits for these parameters to S_{11} data to give incorrect results. Using Markov chain Monte Carlo techniques in this circumstance results in relatively accurate quantification of the uncertainty inherent in the fitting process. Multiple

measurements of differently sized devices can be used to avoid this problem, and in this case MCMC techniques yield more accurate and precise estimates of the distributions of parameter values than least squares. Within the context of a controlled and quantified fabrication process by a foundry, this can lead to more accurate performance and yield predictions for photonic integrated circuits used in applications such as microwave sensing [21] and photonic RF frequency generation [22, 23].

Acknowledgment

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 600207 and the Danish Council for Independent Research agreement 4093-00168. This project has also received funding from the Villum Foundation Young Investigator Program.