Principal Component Analysis Applied in Modeling of Stochastic Electromagnetic Field Propagation

Michael Haider*, Johannes A. Russer*, Andrey Baev†, Yury Kuznetsov†, and Peter Russer*
*Institute for Nanoelectronics, Technical University of Munich, Germany
†Theoretical Radio Engineering Department, Moscow Aviation Institute, Russia
E-Mail: michael.haider@tum.de

***Abstract*—Stochastic electromagnetic (EM) fields can be characterized by auto- and cross correlation spectra. The amount of data, describing stochastic EM fields can easily become burdensome. Therefore, techniques for reducing the dimensionality of data characterizing stochastic EM field problems are important, since correlation matrices describing stochastic EM field correlations scale quadratically with the number of observation points. We present a method for order reduction of stochastic electromagnetic field description, based on principal component analysis (PCA). Furthermore, we investigate how the number of principal components (PCs) relates to the number of independent sources and the transversal coherence as the EM field is propagated using an auto- and cross correlation based numerical propagator.**

***Index Terms*—Stochastic electromagnetic fields, electromagnetic interference, principal component analysis.**

I. Introduction

Stochastic electromagnetic fields are a key concern in the assessment of electromagnetic interference (EMI) and signal integrity (SI). For a large set of problems in electromagnetics, one cannot specify amplitude and phase values for the EM fields. This is either due to a lack of knowledge of the field sources or due to the fact that the investigated electromagnetic field is of intrinsic stochastic origin. Radiated EMI frequently originates from non-deterministic stochastic processes.

A stochastic process is determined by a certain probability density function (PDF) which may vary over time. In the following we consider stationary Gaussian processes, which means that the PDF is constant in time and assumes a Gaussian shape. A Gaussian process is uniquely determined by specifying the first- and second-order statistical moments [1], [2]. Assuming Gaussian statistics is justified due to the central limit theorem if the number of sources is high enough. However, non-stationary stochastic electromagnetic fields are also frequently encountered. Radiation from integrated circuits (ICs), for example, can be considered a stochastic process for the purpose of EMI investigation, since the bit sequence will exhibit a stochastic nature. Due to the clocking, mean values and correlation functions are periodic in time and the process can be described as a cyclostationary stochastic process. Also for such processes, characterizations based on auto- and cross correlations can be given [3], [4], and principal component analysis (PCA) can be applied to the obtained correlation matrices.

For stochastic EM fields, PCA can reduce data storage and handling requirements significantly as discussed in the following. Furthermore, the number of dominant PCs provides an estimate for the number of stochastically independent EM sources involved. PCA as a technique for data reduction and source identification for stochastic EM fields was already treated in [5]–[7]. An efficient algorithm for data reduction based on PCA is presented in [8]. PCA can also be employed for source localization and imaging of noisy EM fields [9]–[12]. Evolution of the transverse correlation in noisy electromagnetic fields has been considered also in [13]. As we will see in this work, the number of PCs that have to be retained for accurate EM field characterization, strongly depends on geometrical considerations of source locations and field sampling planes.

In the following we investigate how the number of PCs scales with the propagation distance, as we propagate correlation matrices in space numerically for different frequencies. We consider a simplified problem, consisting of a two-dimensional array of small dipoles on a source plane. The radiated field is observed at the same $x-y$ locations on planes with different distances z_i from the source plane.

II. Principal Component Analysis

PCA is a well established technique in statistics. It is used to identify the directions of greatest variance in multidimensional data sets, consisting of a very large number of interrelated variables. PCA, originating in the field of psychology and education [14], [15], has found wide spread application in multivariate statistics in recent years. The governing idea behind PCA is to determine linear functionals $\langle \boldsymbol{a}_i, \bullet \rangle$, which maximize the variance of a multivariate random variable $\boldsymbol{X}$. Such a functional is given by the first few eigenvectors of the covariance matrix $\boldsymbol{C}_{\text{cov}}$ of $\boldsymbol{X}$ [16].

III. Evolution of Correlation Information for the Propagated Field

We consider a setup given by a two-dimensional array consisting of $p = m' \times n'$ Hertzian dipoles of length l, oriented in x-direction. The currents $\{I_j\}_{j=1}^{p}$ in the dipoles are governed by Gaussian random processes with zero mean and

Cite as: Michael Haider, Johannes A. Russer, Andrey Baev, Yury Kuznetsov, and Peter Russer, ”Principal Component Analysis Applied in Modeling of Stochastic Electromagnetic Field Propagation,” in Proceedings of the 47th European Microwave Conference (EuMC) 2017 Nuremberg, Germany, Oct. 8-13, 2017, pp. 351–354

are described by the correlation matrix $\boldsymbol{C}_I$ [2]. The location $\left(x'_j, y'_j\right)$ of the j-th dipole on the source-plane $z' = 0$ is

$$x'_j = x'_0 + \left\lfloor \frac{j-1}{m'} \right\rfloor \Delta x' , \tag{1}$$

$$y'_j = y'_0 + \left[(j-1) \mod n'\right] \Delta y' , \tag{2}$$

with $j \in \{1, \ldots, p\}$. Here, $\lfloor \cdot \rfloor$ denotes the floor operation, i.e. the next smaller integer number, and a mod b is the modulo division of a and b. In order to investigate the propagation of stochastic electromagnetic fields, we define a sampling grid, consisting of $q = m \times n$ observations on a plane at a distance of $z = h$ from the source plane. The spatial location (x_j, y_j) of the j-th observation point is analogous to (1) and (2), where the location of the initial point (x_0, y_0) and the horizontal and vertical grid-spacing Δx and Δy may differ from the source-grid parameters.

After choosing a finite set of source and observation points, the method of moments (MoM) can be applied to transfer the field problem to a network problem [17]. The mapping information obtained in form of the moment matrix also provides the information how to transform the correlation information describing stationary stochastic EM fields. For our considerations, we use the free-space dyadic Green's function, also accounting for the near-field contributions, together with point-matching to obtain a generalized impedance matrix $\boldsymbol{Z}(\omega)$ relating the vector of generalized source currents $\boldsymbol{I}_T(\omega)$ to a vector of generalized voltages $\boldsymbol{V}_T(\omega)$ on the observation plane [2]. The subscript $_T$ denotes the time windowed signal for which a spectrum can be defined. The (m, n)-th element Z_{mn} of the generalized impedance matrix $\boldsymbol{Z}$, relating the n-th source-current to the m-th observation, is given by

$$\begin{aligned} Z_{mn}(h, k) &= \frac{l Z_0}{4\pi k} \mathrm{e}^{-jk\sqrt{(x_m - x'_n)^2 + (y_m - y'_n)^2 + h^2}} \\ &\times \Big[g_1\left(x_m - x'_n, y_m - y'_n, h, k\right)\left(x_m - x'_n\right)^2 \\ &+ g_2\left(x_m - x'_n, y_m - y'_n, h, k\right) \Big] , \end{aligned} \tag{3}$$

where Z_0 is the free space wave impedance and $k = 2\pi f / c_0$ is the wave number, c_0 is the speed of light in vacuum. In (3), we use g_1 and g_2 given by

$$g_1(x, y, z, k) = -\frac{3j}{|\boldsymbol{x}|^5} - \frac{3k}{|\boldsymbol{x}|^4} + \frac{jk^2}{|\boldsymbol{x}|^3} , \tag{4}$$

$$g_2(x, y, z, k) = -\frac{jk^2}{|\boldsymbol{x}|} + \frac{k}{|\boldsymbol{x}|^2} + \frac{j}{|\boldsymbol{x}|^3} , \tag{5}$$

with $|\boldsymbol{x}| = \sqrt{x^2 + y^2 + z^2}$. In the following, let us consider uncorrelated currents at a single frequency with unit variance. Using the generalized impedance matrix $\boldsymbol{Z}$ assembled from (3), we can propagate the electric field generated by the source dipole currents to observation-planes at different heights h_i. The observations are related to the sources by

$$\boldsymbol{V}_T^{h_i}(\omega) = \boldsymbol{Z}(h_i, \omega / c_0)\, \boldsymbol{I}_T(\omega) . \tag{6}$$

The subscript $_T$ denotes the spectrum of time-windowed signal. Figure 1 shows the amplitudes of the propagated electric field at different heights h_i for a single realization of the stochastic source currents $\boldsymbol{I}_T$. To compute the stochastic field an ensemble average of the propagation of different realizations of $\boldsymbol{I}_T$ has to be formed.

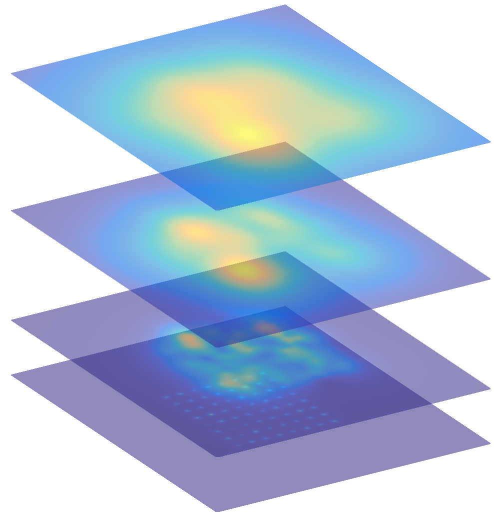

Fig. 1. $|\boldsymbol{E}| / |\boldsymbol{E}_{\max}|$ of the propagated EM field for a single realization of $\boldsymbol{I}_T$, normalized within each observation plane, at heights $h_0 = 0$ mm, $h_1 = 10$ mm, $h_2 = 30$ mm, and $h_3 = 55$ mm.

Since stochastic EM fields with Gaussian probability distribution can be described by second-order statistics, applying auto- and cross correlation spectra, we use a MoM based propagation scheme for correlation matrices using the deterministic impedance matrix assembled from (3). Correlation matrices for generalized voltages $\boldsymbol{V}_T$ and generalized currents $\boldsymbol{I}_T$, as defined in [2], can be obtained by the ensemble averages

$$\boldsymbol{C}_I(\omega) = \lim_{T \to \infty} \frac{1}{2T} \left\langle\!\left\langle \boldsymbol{I}_T(\omega)\, \boldsymbol{I}_T^\dagger(\omega) \right\rangle\!\right\rangle , \tag{7}$$

$$\boldsymbol{C}_V(\omega) = \lim_{T \to \infty} \frac{1}{2T} \left\langle\!\left\langle \boldsymbol{V}_T(\omega)\, \boldsymbol{V}_T^\dagger(\omega) \right\rangle\!\right\rangle , \tag{8}$$

where $\langle\!\langle \cdot \rangle\!\rangle$ denotes the ensemble average. Using (6), (7) and (8), we obtain a propagation rule for the correlation matrices [2], given by

$$\boldsymbol{C}_V^{h_i} = \boldsymbol{Z}(h_i, \omega / c_0)\, \boldsymbol{C}_I(\omega)\, \boldsymbol{Z}^\dagger(h_i, \omega / c_0) . \tag{9}$$

For $\boldsymbol{C}_I(\omega)$ is equal to unity, which is the case for completely uncorrelated source currents I_m, we get

$$\boldsymbol{C}_V^{h_i} = \boldsymbol{Z}(h_i, \omega / c_0)\, \boldsymbol{Z}^\dagger(h_i, \omega / c_0) . \tag{10}$$

Using (3), the elements of the correlation matrix $\boldsymbol{C}_V^{h_i}(\omega)$ can be calculated by

$$C_{V,mn}^{h_i}(\omega) = \sum_{\nu=1}^{N} Z_{m\nu}(h_i, \omega / c_0)\, Z_{\nu n}^*(h_i, \omega / c_0) . \tag{11}$$

A. Required PCs and Energy Considerations

We specify an array of source points, modeled by Hertzian dipoles oriented in x-direction, on an $m' = 8$ by $n' = 8$ grid with a grid point spacing of $\Delta x' = \Delta y' = 1\,\mathrm{cm}$.

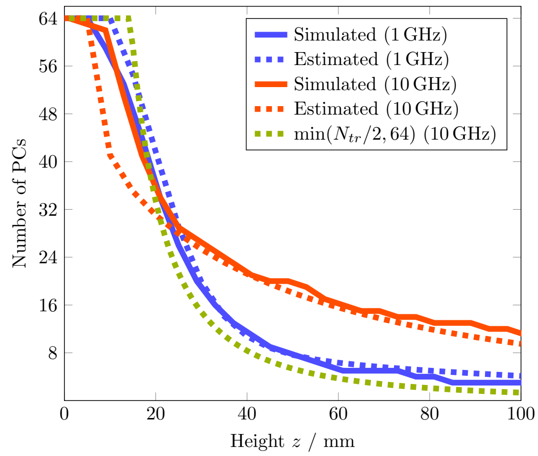

Fig. 2. Number of dominant PCs vs. height of observation plane obtained numerical propagation of the field-field correlations on a constant sized observation grid, by estimation of field energy within the observation window, and by far-field estimate based on transverse mode counting.

With all source dipole currents chosen to be uncorrelated, the correlation matrix describing these sources is a 64×64 matrix of full rank. Hence, we require 64 PCs to account for 100% of the variance. We consider sampling grids at various heights h_i above the source plane where sample the E_x-field. Figure 2 shows the estimated numbers of PCs to retain in order to explain 99% of total variance for each height for source excitations at 1 GHz and 10 GHz. To perform this estimate on the number of PCs, we consider the total EM field energy in each observation plane at $z = h_i$. For this estimation, we numerically propagate the field correlations of the noisy sources and use a finely resolved grid on the observation plane which considerably exceeds the 7×7 cm^2 area used for the sources at $z = h_0$, such that effectively all energy radiated into the observation plane at h_i is also sampled. The spectral energy density (SED) is closely related to the autocorrelation spectrum for each field sampling point on the observation plane. The EM field energy obtained by integrating the energy density over the 7×7 cm^2 area of interest at $z = h_i$ is compared to the total energy on the observation plane $z = h_i$. This ratio between energy on the 7×7 cm^2 area to total energy in the plane, gives a proper estimate for the number of PCs to retain, in order to account for 99% of total variation for each height h_i.

Fig. 2 also shows the actual number of PCs which need to be retained in order to account for 99% of the variance, and hence also for 99% of the SED of the stochastic EM field, obtained by performing PCA on the matrices $\boldsymbol{C}_V^{h_i}$, given by (10). Here, we consider sampling grids of the same size and resolution as the source grid. For terminating the PCA algorithm after a certain percentage of total variation we use the cumulative percentage of total variance (CPTV) criterion from [18].

B. Transverse Coherence

With increasing distance from the source plane, which can be considered an aperture, the number of PCs required to explain the variance on a sampling grid of constant size decreases, while at the same time the spatial angle observed, and hence, the number of transverse modes to be resolved, decreases as well. To estimate the number of PCs required for explaining 99.9% of the variance at sampling grids at a height $z = h_i$ and at height $z = h_0$, we give an estimate on how many transverse modes we can resolve on the given sampling point grid. The EM field is originating from an aperture A_s. This is a worst case estimate, i.e. indicating how many PCs we will need at most to give an accurate description of the correlation matrix. The estimate will be good in the far field and we assume the transverse component of the propagation vector $\mathbf{k}$ to be small in magnitude compared to its overall wave number k_0. The space angle containing one emitted mode is given by $\Omega_c = \lambda_0^2/A_s$, where $\lambda_0 = 2\pi/k_0$ and, for our case, $A_s = 7 \times 7$ cm^2. The number of transverse modes N_{tr} which can be resolved in a distance r from the aperture on an area A_s will be

$$N_{tr} = \frac{A_s^2}{r^2 \lambda_0^2} \,. \tag{12}$$

For our numerical example, where only the E_x-field is sampled, the number of transverse modes to be detected is $N_{tr}/2$. Hence, at a distance r, the number of PCs required for a full description of our correlation matrix is given by $N_{tr}/2$ while at the same time the number of PCs in our estimate cannot exceed the maximum rank of the correlation matrix, which is in our example 64. The function

$$\min(N_{tr}/2, 64) \,, \tag{13}$$

plotted in Fig. 2, shows that it also provides a qualitative good estimate for the maximum number of PCs required, however, with some discrepancy which is not yet explained. At lower frequencies, when the far-field assumption is less justified, the number of PCs required may exceed the number from this estimate based on counting transverse modes.

IV. Efficient Approximation of Measured Data

The amount of data describing stochastic electromagnetic fields obtained either by simulation or by measurement can become burdensome. Besides estimating the number of independent sources, PCA can be used for reducing the amount of data significantly by only storing the most dominant eigenvectors of a given set of correlation matrices. This must be done for each frequency, which makes the PCA itself costly in terms of computation time. Efficient PCA algorithms in terms of power iterations exist [8]. We present results from an actual measurement, obtained by two-probe scanning of a small FPGA board. The tangential magnetic field components were measured on a grid of 19×13 points with two probes simultaneously in time-domain. Determining all correlations requires measurement at about 30,000 point pairs. Subsequently, we calculated all possible auto- and cross correlation functions and obtained the correlation spectra by Fourier transform. Figure 3 shows the cumulative spectral energy density of the field scanned directly above the FPGA board, together with the number

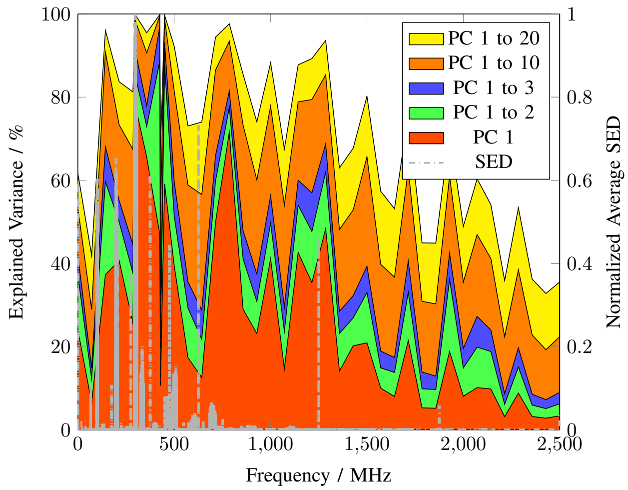

Fig. 3. Number of PCs vs. frequency.

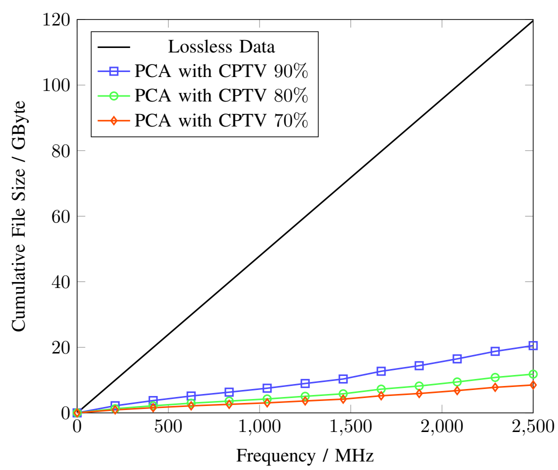

Fig. 4. Cumulative file-size vs. frequency for different CPTV.

of PCs accounting for a certain percentage of total variance. One can see, that for certain frequencies the first PC already accounts for more than 80% of the total variation. By setting a certain threshold for a percentage of variance to be retained, we can adaptively reduce the amount of data to be stored, with full control of the approximation error.

The savings in memory become significant, when we plot the cumulative file-size over frequency, as shown Fig. 4. The total file size was reduced from roughly 120 GB to approximately 20 GB while preserving at least 90% of total energy. The amount of data can be further reduced, allowing for a greater error as a trade off. Thus, the total file size for 80% of total variation is about 11.8 GB and for 70% of total variation we can compress down to 8.5 GB.

V. Conclusion

We investigated how the number of principal components scales for different distances and frequencies. The number of independent sources, that can be resolved at some distance strongly depends on the size of the chosen observation grid. Assuming a constant grid spacing for different distances, the decrease in the number of resolvable stochastic sources can be estimated by the ratio of the energy collected by the observation grid to the total radiated energy. In addition, we used PCA for reducing the complexity of a data set obtained by two-probe scanning of a micro controller board and could achieve substantial data storage savings.

Acknowledgment

This work was supported in part by COST ACTION IC1407, and by the European Unions Horizon 2020 research and innovation programme under grant no. 664828 (NEMF21).

References

[1] J. A. Russer and P. Russer, “An efficient method for computer aided analysis of noisy electromagnetic fields,” in *2011 IEEE MTT-S Int. Microw. Symp. Digest*. IEEE, Jun. 2011, pp. 1–4.

[2] ——, “Modeling of noisy EM field propagation using correlation information,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 1, pp. 76–89, Jan 2015.

[3] J. A. Russer, P. Russer, M. Konovalyuk, A. Gorbunova, A. Baev, and Y. Kuznetsov, “Analysis of cyclostationary stochastic electromagnetic fields,” in *2015 Int. Conf. on Electromagnetics in Advanced Applications (ICEAA)*, Sep 2015, pp. 1452–1455.

[4] ——, “Near-field propagation of cyclostationary stochastic electromagnetic fields,” in *Int. Conf. on Electromagnetics in Advanced Applications (ICEAA), 2015*, Torino, Italy, Sep 7-11 2015, pp. 1456–1459.

[5] L. R. Arnaut and C. S. Obiekezie, “Comparison of complex principal and independent components for quasi-gaussian radiated emissions from printed circuit boards,” *IEEE Transactions on Electromagnetic Compatibility*, vol. 56, no. 6, pp. 1598–1603, Dec 2014.

[6] L. R. Arnaut, C. S. Obiekezie, and D. W. P. Thomas, “Empirical emission eigenmodes of printed circuit boards,” *IEEE Transactions on Electromagnetic Compatibility*, vol. 56, no. 3, pp. 715–725, June 2014.

[7] T. Asenov, J. A. Russer, and P. Russer, “Efficient characterization of stochastic electromagnetic fields using eigenvalue decomposition and principal component analysis methods,” in *2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS)*, Aug 2014, pp. 1–4.

[8] M. Haider and J. A. Russer, “Principal component analysis for efficient characterization of stochastic electromagnetic fields,” *Int. J. of Numerical Modelling: Electronic Networks, Devices and Fields*, p. e2246, 2017.

[9] ——, “Equivalent source localization for stochastic electromagnetic fields,” in *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, Verona, Sep 2017.

[10] Y. Kuznetsov, A. Baev, A. Gorbunova, M. Konovalyuk, D. Thomas, C. Smartt, M. H. Baharuddin, J. A. Russer, and P. Russer, “Localization of the equivalent sources on the PCB surface by using ultra- wideband time domain near-field measurements,” in *Proc. EMC Europe*, Wroclaw, Poland, 2016.

[11] A. Baev, A. Gorbunova, M. Konovalyuk, Y. Kuznetsov, and J. A. Russer, “Planar stochastic sources localization algorithm in EMC problems,” in *Int. Conf. on Electromagnetics in Advanced Applications (ICEAA), 2013*. IEEE, Sep 9-13 2013, pp. 440–443.

[12] J. A. Russer and P. Russer, “Imaging of sources of radiated electromagnetic interference,” *Frequenz*, vol. 65, pp. 261–265, Sep. 2011.

[13] J. A. Russer, G. Gradoni, G. Tanner, S. C. Creagh, D. Thomas, C. Smartt, and P. Russer, “Evolution of transverse correlation in stochastic electromagnetic fields,” in *IEEE Int. Microw. Symp. (IMS)*, May 2015, pp. 1–3.

[14] K. Pearson, “On lines and planes of closest fit to systems of points in space,” *Philosophical Magazine*, vol. 2, no. 11, pp. 559–572, 1901.

[15] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *J. of Educational Psychology*, vol. 24, pp. 417–441, 1933.

[16] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., ser. Springer Series in Statistics. Springer New York, 2002.

[17] J. A. Russer and P. Russer, “Network methods applied to the computation of stochastic electromagnetic fields,” in *2011 Int. Conf. on Electromagnetics in Advanced Applications (ICEAA)*, Sep 2011, pp. 1152–1155.

[18] J. E. Jackson, *A User’s Guide to Principal Components*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, 2004.