A Novel Methodology to Create Generative Statistical Models of Interconnects

Simon De Ridder*, Paolo Manfredi*, Jan De Geest†, Tom Dhaene*, Daniël De Zutter*, Dries Vande Ginste*

*IBCN/Electromagnetics Group, Department of Information Technology, Ghent University/iMinds
Technologiepark-Zwijnaarde 15, 9052 Gent, Belgium
E-mail: `Simon.DeRidder@UGent.be`
†Amphenol-FCI

Abstract—**This paper addresses the problem of constructing a generative statistical model for an interconnect starting from a limited set of S-parameter samples, which are obtained by simulating or measuring the interconnect for a few random realizations of its stochastic physical properties. These original samples are first converted into a pole-residue representation with common poles. The corresponding residues are modeled as a correlated stochastic process by means of principal component analysis and kernel density estimation. The obtained model allows generating new samples with similar statistics as the original data. A passivity check is performed over the generated samples to retain only passive data. The proposed approach is applied to a representative coupled microstrip line example.**

Index Terms—**Statistical model, interconnect, Principal Component Analysis, Kernel Density Estimation**

I. Introduction

The ever-growing impact of manufacturing variability on the performance of electronic devices is prompting an increasing interest in stochastic modeling techniques for signal integrity investigations. Whereas very efficient modeling techniques were proposed based on the framework of generalized polynomial chaos (gPC) [1, 2], these approaches are hardly applicable when the structure affected by variability is characterized by means of measurements or by simulations involving hundreds of uncertain parameters, simply because the data required to construct the gPC models cannot be generated in a time- and cost-effective way.

An alternative approach was proposed in [3], where a generative statistical model was derived that is capable of producing new random samples by suitably interpolating measured S-parameter data in a statistically valid manner. However, this method is not applicable to multiport passive structures, since it models each S-parameter independently, as such not preserving the existing relationship between the S-parameters and (most likely) yielding non-passive samples.

In this paper, a novel strategy for multiports is put forward. A small dataset of "training samples" is obtained by means of Monte Carlo simulation or measurements. By "sample" we mean a S-parameter frequency sweep and/or its corresponding rational macromodel (see section II). These S-parameter samples are first converted into a pole-residue representation via Vector Fitting (VF) [4, 5]. This allows operating with a finite set of frequency-independent variables. For convenience, VF with fixed common poles is adopted, to avoid the problem of

dealing with possibly unstable samples. Second, the stochastic residues are statistically modeled by means of Principal Component Analysis (PCA) and Kernel Density Estimation (KDE). These models allow generating new S-parameter samples with similar statistics and interrelationships as the original ones. Non-passive samples are readily discarded in a post-processing phase.

The approach is applied to the modeling of a coupled microstrip line, for which it is shown that the newly generated S-parameter samples closely resemble the simulated ones. As expected, the agreement further improves as the variability is reduced.

II. Proposed Methodology

Consider a multiconductor transmission line (MTL) with n signal conductors and at least one reference conductor, prone to variability. By means of measurements or simulations, a limited set of K training samples is available in the form of S-parameters. Hence, each such training sample is characterized by a $2n \times 2n$ matrix $\overline{\overline{S}}_k(s), k = 1, \ldots, K$, where s denotes the complex frequency. In a first modeling step, each sample of the MTL is fitted by a rational model with N poles a_i and N residue matrices $\overline{\overline{R}}_{k,i}$, as follows:

$$\overline{\overline{S}}_k \approx \sum_{i=1}^{N} \frac{\overline{\overline{R}}_{k,i}}{s - a_i}, \qquad k = 1, \ldots, K. \qquad (1)$$

This fitting is done by means of Vector Fitting [4, 5]. The poles and residues are either real or come in complex conjugate pairs. As seen from (1), all training samples are fitted with a common set of poles $\{a_i\}_{i=1}^{N}$, while the residues pertain to a specific sample k. This is achieved by first fitting all training samples at once to obtain the common poles, and then, by identifying the residues of each training sample separately. Because the VF algorithm ensures stability by keeping the poles in the left half of the complex plane (i.e., $Re\{a_i\} < 0$), any additional sample generated with these poles will also be stable.

Owing to the reciprocity of the MTL, the K S-parameter matrices are symmetrical. Hence, in total $KN2n(2n + 1)/2$ elements of the residue matrices need to be modeled. Given this high dimensionality, the modeling problem may become intractable. Therefore, a Principal Component Analysis

(PCA) [6] is applied to these $KN2n(2n + 1)/2$ random variables (after rescaling to unit variance) to reduce the dimensionality. At the same time, this technique also produces linearly decorrelated variables. The samples are then again rescaled by the square root of their latent variance to obtain a more spherical distribution.

Whereas the VF and PCA steps help to reduce the problem dimensionality, the actual modeling is now done by means of a Kernel Density Estimation (KDE) [7, 8]. This technique approximates the probability density function (PDF) of the residues — which are now already projected onto the reduced space — by a sum of kernels centered on each training point. In this case, Gaussian kernels were used, and their covariances were estimated using the algorithm detailed in [8]. As the distribution, in general, features nonlinear correlations, KDE tends to provide a better estimate than a parametric fit of any elliptical distribution would.

Generating a new sample of the MTL is now performed as follows. First, one training point is selected at random. Next, from the kernel centered on that point, a new point is sampled. Then the inverse PCA transformation is applied to project this new point back to a new set of residues. Finally, these residues are combined with the common poles to yield a new sample of the S-parameters via equation (1).

If the newly generated sample is not passive, it is simply rejected, and a new one is generated. Enforcing passivity as in [9] is also a possibility, but may cause a bias in the generated sample population and it is therefore undesired.

III. APPLICATION EXAMPLES AND DISCUSSION

To demonstrate its validity and appositeness, the above method is applied to the case of two coupled symmetric microstrip lines. In the nominal design, these lines have a length of $10\,\text{cm}$, a trace width of $50\,\mu\text{m}$, a gap of $40\,\mu\text{m}$ between them, and they reside on a dielectric substrate of thickness $60\,\mu\text{m}$ and with relative permittivity $\epsilon_r = 3.7$ and $\tan\delta = 0.02$. A dataset was constructed using Synopsis' HSPICE, where the distance between the lines, the substrate thickness and the substrate's ϵ_r were varied according to a normal distribution. For the first dataset, a low standard deviation of $1\,\%$ was assumed, while for the second a high standard deviation of $10\,\%$ was applied. Each dataset contains 1000 simulated samples, 50 of which are used for training, the other 950 for validation.

As a robust measure of accuracy, the sum of the areas between the cumulative distribution function (CDF) of the generated samples and the CDF of the validation samples is calculated for each frequency point. This allows comparison to other generative models, and here in particular, a multivariate Gaussian model of the residues is leveraged as a reference. By applying the above methods with $K = 50$ training samples and $N = 20$ terms in the VF expansion (1), the S-parameter results shown in Fig. 1 were obtained for the $1\,\%$ standard deviation dataset. The blue lines correspond to 500 out of 1000 generated new samples, while the green lines show 475 out of 950 validation samples. The red lines are superimposed
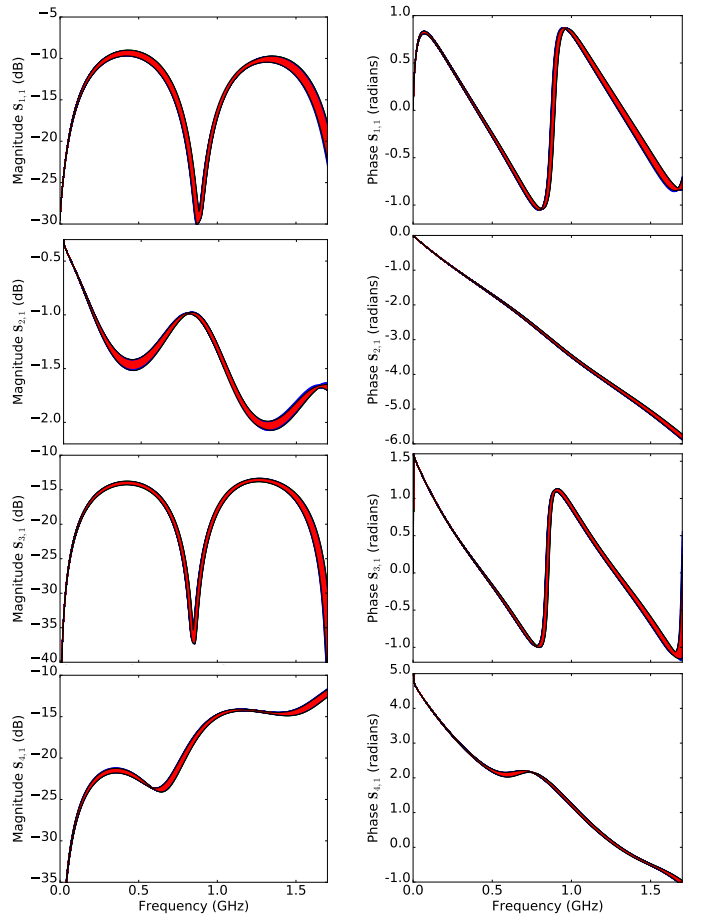

Figure 1. S-parameters of 500 of the newly generated samples (blue lines), and 475 simulated validation samples (green lines) for the case of $1\,\%$ standard deviation. The 50 training samples (50) are superimposed in red. The black lines indicate the minimum and maximum of the validation samples.

and denote the 50 training samples. The black lines indicate the minimum and the maximum of the validation samples. at $1\,\%$ standard deviation it is difficult to see the difference between the various lines in Fig. 1. The reader might first want to look at Fig. 2 for a better understanding. It is noticed that all newly generated samples reside within these two black lines (bounds) and that there is a good correspondence between the validation samples and the newly generated samples. The above proposed accuracy measure for these samples amounts to 0.05, while for those generated by the reference multivariate Gaussian model it is 0.04. This shows that for small variations of the varied dimensions, the distribution of the residues is sufficiently Gaussian. Consequently, both our novel advocated technique and the multivariate Gaussian approach provide a very good model for the actual distribution of the S-parameters. The result of applying the proposed method to the $10\,\%$ standard deviation dataset yields the results shown in Fig. 2, using the same colors for the samples as in Fig. 1. Due to the increased variance, the generated samples do no longer exactly follow the distribution of the simulated samples. In the neighborhood of peaks, there is some overshoot of the
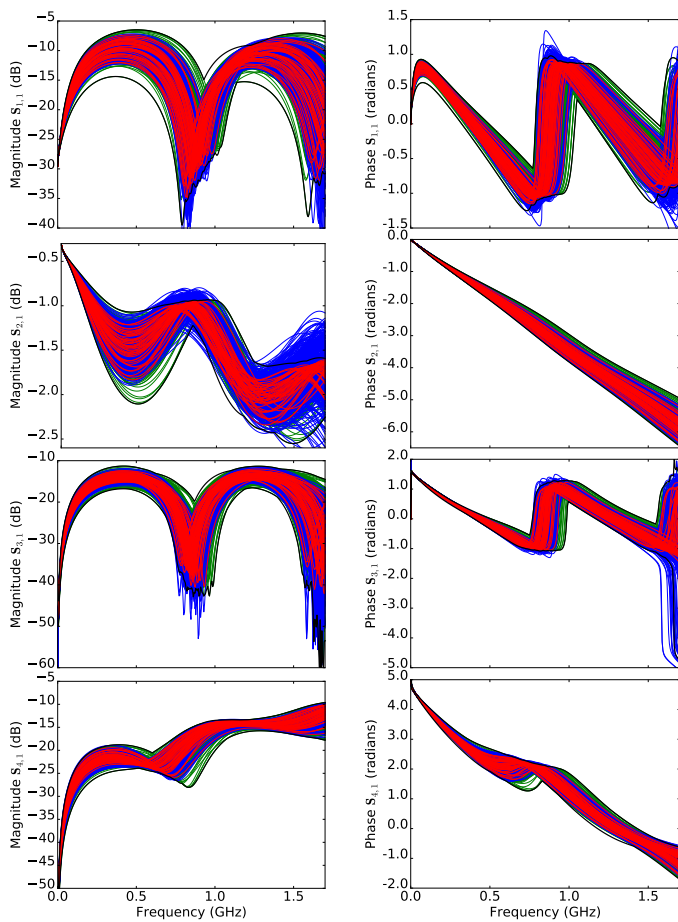
Figure 2. S-parameters of 500 of the newly generated samples (blue lines), and 475 simulated validation samples (green lines) for the case of 10 % standard deviation. The 50 training samples (50) are superimposed in red. The black lines indicate the minimum and maximum of the validation samples.

generated samples. This is caused by the limited number of training samples K combined with a limited set of terms N in the rational expansion, leading to a non-perfect approximation. Nevertheless, as can be seen from the figure, there is still a very good correspondence between the validation samples and the newly generated samples. Moreover, the accuracy of the generated samples as indicated by the aforementioned measure, is 2.3, while for the multivariate Gaussian model it is 3.3. Clearly, for these larger variations, the distribution of the residues is no longer well approximated by a Gaussian, necessitating our proposed method.

IV. CONCLUSIONS

This paper proposes a novel technique to generate random S-parameter samples starting from a small set of simulated data. The technique first converts the original S-parameters into a pole-residue representation with common poles. The random residues are then modeled as a stochastic process by means of PCA and KDE. The obtained statistical model allows generating new samples that closely match the statistical features of the original data. A passivity selection is performed in a post-processing step to ensure that only passive data

is retained. The feasibility and applicability is demonstrated by modeling a pair of coupled microstrip lines with random geometrical and material properties. During the design of novel high-speed links, the efficient generation of a statistically correct population of samples of the interconnect structures prone to variability, based on a limited set of training samples, is of high importance and successfully achieved with our novel method.

REFERENCES

[1] P. Manfredi, D. Vande Ginste, D. De Zutter, and F. G. Canavero, "Uncertainty assessment of lossy and dispersive lines in SPICE-type environments," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 3, no. 7, pp. 1252–1258, July 2013.

[2] M. R. Rufuie, E. Gad, M. Nakhla, and R. Achar, "Generalized Hermite polynomial chaos for variability analysis of macromodels embedded in nonlinear circuits," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 4, no. 4, pp. 673–684, April 2014.

[3] L. L. Campbell and J. E. Purviance, "Interpolation modeling of S-parameter data bases for use in Monte Carlo simulations," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 4, no. 3, pp. 282–296, July 1994.

[4] B. Gustavsen and A. Semlyen, "Rational approximation of frequency domain responses by vector fitting," *IEEE Transactions on Power Delivery*, vol. 14, no. 3, pp. 1052–1061, July 1999.

[5] D. Deschrijver, M. Mrozowski, T. Dhaene, and D. De Zutter, "Macromodeling of multiport systems using a fast implementation of the vector fitting method," *IEEE Microwave and Wireless Components Letters*, vol. 18, no. 6, pp. 383–385, June 2008.

[6] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441, 1933.

[7] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Statist.*, vol. 27, no. 3, pp. 832–837, 09 1956. [Online]. Available: http://dx.doi.org/10.1214/aoms/1177728190

[8] M. Kristan, A. Leonardis, and D. Skoaj, "Multivariate online kernel density estimation with gaussian kernels," *Pattern Recognition*, vol. 44, no. 1011, pp. 2630 – 2642, 2011, semi-Supervised Learning for Visual Content Analysis and Understanding. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2011.03.019

[9] T. Dhaene, D. Deschrijver, and N. Stevens, "Efficient algorithm for passivity enforcement of S-parameter-based macromodels," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 2, pp. 415–420, Feb. 2009.