Emulation and Inversion of Polarization Mode Dispersion: A Lumped System and Pade Approximation Perspective

Alper Demir, *Member, IEEE*, and Alper T. Erdogan, *Member, IEEE*

***Abstract*—Polarization mode dispersion (PMD) has been a major impediment in achieving higher speeds in optical fiber communication systems. Various approaches and techniques for the basic theoretical analysis, modeling, simulation, and compensation of PMD have been proposed in the literature. In this paper, we present a novel and fresh approach to the modeling, emulation and inversion of PMD in optical fibers. In our approach, we first build a full PMD model based on coupled mode theory as a continuous-time, lumped and lossless system. We then develop systematic model dimension reduction techniques, adapted from Krylov-subspace based methods for large electronic systems, in order to obtain compact and low complexity PMD models. The reduced complexity models produced by our technique match the full PMD model over a specifiable frequency range of interest, have the same structure as the full model, and are amenable to efficient software and low complexity hardware implementations for emulation and inversion. Furthermore, the reduced compexity PMD modeling framework we develop in this paper can serve as a general formalism for studying the compressibility of PMD models and emulators.**

***Index Terms*—Lossless systems, Pade approximation, polarization mode dispersion (PMD), reduced complexity modeling.**

I. Introduction

SINGLE-MODE optical fibers are used as the transmission medium in long-haul, high-speed optical fiber communication systems. The two polarization modes of light, which normally have identical propagation characteristics in a single-mode fiber, travel at different speeds due to random imperfections and asymmetries in the fiber. Moreover, the polarization axes of a fiber change randomly along the fiber causing a random coupling between the two polarization modes. The speed difference and random coupling of the polarization modes give rise to the spreading of the pulses that encode the digital information that is being transmitted. This phenomenon, called polarization mode dispersion (PMD), ultimately limits the speed at which data can be transmitted over a single-mode fiber, unless it is somehow compensated. In systems that use polarization multiplexing, PMD manifests itself by also causing cross-talk between the two polarization channels in addition to pulse spreading [1].

Since the seminal paper of Poole *et al.* [2] two decades ago, a great deal of effort has been expended in understanding PMD and its implications on communication system performance. At this point, basic theory of PMD is well established, and various techniques for the modeling and analysis of PMD, and for its emulation and compensation in practical systems have been proposed in the literature. A number of excellent reviews on various aspects of PMD are available [3]–[5].

The most commonly used PMD model is formed as a concatenation of a number of birefringent fiber sections with randomly varying phase retardations and random angular orientations. This model has been used for the simulation of PMD in computational investigations. Hardware realizations of the model in the optical or electrical domain have been employed for experimental PMD emulation, inversion and compensation. It is generally believed that the verity of PMD emulation and inversion can be improved by using a larger number of birefringent sections in this model for a given fiber length, which also widens the frequency bandwidth over which the model is accurate and realistic. However, the use of a large number of sections increases the computational cost of model evaluation in simulation studies. This increased cost would have been tolerable if only a few evaluations of the model are needed. However, PMD is a statistical phenomenon and the evaluation of its impact on system performance is typically done by evaluating the model many times in a Monte Carlo simulation. More importantly, the increased number of sections in the model results in greater and prohibitive complexity for hardware realizations. As a result, there is great interest in *reduced complexity* PMD models with only a few number of sections. However, the use of only a few sections in the commonly used PMD model usually has an unacceptable detrimental effect on the model accuracy and bandwidth. As discussed in [6], the recent interest in new PMD representations and formalisms seem to partially stem from the desire to have a *systematic* methodology for the construction of compact and low complexity PMD models that can be used for emulation and inversion purposes. In this paper, we introduce a fresh and systematic formalism to model complexity reduction for PMD. The model complexity reduction methodology we develop is based on well established Krylov-subspace based techniques (to be described in Section IV) that have been very successfully used in obtaining reduced complexity models

Manuscript received January 18, 2008; revised May 6, 2008. Current version published December 19, 2008. This work was supported in part by the Turkish Academy of Sciences GEBIP program and in part by two TUBITAK (Scientific and Technological Research Council of Turkey) Career Awards (104E057 and 104E073).

The authors are with the Department of Electrical and Electronics Engineering, Koc University, Rumeli Feneri Yolu, 34450 Sariyer-Istanbul, Turkey (e-mail: aldemir@ku.edu.tr; alperdogan@ku.edu.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2008.926920

of large electronic systems, in particular the interconnect networks on integrated circuits [7]. In Section II, we first build a full PMD model, based on coupled mode theory [8], as a continuous-time, lumped and lossless system [9]. We then, in Section III, express this full model in a canonical, state-space form in order to make it suitable for Krylov-subspace based model reduction. We show that the full PMD model we derive is in fact equivalent to the commonly used PMD model built from a concatenation of a number of birefringent fiber sections. Next, in Section IV, we first provide a brief review of model dimension reduction techniques for continuous-time, linear time-invariant, lumped systems, and then develop a reduced complexity modeling methodology for PMD. The reduced models that are obtained with this methodology can be specified to match the full PMD model over a desired bandwidth. The structure and dimension of the reduced models this technique produces are not postulated *a priori* but are deduced from the full model in a transparent, formal and systematic manner. This is a key benefit of this reduced complexity PMD modeling technique which separates it from some other approaches to low complexity PMD modeling. We believe that most low complexity PMD models proposed previously (reviewed in [6]), in the form of a concatenation of several birefringent elements and other components, can be systematically derived using the methodology we present in this paper. In fact, we show that the reduced models produced have a similar structure as the full model, as a concatenation of birefringent sections and random rotations, even though they are obtained through a numerical procedure. We present results obtained by the proposed reduced complexity PMD modeling methodology in Section V. One of the distinguishing features of PMD that separates it from other performance degrading phenomena in optical fiber communications is its statistical and time-varying nature. In Section VI, we analyze the statistical properties of the reduced models and describe how their parameters may be randomized in order to obtain the desired PMD statistics in a Monte Carlo sampling.

In the PMD literature, compact models of PMD have been developed based on various techniques. In most of these approaches, the compact model derivation is based on approximating (with a Taylor series expansion around a specific frequency) the frequency dependence of the so-called PMD vector using Poole's PMD representation [6]. For instance, if the frequency dependence of the PMD vector is modeled with a first-order, linear Taylor series expansion, then one obtains a "first-order PMD" model. Compact models obtained with this approach in the literature are usually limited to several orders, above which the model generation process becomes cumbersome and unwieldy. As discussed in [6], the meaningful bandwidth of Taylor series and Poole representation based approximations grow very slowly with model order, casting doubt on this approach overwhelmingly adopted in the PMD literature. Several approaches in the literature can be identified that deviate from this mainstream. Kogelnik *et al.* in [6] review some of these approaches, in which compact models for PMD are obtained based on Fourier and Taylor series expansions of directly the Jones transfer matrix of the fiber instead of the PMD vector based on Poole's representation. More recently in [10], Yevick *et al.* present an approach in which compact models for PMD with improved bandwidth are obtained based on Chebyshev polynomial approximations of the Jones transfer matrix of the fiber. The reduced complexity modeling methodology we develop in this paper is similar to these approaches, because the compact models obtained are in effect based on an approximation directly for the fiber transfer matrix as opposed to the PMD vector. On the other hand, our approach is based on a multipoint (in frequency) Pade approximation [11]. Compact models obtained using rational Pade approximations based on multiple expansion points can yield better model accuracy over a wider bandwidth when compared with power series expansions around a single center frequency [12]. Moreover, as it will be shown in Section IV, the use of Krylov-subspace based dimension reduction and a state-space formulation in performing multipoint Pade approximations results in a simple and systematic technique that can be used to generate compact models with varying levels of accuracy over any desired frequency bandwidth in a noncumbersome and straightforward manner.

We stress that the main contribution of the work presented in this paper is in proposing a completely novel methodology and exploring a new path for the reduced complexity modeling of PMD that deviates from the mainstream. In this current paper, we present only exploratory results that illustrate the workings of the proposed methodology and, by no means are able to investigate the full potential of the technique. We think that the methodology proposed can serve as a general and systematic formalism for studying reduced compexity PMD models. We certainly hope that it will receive critical attention from other PMD researchers, and further work will explore its full potential and try to go beyond some restrictions it may have.

In the model reduction literature, the term *order* is used to refer to the degrees of freedom, the dimension or the complexity of a model (to be made precise later), in contrast with its use in the PMD literature as the approximation *order* for the frequency dependence of the PMD vector in Poole's representation. In order not to create a confusion for the readers with a background in PMD, we avoid the term *order* in this paper and instead use model *dimension* or *complexity* where appropriate.

II. Modeling PMD Using Coupled Mode Theory as a Continuous-Time Lumped Lossless System

The following coupled partial differential equations (PDEs) (expressed in matrix form) govern the dynamics of the complex envelopes of the electric field for the two polarization components for a light wave propagating in a single-mode optical fiber [8], [13]

$$\frac{\partial \mathbf{A}(z,t)}{\partial z} = \mathbf{D}\frac{\partial \mathbf{A}(z,t)}{\partial t} + \mathbf{R}(z)\mathbf{A}(z,t) \tag{1}$$

where t is time, z is the position along the propagation direction, with

$$\mathbf{A}(z,t) = [A_1(z,t)A_2(z,t)]^T \tag{2}$$

and

$$\mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & -\Delta\beta \end{bmatrix} \tag{3}$$

The term with the time derivative in the RHS of (1) accounts for the speed difference for the two polarization components, whereas the other term models the coupling between them. The loss of the fiber (assumed polarization independent) has been factored out of the PDE in (1), and, hence, it represents a lossless 2-D system, which implies that $\mathbf{R}(z)$ in (1) should be a skew-Hermitian, 2×2 matrix, i.e., $\mathbf{R} = -\mathbf{R}^\dagger$, where $\cdot^\dagger$ denotes conjugate-transpose [8]. Moreover, the PDE in (1) uses a retarded frame of reference that moves with the group velocity of the faster polarization component. As such, (1) represents a causal system. In summary, the linear system described by (1) is a lossless, causal, distributed, continuous-time system. The coupled equations in (1) can be viewed as a special case of the coupled nonlinear Schrodinger equations that are usually used to model signal propagation in single-mode optical fibers, with the terms for the effects of chromatic dispersion and fiber nonlinearities removed [14].

We now define

$$\begin{aligned}\mathbf{B}(z,t) &= \exp\left[-\int_0^z \mathbf{R}(u)\,du\right]\mathbf{A}(z,t)\\ &= \exp\left[-\mathbf{Q}(z)\right]\mathbf{A}(z,t)\end{aligned} \tag{4}$$

and assume that the matrices $\mathbf{R}(z)$ and $\mathbf{Q}(z) = \int_0^z \mathbf{R}(u)\,du$ commute. This commutativity condition is equivalent to the condition that the product of $\mathbf{R}(z)$ and $\mathbf{Q}(z)$ is Hermitian. This follows easily from the fact that $\mathbf{Q}(z)$ is also skew-Hermitian when $\mathbf{R}(z)$ is skew-Hermitian [15]. The commutativity of $\mathbf{R}(z)$ and $\mathbf{Q}(z)$ is guaranteed if the matrices $\mathbf{R}(z)$ for all z are simultaneously diagonalizable by the same z-independent unitary transformation.

We differentiate (4) with respect to z

$$\frac{\partial \mathbf{B}(z,t)}{\partial z} = -\mathbf{R}(z)\mathbf{B}(z,t) + \exp\left[-\mathbf{Q}(z)\right]\frac{\partial \mathbf{A}(z,t)}{\partial z} \tag{5}$$

and substitute (1) in the above to obtain

$$\begin{aligned}&\frac{\partial \mathbf{B}(z,t)}{\partial z}\\ &\quad= -\mathbf{R}(z)\mathbf{B}(z,t)\\ &\qquad+ \exp\left[-\mathbf{Q}(z)\right]\left[\mathbf{D}\frac{\partial \mathbf{A}(z,t)}{\partial t} + \mathbf{R}(z)\mathbf{A}(z,t)\right]\\ &\quad= -\mathbf{R}(z)\mathbf{B}(z,t) + \exp\left[-\mathbf{Q}(z)\right]\\ &\qquad\times\left[\mathbf{D}\exp\left[\mathbf{Q}(z)\right]\frac{\partial \mathbf{B}(z,t)}{\partial t}\right.\\ &\qquad\left.+\ \mathbf{R}(z)\exp\left[\mathbf{Q}(z)\right]\mathbf{B}(z,t)\right]\\ &\quad= \exp\left[-\mathbf{Q}(z)\right]\mathbf{D}\exp\left[\mathbf{Q}(z)\right]\frac{\partial \mathbf{B}(z,t)}{\partial t}\end{aligned} \tag{6}$$

where we have used the commutativity of $\exp\left[-\mathbf{Q}(z)\right]$ with $\mathbf{R}(z)$, which immediately follows from the commutativity of $\mathbf{Q}(z)$ with $\mathbf{R}(z)$. We note here that $\exp\left[\mathbf{Q}(z)\right] = \mathbf{U}(z)$ is a unitary matrix (i.e., $\mathbf{U}(z)^\dagger\mathbf{U}(z) = \mathbf{I}$, which follows from the fact that $\mathbf{Q}(z)$ is skew-Hermitian) and its inverse is given by $\exp\left[-\mathbf{Q}(z)\right]$.

Next, we discretize the z-dependence in (6) in order to obtain a continuous-time, lumped approximation for the distributed system described by (6). When we perform this discretization, we would like to preserve the losslessness property of the system. Not all discretization schemes satisfy this constraint. Here, we use a modified trapezoidal scheme and prove that the lumped system we obtain is perfectly lossless. We discretize (6) to obtain

$$\begin{aligned}&\frac{\mathbf{B}(z+\Delta z,t) - \mathbf{B}(z,t)}{\Delta z}\\ &\qquad= \frac{\partial}{\partial t}\mathbf{U}(z)^\dagger\mathbf{D}\mathbf{U}(z)\left[\frac{\mathbf{B}(z+\Delta z,t) + \mathbf{B}(z,t)}{2}\right].\end{aligned} \tag{7}$$

The standard trapezoidal scheme preserves losslessness only in an approximate manner when the discretization granularity is fine enough so that $\mathbf{U}(z)$ for two consecutive z values are approximately equal. From (7), we derive

$$\begin{aligned}\mathbf{B}(z+\Delta z,t) = &\left[\mathbf{I} - \frac{\Delta z}{2}\frac{\partial}{\partial t}\mathbf{U}(z)^\dagger\mathbf{D}\mathbf{U}(z)\right]^{-1}\\ &\times\left[\mathbf{I} + \frac{\Delta z}{2}\frac{\partial}{\partial t}\mathbf{U}(z)^\dagger\mathbf{D}\mathbf{U}(z)\right]\mathbf{B}(z,t)\end{aligned} \tag{8}$$

where $(\partial)/(\partial t)$ should be treated as an operator, and $\mathbf{I}$ is the identity matrix. The equation above that relates $\mathbf{B}(z+\Delta z,t)$ to $\mathbf{B}(z,t)$ can be viewed as the description of a linear time-invariant (LTI) system that models one section of fiber with length Δz, between locations z and $z+\Delta z$. We rewrite (8) in the frequency domain using the complex (Laplace) frequency variable s and obtain the transfer function of this LTI system

$$\begin{aligned}\mathbb{H}^{\mathbf{B}}_{\Delta z}(z,s) = &\left[\mathbf{I} - s\frac{\Delta z}{2}\mathbf{U}(z)^\dagger\mathbf{D}\mathbf{U}(z)\right]^{-1}\\ &\times\left[\mathbf{I} + s\frac{\Delta z}{2}\mathbf{U}(z)^\dagger\mathbf{D}\mathbf{U}(z)\right]\end{aligned} \tag{9}$$

where

$$\mathbb{B}(z+\Delta z,s) = \mathbb{H}^{\mathbf{B}}_{\Delta z}(z,s)\mathbb{B}(z,s) \tag{10}$$

where $\mathbb{B}(z,s)$ is the Laplace transform of $\mathbf{B}(z,t)$ with respect to t.

The transfer function above can be rewritten as

$$\mathbb{H}^{\mathbf{B}}_{\Delta z}(z,s) = \mathbf{U}(z)^\dagger\begin{bmatrix}1 & 0\\ 0 & \frac{1-s\frac{\Delta z}{2}\Delta\beta}{1+s\frac{\Delta z}{2}\Delta\beta}\end{bmatrix}\mathbf{U}(z) \tag{11}$$

using the fact that $\mathbf{U}(z)$ is unitary. We note here that both of the diagonal entries of the matrix in the middle above represent 1-D all-pass, i.e., lossless, systems. By cascading transfer functions of the form in (11), we can compute the overall transfer function for a fiber with a total length of L. This is essentially equivalent to concatenating sections of fiber modeled with (11) to obtain a fiber of length L. We keep it general and assume that the discretization step size Δz is not necessarily uniform, and derive the overall transfer function as follows:

$$\mathbb{H}^{\mathbf{B}}(L,s) = \prod_{k=0}^{N-1}\left\{\mathbf{U}(z_k)^\dagger\begin{bmatrix}1 & 0\\ 0 & \frac{1-s\frac{\Delta z_k}{2}\Delta\beta}{1+s\frac{\Delta z_k}{2}\Delta\beta}\end{bmatrix}\mathbf{U}(z_k)\right\} \tag{12}$$

where

$$\mathbb{B}(L,s) = \mathbb{H}^{\mathbf{B}}(L,s)\mathbb{B}(0,s) \tag{13}$$

and $\Delta z_k = z_{k+1} - z_k$, with $z_0 = 0$ and $z_N = L$. We note that the product of the transfer functions above has to be computed by performing the matrix multiplications implied from the left, starting with $k = 0$. We can then derive an expression for the transfer function in terms of the Laplace transform of the original complex envelope $\mathbb{A}(z,t)$ using (4) as follows:

$$\mathbb{H}^{\mathbf{A}}(L,s) = \mathbf{U}(L)\mathbb{H}^{\mathbf{B}}(L,s)\mathbf{U}(0)^{\dagger} \tag{14}$$

where

$$\mathbb{A}(L,s) = \mathbb{H}^{\mathbf{A}}(L,s)\mathbb{A}(0,s). \tag{15}$$

We note that

$$\mathbb{H}^{\mathbf{B}}(L,s=0) = \mathbf{I}, \quad \mathbb{H}^{\mathbf{A}}(L,s=0) = \mathbf{U}(L)\mathbf{U}(0)^{\dagger}. \tag{16}$$

As seen in (14) and (16), $\mathbb{H}^{\mathbf{A}}(L,s)$ and $\mathbb{H}^{\mathbf{B}}(L,s)$ differ from each other by an s-independent (i.e., frequency independent) unitary transformation, $\mathbb{H}^{\mathbf{B}}(L,s)$ is normalized to be $\mathbf{I}$ at $s = 0$. For the rest of our treatment, we will concentrate on $\mathbb{H}^{\mathbf{B}}(L,s)$ and denote it with $\mathbb{H}(L,s)$ without the superscript $\cdot^{\mathbf{B}}$.

There is a very close connection between the fiber transfer function we have derived above and the commonly used all-order PMD model that is formed as a concatenation of birefringent fiber sections with randomly varying phase retardations and random angular orientations. The correspondence between the two models becomes apparent once we make the observation

$$\exp(-s\Delta z\,\Delta\beta) \approx \frac{1 - s\frac{\Delta z}{2}\Delta\beta}{1 + s\frac{\Delta z}{2}\Delta\beta}. \tag{17}$$

In fact, the RHS above is a first-order rational Pade approximation [11] of the continuous-time delay $\exp(-s\tau)$ where

$$\tau = \Delta z \Delta\beta. \tag{18}$$

The modified trapezoidal scheme we have used in discretizing (6) in effect results in this Pade approximation.

It is trivial to verify that the transfer functions both in (12) and (14) represent lossless, lumped, continuous-time LTI systems. The losslessness is a result of the fact both of these transfer functions satisfy

$$\mathbb{H}(z,s)^{\dagger}\mathbb{H}(z,s) = \mathbf{I} \quad \text{for } \mathrm{Re}\{s\} = 0. \tag{19}$$

Moreover, these transfer functions are both rational matrix functions in s, and as such they represent lumped continuous-time systems [9]. Each term in the product in (12) in fact represents a degree-one rational matrix and the overall transfer function in (12) represents a lumped, continuous-time LTI system of degree N.

III. State-Space Model of PMD as a Continuous-Time Lumped System

We now form a state-space realization of the two-input, two-output LTI system with the input-output matrix transfer function in (12). From (11) and (12) and using (18) as $\tau_k = \Delta z_k \Delta\beta$, we define

$$\mathbb{H}_k(s) = \mathbf{U}_k^{\dagger}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1-s\frac{\tau_k}{2}}{1+s\frac{\tau_k}{2}} \end{bmatrix}\mathbf{U}_k \tag{20}$$

and

$$\mathbf{U}(z_k)^{\dagger} = \mathbf{U}_k^{\dagger} = [\,\hat{\mathbf{u}}_k \quad \mathbf{u}_k\,] \tag{21}$$

where

$$\hat{\mathbf{u}}_k^{\dagger}\mathbf{u}_k = 0 \quad \text{and} \quad \mathbf{u}_k^{\dagger}\mathbf{u}_k = \hat{\mathbf{u}}_k^{\dagger}\hat{\mathbf{u}}_k = 1 \tag{22}$$

since $\mathbf{U}_k$ is unitary. $\mathbb{H}_k(s)$ in (20) can be expressed as

$$\mathbb{H}_k(s) = \mathbf{I} - \frac{2\,s}{s+\alpha_k}\,\mathbf{u}_k\,\mathbf{u}_k^{\dagger} = \left(\mathbf{I} - 2\,\mathbf{u}_k\,\mathbf{u}_k^{\dagger}\right) + \frac{2\,\alpha_k}{s+\alpha_k}\,\mathbf{u}_k\,\mathbf{u}_k^{\dagger} \tag{23}$$

where

$$\alpha_k = \frac{2}{\tau_k}. \tag{24}$$

The equivalence of (20) and (23) can be easily seen by observing that $\{(\alpha_k - s)/(\alpha_k + s), \mathbf{u}_k\}$ and $\{1, \hat{\mathbf{u}}_k\}$ are $\{\text{eigenvalue}, \text{eigenvector}\}$ pairs for $\mathbb{H}_k(s)$. A minimal state-space realization for the two-input, two-output, degree-one transfer function in (23) can be formulated as follows:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}x_k(t) &= -\alpha_k x_k(t) + \mathbf{u}_k^{\dagger}\mathbf{i}_k(t) \\ \mathbf{y}_k(t) &= 2\alpha_k\,\mathbf{u}_k x_k(t) + (\mathbf{I} - 2\,\mathbf{u}_k\mathbf{u}_k^{\dagger})\mathbf{i}_k(t) \end{aligned} \tag{25}$$

where x_k is the scalar state variable, $\mathbf{i}_k$ is the 2×1 input and $\mathbf{y}_k$ is the 2×1 output. The state-space model above is for one Δz_k-long section of a fiber. A state-space realization for a fiber of length L can be obtained with a cascade connection of N degree-one systems, each one described by (25) with $k = 0$ to $N-1$ as in (11). One obtains

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t) &= \mathsf{A}\mathbf{x}(t) + \mathsf{B}\mathbf{i}(t) \\ \mathbf{y}(t) &= \mathsf{C}\mathbf{x}(t) + \mathsf{D}\mathbf{i}(t) \end{aligned} \tag{26}$$

where

$$\mathbf{x} = [x_0 \quad x_1 \quad \cdots x_{N-1}]^T \tag{27}$$

and $\mathbf{i} = \mathbf{i}_0$ is the input to the first fiber section, i.e., the input to the fiber, and $\mathbf{y} = \mathbf{y}_{N-1}$ is the output of the last fiber section, i.e., output of the fiber. In (26), B is $N\times 2$, C is $2\times N$ and D is 2×2. A is an $N\times N$ lower-triangular matrix with diagonal entries equal to $-\alpha_k$ for $k = 0$ to $N-1$. The lower-triangular structure for A arises from the fact that (26) is obtained from a cascade connection of N degree-one systems and the state vector is selected as a collection of the states of the degree-one systems. As such, each state variable for the cascaded system depends on only the previous states and the input. The 2×2 transfer function matrix for a fiber of length L, given in (12), can be expressed as follows in terms of the matrices in the state-space representation in (26):

$$\mathbb{H}(L,s) = \mathsf{C}(s\mathbf{I} - \mathsf{A})^{-1}\mathsf{B} + \mathsf{D}. \tag{28}$$

IV. Reduced Complexity Modeling of PMD

The transfer functions and state-space models we described in Sections II and III can be used in building PMD models for emulation and inversion, with an appropriate choice of the discretization step sizes Δz_k, the number of sections N, $\Delta\beta$ in (3) and the unitary matrices $\mathbf{U}_k$ in (21). The complexity of these models can obviously be reduced considerably by simply choosing very large discretization step sizes set to a large fraction of the total fiber length. However, this has an unacceptable detrimental effect on the model accuracy and bandwidth. In this section, we describe a systematic and general technique for obtaining reduced-dimension, low complexity models of PMD without forfeiting model accuracy over a specified and desired bandwidth. The reduced-dimension models we derive cannot be trivially obtained by simply choosing large step sizes in the models described in Sections II and III. However, we show that the reduced-dimension models produced by our methodology can be represented with a similar structure as the one in (12) (as a cascade connection of degree-one sections) but with a more general form. The reduced-dimension modeling technique we develop preserves an important system property, i.e., losslessness, while maintaining model accuracy over a desired bandwidth.

A. Model Reduction for LTI Systems

Model dimension reduction techniques for LTI systems have been developed and used in various disciplines. The first model reduction techniques were developed in control and system theory using state-space formulations, based on truncated balanced realizations [16], Hankel-norm reduction [17] and proper orthogonal decomposition [18]. Later in the 1990s, model-dimension reduction techniques that are based on forming Pade approximations of rational transfer functions were introduced [19], specifically for building reduced models of interconnect, i.e., wiring, on integrated circuits. The observation of the connection between Pade approximations and Krylov subspaces [20], [12] enabled numerically efficient and accurate application of Pade-type model reduction techniques to various problems in electronic circuit and system design [7].

When model reduction is applied to a lumped, continuous-time system described by state-space equations of the form in (26), the goal is to obtain a reduced dimension, smaller system

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mathbf{x}}(t) &= \tilde{\mathsf{A}}\tilde{\mathbf{x}}(t) + \tilde{\mathsf{B}}\mathbf{i}(t)\\ \mathbf{y}(t) &= \tilde{\mathsf{C}}\tilde{\mathbf{x}}(t) + \mathsf{D}\mathbf{i}(t)\end{aligned} \tag{29}$$

while approximately preserving the input-output behavior of the system as quantified by its transfer function $\mathbb{H}(s)$ in (28), i.e.,

$$\mathbb{H}(s) \approx \tilde{\mathbb{H}}(s) \tag{30}$$

where

$$\tilde{\mathbb{H}}(s) = \tilde{\mathsf{C}}(s\mathrm{I} - \tilde{\mathsf{A}})^{-1}\tilde{\mathsf{B}} + \mathsf{D}. \tag{31}$$

For notational simplicity, we omit the fiber length argument L for $\mathbb{H}(s)$. Above in (31), $\tilde{\mathsf{B}}$ is $n\times 2$, C is $2\times n$, and A is an $n\times n$, where $n \ll N$, i.e., the dimension of the reduced system in (29) is much smaller than the dimension of the original system in (26). D in (29) is identical to the D in (26), both 2×2 matrices. Reduction in the size of D is neither possible nor desired. For the rest of our treatment, for notational simplicity, we will omit the frequency independent term D in the transfer function expressions.

Various approaches to model reduction differ from each other in defining the meaning of the approximation in (30). The reduced models produced by approaches based on truncated balanced realizations or Hankel-norm reduction feature a frequency-domain L^∞ error bound for the approximation. Thus, these techniques in general produce reduced models with a globally accurate (in frequency) transfer function [21]. However, truncated balanced realization or Hankel-norm reduction techniques are not useful for lossless systems: It can be shown that Hankel singular values for a lossless system are all equal to each other [22], and, hence, one cannot obtain good reduced models for lossless systems with a globally accurate approximation of the transfer function of the full system. On the other hand, with model dimension reduction approaches based on Pade approximations, one tries to obtain a reduced dimension model with good local accuracy (in frequency) for the matrix rational transfer function by matching the leading coefficients in a power series expansion (in the complex frequency variable s) of the reduced transfer function (also a matrix rational function) to the expansion coefficients of the full transfer function. The model reduction problem from this point of view is the determination of a much lower degree matrix rational transfer function that approximates a higher degree one. We show in this paper that Pade-type model reduction techniques are more meaningful and useful for lossless systems.

B. Model Reduction Based on Pade Approximations

Reduced models based on Pade approximations are typically produced using a projection framework [7], in which (26) is reduced to (29) by using two $N \times n$ projection matrices U_n and V_n that satisfy the bi-orthogonality condition

$$\mathsf{V}_n^\dagger \mathsf{U}_n = \mathrm{I}. \tag{32}$$

The reduced size matrices in (29) are computed from the full size matrices in (26) using the projection relationships as follows:

$$\tilde{\mathsf{A}} = \mathsf{V}_n^\dagger \mathsf{A}\mathsf{U}_n \quad \tilde{\mathsf{B}} = \mathsf{V}_n^\dagger \mathsf{B} \quad \tilde{\mathsf{C}} = \mathsf{C}\mathsf{U}_n. \tag{33}$$

The reduced dimension state vector $\tilde{\mathbf{x}}(t)$ and the full size state vector $\mathbf{x}(t)$ are related to each other by

$$\mathbf{x}(t) = \mathsf{U}_n\tilde{\mathbf{x}}(t). \tag{34}$$

The gist of the projection method lies in the construction or choice of the projection matrices U_n and V_n, which directly determine the approximation properties of the reduced model. This is where Krylov subspaces come into the picture.

An r-dimensional block Krylov subspace associated with an $N \times N$ matrix M and an $N \times m$ matrix P is defined as follows:

$$\mathcal{K}_r(\mathsf{M},\mathsf{P}) = \mathrm{colspan}[\mathsf{P} \quad \mathsf{M}\mathsf{P} \quad \cdots \quad \mathsf{M}^{k-1}\mathsf{P}] \tag{35}$$

where "colspan" above is defined as the span of the set of vectors composed of the columns of all of the $N \times m$ matrices in the set $\{\mathsf{P}, \mathsf{M}\mathsf{P}, \ldots, \mathsf{M}^{k-1}\mathsf{P}\}$, a total of $k\,m$ column vectors. If all of these vectors are linearly independent, then the dimension of the Krylov subspace $\mathcal{K}_r(\mathsf{M}, \mathsf{P})$ is $r = k\,m$ assuming that $r < N$, otherwise $r < km$.

If one constructs or chooses the projection matrices to be equal $\mathsf{U}_n = \mathsf{V}_n$ and such that

$$\mathcal{K}_r\left((\sigma\,\mathrm{I} - \mathsf{A})^{-1}, (\sigma\,\mathrm{I} - \mathsf{A})^{-1}\mathsf{B}\right) \subseteq \mathrm{colspan}\mathsf{U}_n \tag{36}$$

then one can show that the first k coefficients in a power series expansion of the full transfer function $\mathbb{H}(s)$ around the expansion point σ matches the corresponding coefficients in the expansion of the reduced transfer function $\tilde{\mathbb{H}}(s)$, i.e.,

$$\mathbb{H}(s) = \tilde{\mathbb{H}}(s) + \mathcal{O}((s-\sigma)^k). \tag{37}$$

The detailed derivation of the above result is beyond the scope of this paper [7]. However, one arrives at this result by first deriving the power series expansion of $\mathbb{H}(s)$ around σ

$$\begin{aligned}\mathsf{C}\,(s\mathrm{I} - \mathsf{A})^{-1}\,\mathsf{B} &= \mathsf{C}((s-\sigma)\mathrm{I} + \sigma\mathrm{I} - \mathsf{A})^{-1}\mathsf{B} \\ &= \mathsf{C}\left(\mathrm{I} + (s-\sigma)(\sigma\mathrm{I} - \mathsf{A})^{-1}\right)^{-1}\left((\sigma\mathrm{I} - \mathsf{A})^{-1}\mathsf{B}\right) \\ &= \sum_{i=0}^{\infty}(-1)^i[\mathsf{C}(\sigma\mathrm{I} - \mathsf{A})^{-i}((\sigma\mathrm{I} - \mathsf{A})^{-1}\mathsf{B})](s-\sigma)^i.\end{aligned} \tag{38}$$

The above is a matrix power series expansion around σ, and the expansion coefficients are in fact 2×2 matrices for our case. The coefficient matrices in this power series expansion are called the *moments* of the transfer function at σ.

When the two projection matrices are constructed to be equal, i.e., $\mathsf{U}_n = \mathsf{V}_n$ as above, then $\mathsf{U}_n^{\dagger}\mathsf{U}_n = \mathrm{I}_n$ follows from (32) implying that the columns of U_n are orthogonal to each other. We note that there are infinitely many projection matrices U_n that satisfy (36), and the use of any of them results in a reduced model with the mentioned moment matching property. Construction of a projection matrix U_n with orthonormal columns which satisfies (36) can be accomplished in a straightforward manner by simply forming the block matrix

$$[(\sigma\,\mathrm{I} - \mathsf{A})^{-1}\mathsf{B} \quad (\sigma\,\mathrm{I} - \mathsf{A})^{-2}\mathsf{B} \quad \cdots \quad (\sigma\,\mathrm{I} - \mathsf{A})^{-k}\mathsf{B}] \tag{39}$$

and applying the Gramm–Schmidt orthogonalization process to its columns [23]. However, this procedure is numerically not well conditioned, because the columns of the above matrix (computed with repeated multiplications with the matrix $(\sigma\,\mathrm{I} - \mathsf{A})^{-1}$) become (numerically) nearly linearly dependent very quickly. An orthonormal basis for the columns of (39) can instead be constructed via the Arnoldi procedure in a numerically stable manner [12], [24], [25]. The Arnoldi method is essentially a modified Gramm–Schmidt orthogonalization scheme [23], in which the columns of (39) are never explicitly generated. Instead, other vectors that span the same subspace as the columns of (39) are computed iteratively in a numerically well-conditioned manner.

A reduced model generated with a projection matrix U_n that satisfies (36) is naturally more accurate around the expansion point σ and may lose accuracy when s is away from σ. The frequency band around σ over which the reduced model has acceptable accuracy can be widened by increasing the number of moments matched at σ, i.e., k in (39). However, one may have to match many moments for reasonable accuracy away from σ, due to the power series nature of the expansion, which increases the size of the reduced model. When model accuracy over a wide band of frequencies is needed, it is more optimal to match a few moments at a number of expansion points as opposed to matching many moments at a single expansion point. *Multipoint moment matching* [12] can be achieved by simply modifying (36) as follows:

$$\bigcup_{i=1}^{S}\mathcal{K}_{r_i}((\sigma_i\,\mathrm{I} - \mathsf{A})^{-1}, (\sigma_i\,\mathrm{I} - \mathsf{A})^{-1}\mathsf{B}) \subseteq \mathrm{colspan}\mathsf{U}_n. \tag{40}$$

A projection matrix U_n satisfying the above can be constructed using a generalization of the Arnoldi iteration [12]. With S expansion points and k moments matched at each one of them typically results in a reduced model of dimension $n = 2\,S\,k$, in our case, since the column dimension of B is 2. However, if the collection of the columns of (39) for a number of expansion points are linearly dependent, then the reduced model dimension will be smaller. With a suitable choice of the expansion points σ_i in (40) and the number of moments to be matched at each one of them, one can produce compact reduced models with reasonable accuracy over a wide bandwidth of interest.

C. Preserving Losslessness in Model Reduction

In model dimension reduction, it is important to preserve system properties such as stability, passivity, losslessness, other relevant structural properties and at the same time obtain compact reduced models with the smallest possible dimension with good transfer function accuracy over a desired bandwidth [7]. For example, preservation of passivity[1] has been an important consideration in modeling of interconnect structures on integrated circuits [7], [26], [27]. In our case, the most important system property that should be preserved is the losslessness of the PMD model. We next describe how this can be achieved with a projection framework for model reduction. The scheme we develop here also preserves stability.

The results we derive below on preserving losslessness in model reduction are similar to the ones that had been derived in the model reduction literature on preserving passivity [7], [26], [27]. The techniques for preservation of passivity build on a classical result called the *positive real lemma* [28], [29], whereas the results we derive below build on a related result called the *lossless bounded real lemma* to be described right below.

We first state two important results on lossless systems that were developed in the 1960s and 1970s [28], [29].

[1]A system is passive if it does not generate energy.

Result 4.1: Let $(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D})$ be any *minimal* state-space realization of a transfer function $\mathbb{H}(s)$. Then $\mathbb{H}(s)$ is lossless if and only if there exists a unique positive definite, Hermitian matrix P satisfying

$$\mathsf{A}\mathsf{P}+\mathsf{P}\mathsf{A}^{\dagger}+\mathsf{B}\mathsf{B}^{\dagger}=\mathbf{0} \tag{41}$$
$$\mathsf{C}\mathsf{P}+\mathsf{D}\mathsf{B}^{\dagger}=\mathbf{0} \tag{42}$$
$$\mathsf{D}\mathsf{D}^{\dagger}=\mathbf{I}. \tag{43}$$

A version of the above result was first derived in the context of analyzing the realizations of scattering matrices of lossless multiports in classical network theory, and it is known as the *lossless bounded real lemma* [28], [29].

Result 4.2: A minimal state-space realization $(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D})$ for a lossless system can always be *balanced*, and the positive definite matrix P in Result 4.1 becomes equal to a scalar multiple of the identity matrix, i.e., $\mathsf{P}=p\mathrm{I}$, where p is a positive real scalar.

We are now in a position to state and prove the following:

Theorem 4.1: If any $N\times n$ projection matrix U_n satisfying

$$\mathsf{U}_n^{\dagger}\,\mathsf{U}_n=\mathrm{I}_n \tag{44}$$

is used to generate an n-dimensional reduced model with

$$\tilde{\mathsf{A}}=\mathsf{U}_n^{\dagger}\mathsf{A}\mathsf{U}_n\quad \tilde{\mathsf{B}}=\mathsf{U}_n^{\dagger}\mathsf{B}\quad \tilde{\mathsf{C}}=\mathsf{C}\mathsf{U}_n\quad \tilde{\mathsf{D}}=\mathsf{D} \tag{45}$$

from a balanced, minimal realization $(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D})$ of a lossless system of dimension N, then the reduced system realization $(\tilde{\mathsf{A}},\tilde{\mathsf{B}},\tilde{\mathsf{C}},\tilde{\mathsf{D}})$ is also lossless, minimal and balanced.

Proof: Due to Result 4.1 and Result 4.2 above, there exists a positive definite, Hermitian matrix $\mathsf{P}=p\mathrm{I}_N$, a scalar multiple of an $N\times N$ identity matrix that satisfies (41). Then, we can derive

$$\begin{aligned}\mathsf{U}_n^{\dagger}\mathsf{P}&=p\mathsf{U}_n^{\dagger}\mathrm{I}_N=p\mathsf{U}_n^{\dagger}\\&=p\mathrm{I}_n\mathsf{U}_n^{\dagger}=p\mathsf{U}_n^{\dagger}\mathsf{U}_n\mathsf{U}_n^{\dagger}\\&=p\mathsf{U}_n^{\dagger}\mathrm{I}_N\mathsf{U}_n\mathsf{U}_n^{\dagger}=\mathsf{U}_n^{\dagger}\mathsf{P}\mathsf{U}_n\mathsf{U}_n^{\dagger}.\end{aligned} \tag{46}$$

We multiply (41) from the left with $\mathsf{U}_n^{\dagger}$ and from the right with U_n and obtain

$$\mathsf{U}_n^{\dagger}\mathsf{A}\mathsf{P}\mathsf{U}_n+\mathsf{U}_n^{\dagger}\mathsf{P}\mathsf{A}^{\dagger}\mathsf{U}_n+\mathsf{U}_n^{\dagger}\mathsf{B}\mathsf{B}^{\dagger}\mathsf{U}_n=\mathbf{0}. \tag{47}$$

We use (46) in the above

$$\mathsf{U}_n^{\dagger}\mathsf{A}\mathsf{U}_n\mathsf{U}_n^{\dagger}\mathsf{P}\mathsf{U}_n+\mathsf{U}_n^{\dagger}\mathsf{P}\mathsf{U}_n\mathsf{U}_n^{\dagger}\mathsf{A}^{\dagger}\mathsf{U}_n+\mathsf{U}_n^{\dagger}\mathsf{B}\mathsf{B}^{\dagger}\mathsf{U}_n=\mathbf{0} \tag{48}$$

define

$$\tilde{\mathsf{P}}=\mathsf{U}_n^{\dagger}\mathsf{P}\mathsf{U}_n=p\mathsf{U}_n^{\dagger}\mathrm{I}_N\mathsf{U}_n=p\mathsf{U}_n^{\dagger}\mathsf{U}_n=p\mathrm{I}_n \tag{49}$$

$\tilde{\mathsf{P}}$ defined above is an $n\times n$ positive-definite matrix, in fact a scalar multiple of an $n\times n$ identity matrix. We substitute (45) and (49) in (48) to obtain

$$\tilde{\mathsf{A}}\tilde{\mathsf{P}}+\tilde{\mathsf{P}}\tilde{\mathsf{A}}^{\dagger}+\tilde{\mathsf{B}}\tilde{\mathsf{B}}^{\dagger}=\mathbf{0}. \tag{50}$$

We multiply (42) from the right with U_n and obtain

$$\mathsf{C}\mathsf{P}\mathsf{U}_n+\mathsf{D}\mathsf{B}^{\dagger}\mathsf{U}_n=\mathbf{0}. \tag{51}$$

We use (46) in the above

$$\mathsf{C}\mathsf{U}_n\mathsf{U}_n^{\dagger}\mathsf{P}\mathsf{U}_n+\mathsf{D}\mathsf{B}^{\dagger}\mathsf{U}_n=\mathbf{0} \tag{52}$$

and substitute (45) and (49) in the above to obtain

$$\tilde{\mathsf{C}}\tilde{\mathsf{P}}+\tilde{\mathsf{D}}\tilde{\mathsf{B}}^{\dagger}=\mathbf{0}. \tag{53}$$

The proof is complete with (45), (49), (50), and (53), due to Result 4.1 and Result 4.2. ■

For Theorem 4.1 stated above to be applicable, the right and left projection matrices U_n and V_n in (32) and (33) need to be equal to each other, and more importantly, the state-space realization $(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D})$ for the lossless system needs to be *balanced*. It can be shown that the state-space realization given by (25) and (26) is in fact balanced with P as a scalar multiple of the identity matrix. We can remove the balancing requirement by choosing a different left projection matrix. In order to demonstrate how this can be accomplished, we first state a result that is equivalent to Result 4.2, another incarnation of the lossless bounded real lemma [29], [30]:

Result 4.3: Let $(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D})$ be any *minimal* state-space realization of a transfer function $\mathbb{H}(s)$. Then $\mathbb{H}(s)$ is lossless if and only if there exists a unique positive definite, Hermitian matrix P satisfying

$$\mathsf{P}\mathsf{A}+\mathsf{A}^{\dagger}\mathsf{P}+\mathsf{C}^{\dagger}\mathsf{C}=\mathbf{0} \tag{54}$$
$$\mathsf{P}\mathsf{B}+\mathsf{C}^{\dagger}\mathsf{D}=\mathbf{0} \tag{55}$$
$$\mathsf{D}\mathsf{D}^{\dagger}=\mathbf{I}. \tag{56}$$

We now state and prove the following.

Theorem 4.2: If any $N\times n$ U_n and V_n satisfying

$$\mathsf{V}_n^{\dagger}=(\mathsf{U}_n^{\dagger}\,\mathsf{P}\,\mathsf{U}_n)^{-1}\mathsf{U}_n^{\dagger}\,\mathsf{P} \tag{57}$$

are used to generate an n-dimensional reduced dimension model with

$$\tilde{\mathsf{A}}=\mathsf{V}_n^{\dagger}\mathsf{A}\mathsf{U}_n\quad \tilde{\mathsf{B}}=\mathsf{V}_n^{\dagger}\mathsf{B}\quad \tilde{\mathsf{C}}=\mathsf{C}\mathsf{U}_n\quad \tilde{\mathsf{D}}=\mathsf{D} \tag{58}$$

from a minimal realization $(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D})$ of a lossless system of dimension N, then the reduced system realization $(\tilde{\mathsf{A}},\tilde{\mathsf{B}},\tilde{\mathsf{C}},\tilde{\mathsf{D}})$ is also lossless and minimal. P in (57) is the positive definite solution of (54).

Proof: We first note that $\mathsf{V}_n^{\dagger}\,\mathsf{U}_n=\mathrm{I}_n$ with V_n as in (57). We then define

$$\tilde{\mathsf{P}}=\mathsf{U}_n^{\dagger}\mathsf{P}\mathsf{U}_n. \tag{59}$$

Then

$$\tilde{\mathsf{P}}\tilde{\mathsf{A}}=\mathsf{U}_n^{\dagger}\mathsf{P}\,\mathsf{A}\mathsf{U}_n \tag{60}$$

immediately follows from (57)–(59), using which we obtain

$$\tilde{\mathsf{P}}\tilde{\mathsf{A}}+\tilde{\mathsf{A}}^{\dagger}\tilde{\mathsf{P}}+\tilde{\mathsf{C}}^{\dagger}\tilde{\mathsf{C}}=\mathbf{0} \tag{61}$$

by multiplying (54) with $\mathsf{U}_n^{\dagger}$ from the left and with U_n from the right. We also obtain

$$\mathsf{U}_n^{\dagger}\mathsf{P}\mathsf{B}+\mathsf{U}_n^{\dagger}\mathsf{C}^{\dagger}\mathsf{D}=\mathbf{0} \tag{62}$$

and

$$\tilde{\mathsf{P}}\tilde{\mathsf{B}}+\tilde{\mathsf{C}}_n^{\dagger}\tilde{\mathsf{D}}=\mathbf{0} \tag{63}$$

from (55) and (57)–(59). The proof is complete with (58), (59), (61), and (63), due to Result 4.3. ■

We should emphasize here that Theorem 4.1 and Theorem 4.2 are valid for *any* $N \times n$ orthonormal projection matrix U_n, no matter how it is constructed. In fact, for Theorem 4.2 to be valid, any full rank U_n (not necessarily orthonormal) suffices. However, if U_n is constructed as described in Section IV-B, then the reduced model obtained will not only be lossless but also be an accurate representation of the full system over a bandwidth of interest. If the projection matrices are chosen as prescribed in Theorem 4.2, then it can be shown that the stability of the system is also preserved [31].

D. Factorization of the Reduced Transfer Function

It would be desirable if a factorization of the reduced transfer function in (31) can be performed and expressed in a similar structure as the cascade of degree-one transfer functions of the form in (20) for the full system. This would be, for instance, very useful if the reduced transfer function is to be realized in hardware for emulation and compensation purposes.

The factorization algorithm we develop here uses a modified version of another result on lossless systems that was derived back in late 1960s and early 1970s [28].

Result 4.4: Any lossless rational matrix transfer function of degree n can be factorized as a product (cascade) of n lossless rational matrix transfer functions of degree-one. The general form of a degree-one lossless, square, rational matrix transfer function is

$$\mathbb{H}_o(s) = \left(\mathbf{I} - \frac{2\,\alpha}{\alpha - j\,\beta}\,\mathbf{u}\,\mathbf{u}^\dagger\right) + \frac{2\,\alpha}{(s+\alpha+j\,\beta)}\,\frac{(\alpha+j\,\beta)}{(\alpha-j\,\beta)}\,\mathbf{u}\,\mathbf{u}^\dagger \quad (64)$$

where $j = \sqrt{-1}$, the complex (column) vector $\mathbf{u}$ satisfies $\mathbf{u}^\dagger\,\mathbf{u} = 1$, α and β are real numbers. The above transfer function is normalized to satisfy $\mathbb{H}_o(0) = \mathbf{I}$. The most general form for the degree-one transfer function can be obtained by multiplying $\mathbb{H}_o(s)$ with any s-independent unitary matrix.

The degree-one transfer function in (64) is a generalized version of the one in (23). The pole of the transfer function in (23) is real, however, the pole in (64) above is allowed to be complex. In fact, if the imaginary part β of the pole is set to 0 in (64), one obtains (23). The transfer function in (64) can also be expressed in a similar form as the one in (20)

$$\mathbb{H}_o(s) = \mathbf{U}^\dagger \begin{bmatrix} 1 & 0 \\ 0 & \frac{(-s+\alpha-j\,\beta)}{(s+\alpha+j\,\beta)}\,\frac{(\alpha+j\,\beta)}{(\alpha-j\,\beta)} \end{bmatrix} \mathbf{U} \quad (65)$$

with

$$\mathbf{U}^\dagger = [\hat{\mathbf{u}} \quad \mathbf{u}] \quad (66)$$

where

$$\hat{\mathbf{u}}^\dagger\,\mathbf{u} = 0 \quad \text{and} \quad \mathbf{u}^\dagger\,\mathbf{u} = \hat{\mathbf{u}}^\dagger\,\hat{\mathbf{u}} = 1. \quad (67)$$

The equivalence of (64) and (65) can be easily seen by observing that $\{((-s+\alpha-j\,\beta)/(s+\alpha+j\,\beta))((\alpha+j\,\beta)/(\alpha-j\,\beta)), \mathbf{u}\}$ and $\{1, \hat{\mathbf{u}}\}$ are $\{\text{eigenvalue}, \text{eigenvector}\}$ pairs for $\mathbb{H}_o(s)$. A minimal state-space realization for the degree-one transfer function in (64) can be formulated with

$$\mathsf{A} = -(\alpha + j\,\beta) \quad \mathsf{B} = \mathbf{u}^\dagger$$
$$\mathsf{C} = \frac{2\,\alpha\,(\alpha+j\,\beta)}{\alpha - j\,\beta}\,\mathbf{u} \quad \mathsf{D} = \mathbf{I} - \frac{2\,\alpha}{\alpha - j\,\beta}\,\mathbf{u}\,\mathbf{u}^\dagger. \quad (68)$$

We note that the pole $(\alpha + j\,\beta)$ and the unity norm vector $\mathbf{u}$ completely specify the above realization. When n degree-one state-space realizations of the form in (68) are cascaded to obtain the state-space realization for a system of dimension n, the matrix A_c for the cascaded system can be constructed as a lower triangular matrix with a proper ordering of the state variables. The diagonal entries of A_c are then equal to $-(\alpha_k + j\,\beta_k)$ for $k = 0 \ldots n-1$. The matrix B_c for the cascaded realization has the following particular form

$$\mathsf{B}_c = \begin{bmatrix} \mathbf{u}_0^\dagger \\ \mathbf{u}_1^\dagger \left(\mathbf{I} - \frac{2\,\alpha_0}{\alpha_0 - j\,\beta_0}\,\mathbf{u}_0\,\mathbf{u}_0^\dagger\right) \\ \vdots \\ \mathbf{u}_{n-1}^\dagger \prod_{k=0}^{n-2} \left(\mathbf{I} - \frac{2\,\alpha_k}{\alpha_k - j\,\beta_k}\,\mathbf{u}_k\,\mathbf{u}_k^\dagger\right) \end{bmatrix} \quad (69)$$

where the matrix product Π above is from the left.

We now describe an algorithm for computing the factorization of lossless rational matrix transfer functions as the cascade of degree-one sections in the form given in (64) or (65), starting from a state-space realization for the transfer function. The factorization procedure we describe is general, i.e., it can be applied to any lossless transfer function. However, our intention is to use it on the lossless reduced state-space realization we obtain using the model reduction techniques we discussed before.

In the factorization algorithm, we start with a minimal and balanced state-space realization for a lossless system, with otherwise arbitrary $(\tilde{\mathsf{A}}, \tilde{\mathsf{B}}, \tilde{\mathsf{C}}, \tilde{\mathsf{D}})$. In such a realization, $\tilde{\mathsf{A}}$ is not necessarily lower triangular. For instance, if $\tilde{\mathsf{A}}$ is obtained by a unitary projection operation as in (45) from a larger possibly lower triangular A, $\tilde{\mathsf{A}}$ in general will not be lower triangular. We first compute a (complex) Schur decomposition [32] of $\tilde{\mathsf{A}}$

$$\tilde{\mathsf{A}} = \mathsf{T}\tilde{\mathsf{A}}_c\mathsf{T}^\dagger \quad (70)$$

where $\tilde{\mathsf{A}}_c$ is now lower triangular and T is unitary with $\mathsf{T}^\dagger\mathsf{T} = \mathrm{I}$. We then perform a similarity transformation on $(\tilde{\mathsf{A}}, \tilde{\mathsf{B}}, \tilde{\mathsf{C}}, \tilde{\mathsf{D}})$ with

$$\tilde{\mathsf{A}}_c = \mathsf{T}^\dagger\tilde{\mathsf{A}}\mathsf{T} \quad \tilde{\mathsf{B}}_c = \mathsf{T}^\dagger\tilde{\mathsf{B}} \quad \tilde{\mathsf{C}}_c = \tilde{\mathsf{C}}\mathsf{T} \quad \tilde{\mathsf{D}}_c = \tilde{\mathsf{D}} \quad (71)$$

which preserves the transfer function. We note here that the unitary similarity transformation in (71) above is similar to the unitary projection in (45) albeit with some crucial differences. The operation in (45) is performed with a rectangular projection matrix, and, hence, it reduces the dimension of the system and approximates the transfer function, whereas the unitary transformation in (71) above is performed with a square matrix T and it preserves the system dimension and the exact transfer function. It should also be noted that the unitary transformation in (71) does not destroy the balanced property of the realization [30].

The complex poles $-(\alpha_k + j\,\beta_k), k = 0 \ldots n-1$ for the degree-one sections of the factored system are given by the diagonal entries of $\tilde{\mathsf{A}}_c$. We can then compute the vectors $\mathbf{u}_k$ for the sections in a successive manner, from $k = 0$ to $k = n-1$, based on the structure in (69). During this process, after each $\mathbf{u}_k$ is computed, it needs to be normalized so that it satisfies $\mathbf{u}_k^\dagger\,\mathbf{u}_k = 1$. We observe here that the off-diagonal entries of $\tilde{\mathsf{A}}_c$, $\tilde{\mathsf{C}}_c$ and $\tilde{\mathsf{D}}_c$ are not needed in determining the parameters of the degree-one sections for the factorization.

The factorization procedure described above enables us to represent the reduced transfer function for PMD in exactly the same structure as the full size model, as cascaded alternating s-independent unitary matrices and s-dependent diagonal matrices of the form in (20) and (65). The pole that defines the s-dependent sections in the full PMD model derived from coupled mode theory is always real, and the pole frequency α_k is in fact equal to twice the inverse of the delay difference between the two polarization components of one section of fiber, based on (17) and (24). On the other hand, the poles of the s-dependent sections in the reduced PMD model will be in general complex, and there is no apparent relationship between these poles and any physical delay in the fiber. However, when we present the applications of the theory and techniques we have developed for reduced complexity PMD modeling, we will be in a position to form an interpretation for the physical meaning of the complex poles of the reduced model.

V. Results for Reduced Complexity PMD Modeling

A. Setup and Preliminaries

We first explain the technical issues related to our setup in order to help interpret our results better.

The full model we use is constructed with $N = 200$ sections, i.e., with $\tau_1 = \ldots = \tau_N = \tau$, with identical real poles at $-\alpha = -(2/\tau)$. α is chosen in such a way so that the mean differential group delay (DGD) for the full model is $\mathrm{meanDGD} = 10$ ps.

In order to generate different realizations of the full model, we choose the vector $\mathbf{u}_k$ in (23) for each section randomly, using (81) to be explained in Section VI where we investigate the statistical properties of the full as well as the reduced models.

In order to quantify the difference (residual error) between the full model and the reduced model, we use an error measure originally proposed in [10]. We define

$$\mathbb{H}_{\mathrm{res}}(\omega) = \tilde{\mathbb{H}}_n(\omega)^\dagger \mathbb{H}_N(\omega) \tag{72}$$

as the residual transfer matrix, where $\tilde{\mathbb{H}}_n(\omega)$ and $\mathbb{H}_N(\omega)$ are the Jones transfer function matrices for the reduced model with n sections and the full model with $N \gg n$ sections respectively. If the transfer function matrices for the full and reduced model are identical then the residual transfer matrix $\mathbb{H}_{\mathrm{res}}(\omega)$ will be equal to an identity matrix. In order to quantify the deviation of $\mathbb{H}_{\mathrm{res}}(\omega)$ from an identity matrix, we define the *residual error* (RE) as the magnitude of the difference between the two eigenvalues of the residual transfer matrix

$$\mathrm{RE}(\omega) = |\lambda_1\{\mathbb{H}_{\mathrm{res}}(\omega)\} - \lambda_2\{\mathbb{H}_{\mathrm{res}}(\omega)\}|. \tag{73}$$

TABLE I
Reduced Dimension PMD Models

Model Name	Expansion Points (GHz)					Model Dimension
	0	-10	+10	-19	+19	
RDM 1	1	0	0	0	0	2
RDM 2	2	0	0	0	0	4
RDM 3	3	0	0	0	0	6
RDM 4	1	1	1	0	0	6
RDM 5	2	2	2	0	0	12
RDM 6	2	2	2	1	1	16

Justification for this error measure and a short discussion on its meaning can be found in [10]. The residual transfer matrix $\mathbb{H}_{\mathrm{res}}(\omega)$ is a 2×2 unitary matrix, its eigenvalues lie on the unit circle in the complex plane. Hence, the residual error in (73) satisfies

$$0 \le \mathrm{RE}(\omega) \le 2. \tag{74}$$

We evaluate the reduced models in the frequency range between -20 and 20 GHz, a 40-GHz bandwidth corresponding to a 40-Gbit/sec system with a bit period of 25 ps, assuming binary signaling. The total mean DGD of 10 ps we consider amounts to 40% of this bit period.

The reduced complexity PMD models reported in this section are obtained using the multipoint moment matching technique described in Section IV-B with an implementation based on the Arnoldi process [12], [24], [25]. The reduced PMD model for a particular fiber realization is computed by using this technique on the random instance of the full model generated as described above. The model reduction process we have implemented preserves the losslessness and stability of the full model, based on the results presented in Section IV-C. The reduced models are factorized into a cascade connection of degree-one sections using the algorithm developed in Section IV-D.

B. Results for One Fiber Realization

In order to illustrate the workings of the model reduction methodology described in the paper, we report results obtained on six reduced dimension models (RDMs). Table I provides information on these models, in particular, the frequencies for the expansion points used for the multipoint Pade approximation, the number of moments matched at each one of these expansion points and the resulting model dimension (number of degree-one sections). The expansion points [σ_i's in (40)] were chosen to be purely imaginary at the frequencies shown in Table I.

In Fig. 1, the residual error [defined by (73)] for all of the six RDMs in Table I are shown for a particular realization of the full model. The impact of the expansion point(s) for the multipoint Pade approximation and the number of moments matched at each expansion point is clearly observable in this plot. The RDMs match the full model exactly at the expansion points. The accuracy of the RDMs away from the expansion points can be improved by either matching more moments at an expansion point or by introducing more expansion points. It is easily discernible that the optimal reduced model should have several expansion points distributed in the frequency range of

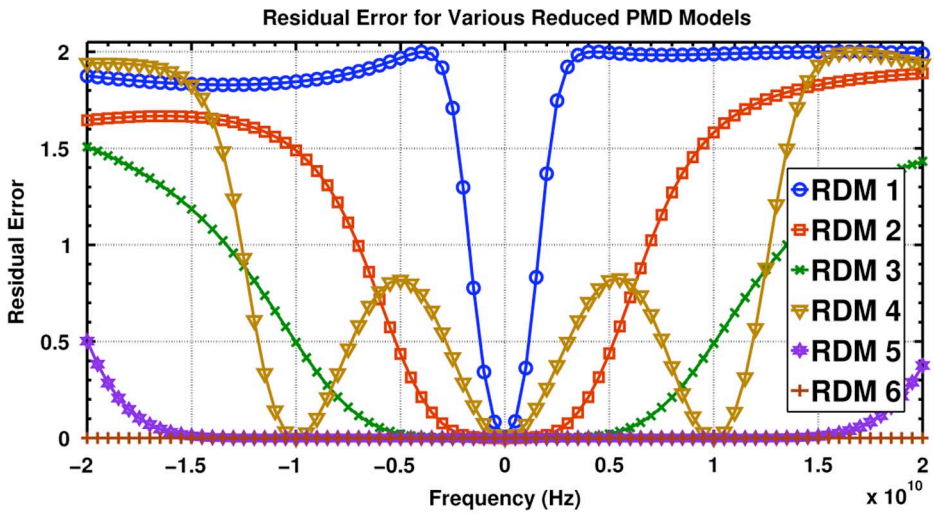

Fig. 1. Residual error for the reduced models in Table I: One fiber realization.

interest with a few moments matched at each of them, as opposed to a single expansion point in the middle with all moments matched there. The multipoint aspect of the reduced complexity PMD modeling technique we propose in this paper offers great flexibility in tailoring the accuracy of RDMs throughout a frequency band of interest. In order to better appreciate the accuracy of RDM 5 and RDM 6, we also provide a version of the residual error graph with a logarithmic residual error axis in Fig. 2. In Figs. 3 and 4, a diagonal entry of the fiber transfer function and the frequency dependent DGD for both the full model and RDM 5 are shown for a particular fiber realization. We observe by comparing these graphs with the ones in Figs. 1 and 2 that the residual error defined by (73) provides a meaningful measure for the accuracy of a reduced model in the sense of matching both the transfer function and the DGD frequency profile. We have chosen to use RDM 5 from Table I for generating the graphs presented in Figs. 3 and 4 in order to demonstrate the fact that a low residual error (in the frequency range from -15 to $+15$ GHz) corresponds to an excellent match between the RDM and the full model in both the transfer function and the frequency dependence of DGD, whereas for higher values of residual error (for offset frequencies larger than 15 GHz in absolute value), the agreement between the RDM and the full model degrades. However, RDM 6 with dimension 16 (not RDM 5 with dimension 12) is the most accurate RDM we consider in Table I. If similiar graphs for the fiber transfer function and the frequency dependence of DGD are generated for the more accurate RDM 6, the curves for the RDM and the full model would have been indistinguishable for the entire frequency band from -20 to $+20$ GHz.

C. Reduced Models With Wider Accuracy Bandwidth

The parameters for the most accurate model in Table I, RDM 6 presented in the previous section, were chosen so that it is acccurate in a 40-GHz bandwidth. Using our methodology, one can, easily and in a systematic manner, generate models which are accurate over a wider bandwidth if needed, albeit at the expense of an increase in the model dimension. As an example, we present here a model which is accurate over an 80-GHz bandwidth. For this model, the expansion points [σ_i's in (40)] are chosen to be purely imaginary at the following frequencies (with number of moments matched): 0 GHz (2), ± 10 GHz (1), ± 20 GHz (2), ± 30 GHz (1), ± 39 GHz (1). The dimension of this new model is 24, in contrast with 16 for RDM 6 in Table I. In Fig. 5, the residual error for RDM 6 and the new model with wider bandwidth are shown for a particular realization of the full model. As it can be observed in this figure, by using a larger number of expansions points distributed over a wider frequency range, the accuracy bandwidth of the models produced by our methodology can be enlarged at the cost of an increase in model dimension.

D. Results Over Many Fiber Realizations

In order to demonstrate the statistical performance of the model reduction methodology, we have applied the technique individually to 40000 randomly generated full models with 200 sections and a mean DGD of 10 ps. The results we report in Fig. 6 are for RDM 6 in Table I, where the mean and standard deviation of the residual error of (73) computed over 40000 realizations are shown. In Fig. 6, we also provide a curve for mean plus three times the standard deviation as an indicator of the statistical confidence level for the reduced model. As seen here, the reduced model is accurate not only for one particular realization of the full model but for all of the realizations over the whole frequency range of interest.

E. Poles of the Reduced Models

For the full model, the pole positions (all real) for the sections [i.e., α_k in (23)] are fixed, all equal to each other and they result from a first-order rational Pade approximation for the DGD of a section, as shown in (17). It would be interesting to see the pole positions for the reduced dimension model, which can be computed using the factorization algorithm described in Section IV-D. In Fig. 7, the poles of reduced model RDM 6 over 40000 realizations are shown. We observe that all of the poles are in the left half plane, but they are complex and the poles for the different sections in the model are nonequal and distributed roughly on an half ellipse centered at the origin. The poles of the reduced model exhibit slight variations from realization to realization. However, the median positions for the poles are in

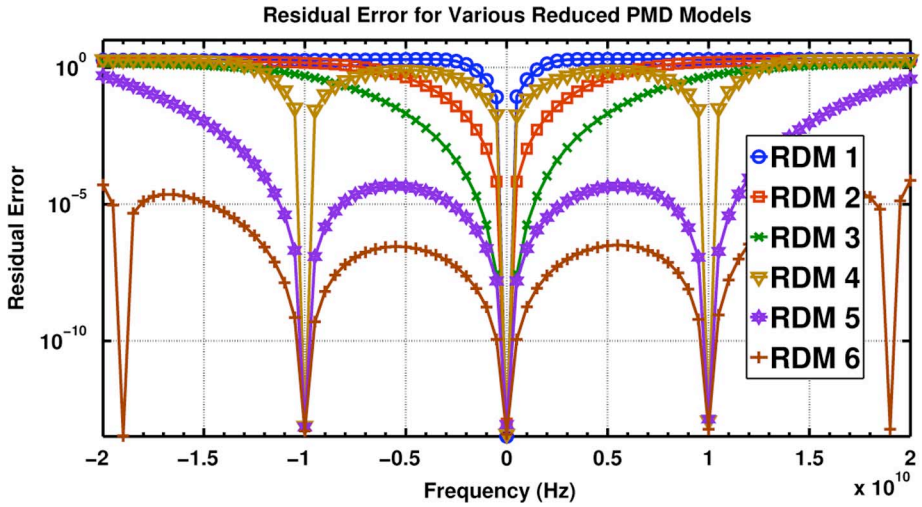

Fig. 2. Log error for the reduced models in Table I: One fiber realization.

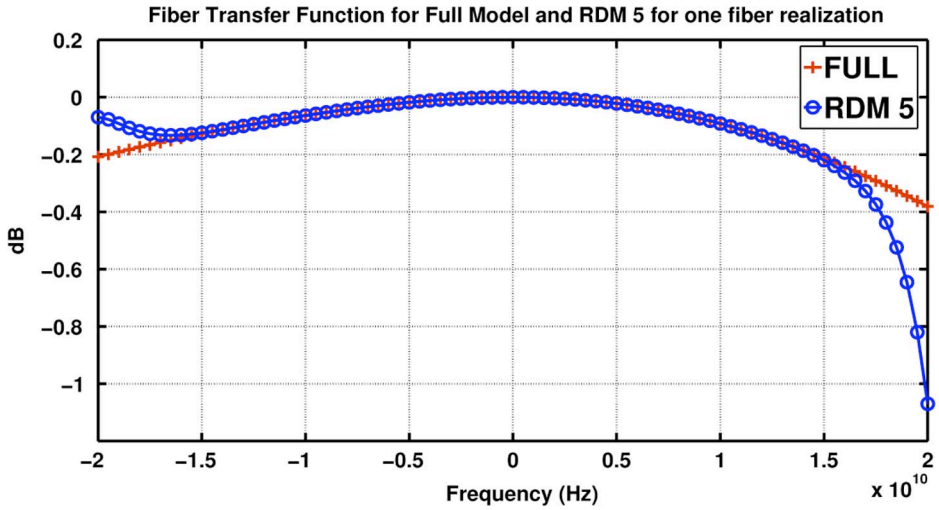

Fig. 3. Transfer function for full model and RDM 5: One fiber realization.

proximity to the pole locations indicated by white stars in Fig. 7. These starred pole locations were determined by computing the poles of a multipoint Pade approximation [11] for an exponential function given by

$$\mathbb{H}_D(s) = \exp\left(-s\mathrm{DGD}_{\mathrm{MAX}}\right) \tag{75}$$

where $\mathrm{DGD}_{\mathrm{MAX}} = N\tau$. The above represents the transfer function for a pure delay system for the total differential delay of the entire fiber (corresponding to the maximum worst-case total DGD). The expansion points and the number of moments matched at each one of the expansion points for the multipoint Pade approximation of (75) were chosen as for RDM 6 given in Table I. A little bit of tinkering reveals that the resemblance between the poles of the reduced PMD model and the ones for the Pade approximation of a pure delay system for the total maximum differential fiber delay is not surprising. The determinant of the fiber transfer function in (12)

$$\det[\mathbb{H}(s)] = \left[\frac{1 - s\tau/2}{1 + s\tau/2}\right]^N \approx \exp\left(-sN\tau\right) \tag{76}$$

is, in fact, equal to the pure delay in (75). When we try to approximate the transfer function $\mathbb{H}(s)$, we are in effect also approximating its determinant. Even though the poles of the reduced model are in proximity to the poles of the Pade approximation of the determinant in (76), they are not identical. The reason behind this is somewhat subtle. Given a number of poles in the Pade approximation, the number of moments one can match in computing the Pade approximation of a 2×2 transfer function is half as much as the moments that can be matched for a scalar (i.e., the determinant) transfer function [7]. As such, the poles of the reduced model exhibit slight realization dependence, but the distribution of the poles roughly on an ellipse is realization independent.

VI. Statistical Properties of Reduced PMD Models

In our discussion of reduced complexity modeling of PMD so far, we have focused on the accuracy of the obtained reduced model for a given random realization of the full model. One could in principle apply the model reduction technique we have developed to many realizations of the full model and obtain realizations of the reduced model in the probability space, as we

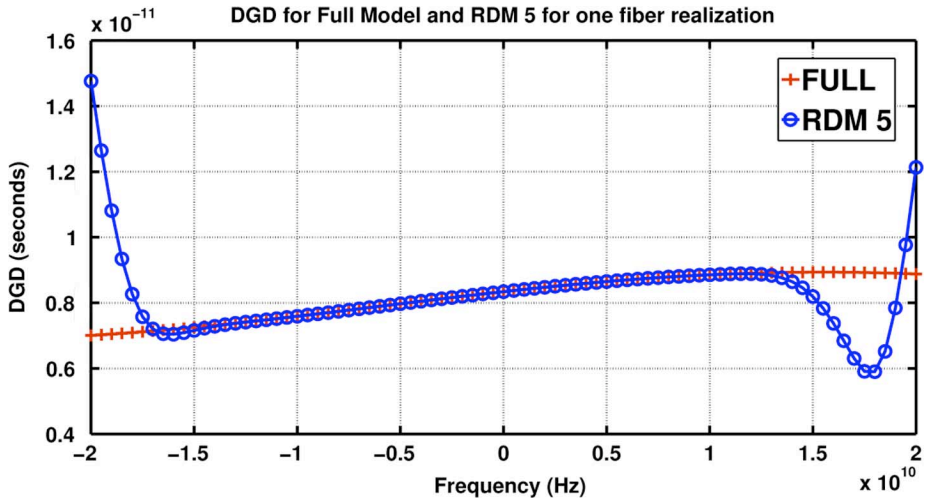

Fig. 4. DGD for full model and RDM 5: One fiber realization.

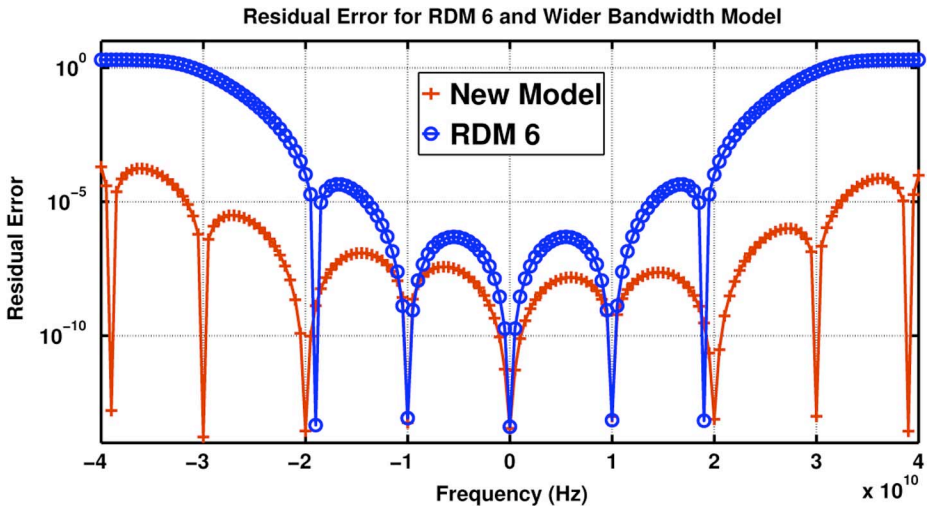

Fig. 5. Log error for RDM 6 and model with wider accuracy bandwidth.

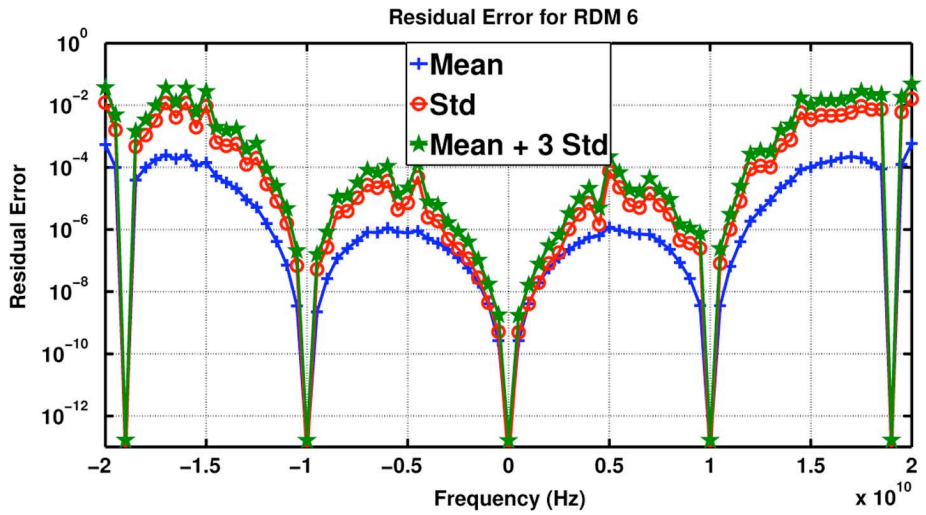

Fig. 6. Log error for RDM 6: Statistics over 40000 fiber realizations.

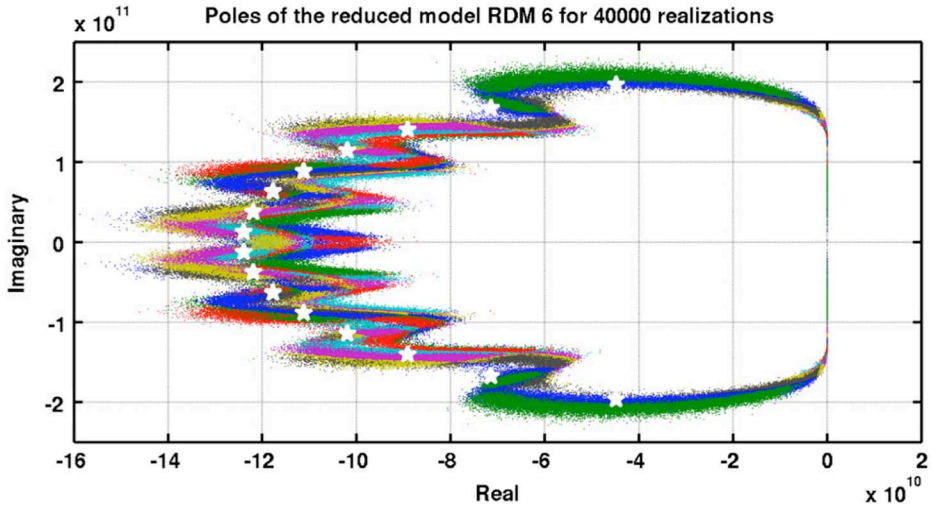

Fig. 7. Poles of RDM 6 for 40000 fiber realizations.

have done to generate the results reported in Section V-D. However, it would be best if different statistical realizations of the reduced model can be obtained directly, with a random parameterization of the reduced model itself, without having to apply the model reduction process on many realizations of the full model. In this section, we first investigate the statistical properties of the reduced models we obtain, and then describe how they may be parameterized for the generation of desirable PMD statistics as it can be done with the full model. Before we start our investigation into the statistical properties of the reduced complexity PMD models produced by our methodology, we make the following conjecture: If the reduced models are accurate in the sense of matching/inverting the Jones transfer function matrix of the full model (for a frequency range of interest) over the whole statistical PMD space, then it will be possible to form a parameterization for the reduced model (with associated probability distributions for the random parameters) so that it can be used directly to generate the correct statistical PMD behavior same as in the full model. The validity of this conjecture implies that the problem we are addressing here indeed has a solution. Our goal is then to figure out the proper parameterization and the probability distribution of the parameters in the reduced model.

A. DGD for One Section

In developing the PMD model reduction methodology in this paper, we have used a coupled mode theory and transfer function (Jones) matrix formalism. The statistical properties of PMD models are most often studied using a Stokes space formalism, based on the so-called PMD vector and the associated concatenation rules [4]. The length of the PMD vector gives us the DGD of the fiber. In order to perform a statistical characterization of our reduced complexity PMD models, we will use the PMD vector formalism and the concatenation rules. The PMD vector of a section of fiber can be easily obtained from its transfer function matrix [4]. In particular, we will be interested in the length of the PMD vector that corresponds to the transfer function of a section in both the full model in (23)/(20) and the reduced model in (64)/(65). Since the transfer function in (23)/(20) is a special case of the one in (64)/(65) (poles have zero imaginary part), we next compute the length of the PMD vector that corresponds to the transfer function in (64)/(65). The length of the PMD vector for any lossless transfer function matrix $\mathbb{H}(\omega)$ [that satisfies (19)] can be computed using

$$\mathrm{DGD}(\omega) = |\mathtt{Im}\{\lambda_1[\mathbb{G}(\omega)] - \lambda_2[\mathbb{G}(\omega)]\}| \tag{77}$$

where

$$\mathbb{G}(\omega) = \left[\frac{\mathrm{d}\mathbb{H}(\omega)}{\mathrm{d}\omega}\right]\mathbb{H}(\omega)^{\dagger} \tag{78}$$

and $\lambda_k[\cdot]$ denotes the kth eigenvalue of the matrix in the argument. The form of the transfer function in (65) makes it quite trivial to evaluate (78)

$$\begin{aligned}\mathbb{G}_o(\omega) &= \left[\frac{\mathrm{d}\mathbb{H}_o(\omega)}{\mathrm{d}\omega}\right]\mathbb{H}_o(\omega)^{\dagger} \\ &= \mathbf{U}^{\dagger}\begin{bmatrix} 0 & 0 \\ 0 & \frac{(j\,\omega+\alpha+j\,\beta)}{(-j\,\omega+\alpha-j\,\beta)}\frac{\mathrm{d}}{\mathrm{d}\omega}\frac{(-j\,\omega+\alpha-j\,\beta)}{(j\,\omega+\alpha+j\,\beta)} \end{bmatrix}\mathbf{U}\end{aligned} \tag{79}$$

and (77) to obtain the DGD for one section

$$\begin{aligned}&\mathrm{DGD}_s(\omega) \\ &\quad= \left|\mathtt{Im}\left\{\frac{(j\,\omega+\alpha+j\,\beta)}{(-j\,\omega+\alpha-j\,\beta)}\frac{\mathrm{d}}{\mathrm{d}\omega}\left[\frac{(-j\,\omega+\alpha-j\,\beta)}{(j\,\omega+\alpha+j\,\beta)}\right]\right\}\right| \\ &\quad= \frac{2\alpha}{\alpha^2+(\omega+\beta)^2}.\end{aligned} \tag{80}$$

We observe that the length of the PMD vector in (80) corresponding to the transfer function in (65) for a section is frequency (ω) dependent. Moreover, the DGD value in (80) is independent of the unitary, constant matrix $\mathbf{U}$ in (65). $\mathbf{U}$ influences only the direction of the PMD vector for a section, and the DGD of a section is determined only by the real (α) and the imaginary (β) parts of the section pole.

B. Random Parameterization of the Full Model

The DGD in (80) is for one section of the fiber transfer function. The DGD value corresponding to a cascade of a number of

such sections can be computed using the PMD vector concatenation rules [4]. The PMD vector for the cascade of a number of sections is obtained by vectorially adding the PMD vectors corresponding to the sections in the cascade. In this case, the directions as well as the lengths of the PMD vectors are needed. In PMD modeling, it has been usual practice to choose the directions of the PMD vectors for the sections in a uniformly random manner and independent of each other. Uniform randomization of the PMD vector directions can be realized by a proper random parameterization of the full PMD transfer function in (12) that can be factorized as the cascade of degree-one transfer functions as in (20) and (23). The random parameterization of PMD models has been analyzed and discussed extensively in the literature. Based on previous work [33], we perform the randomization of the full PMD transfer function in (12) as follows. We generate the unit norm vector $\mathbf{u}_k$ in (23) with

$$\mathbf{u}_k = [\sqrt{\gamma_k}\exp(j\,\theta_k) \quad \sqrt{1-\gamma_k}\exp(j\,\phi_k)]^T \tag{81}$$

where γ_k is a uniformly distributed random number in the range $[0,1]$, and θ_k and ϕ_k are uniformly distributed random numbers in the range $[0, 2\pi]$. γ_k, θ_k and ϕ_k are independent not only of each other but also the ones for other values of k, i.e., for other sections in the cascaded transfer function. α_k in (20) is not a random parameter and it is chosen deterministically. In the full PMD model, we choose all α_k's to be equal to each other for all sections, based on (18) and (24). The random parameterization of $\mathbf{u}_k$ as in (81) above makes sure that the *direction* of the PMD vector corresponding to one section in (20) has a uniform distribution in three dimensions, i.e., tip of the vector is uniformly random on a sphere.

C. DGD Distribution for the Full Model

The probability distribution of DGD for the full model can be determined based on the fact that the total DGD for a number of cascaded sections is given by the distance of a point in space that is reached after an isotropic random walk in the 3-D space starting from the origin [33]. For the full model, the DGD's of the sections in the cascade are identical and given by

$$\mathrm{DGD}_s(\omega) = \frac{2\alpha}{\alpha^2+\omega^2} = \frac{\tau}{1+\left(\frac{\omega\tau}{2}\right)^2} \tag{82}$$

which follows from (24) and (80), with τ as in (18). The DGD value above is frequency dependent, but this dependence is weak. For frequencies satisfying $\omega\tau \ll 1$, the section DGD is approximately equal to τ. The accuracy of this approximation can be improved by using more sections in the full model (since τ decreases when more sections are used in the model as discussed below), or by narrowing the frequency range of interest. For the full model, the probability distribution of total DGD can be computed using the results in [33], which considers the problem of computing the distribution of the length of a vector in three dimensions that is obtained by adding a number of vectors of equal length but having uniformly random directions. The largest total DGD that can be obtained from the full model is given by

$$\max \mathrm{DGD}_{\mathrm{full}} = \tau N = \Delta z\,\Delta\beta N = \frac{L}{N}\Delta\beta N = L\Delta\beta \tag{83}$$

where N is the number of sections in the full model and L is the total fiber length. Even though the exact expression for the mean of $\mathrm{DGD}_{\mathrm{full}}$ is complicated, a simple expression for the root mean square (RMS) value of $\mathrm{DGD}_{\mathrm{full}}$ can be obtained

$$\sqrt{\mathrm{E}\left[\mathrm{DGD}_{\mathrm{full}}^2\right]} = \tau\sqrt{N} \tag{84}$$

where $\mathrm{E}[\cdot]$ denotes expectation. The mean and RMS values of DGD are quite close to each other, mean value is approximately $\sqrt{8/3\,\pi} \approx 0.92$ times the RMS value, which is the ratio between the mean and RMS value of a Maxwellian distribution. When this model is used as a PMD emulator, a specification for mean (or RMS) DGD is usually given. τ in (84) is given by (18), i.e., $\tau = \Delta z\,\Delta\beta = L/N\Delta\beta$. Thus, the RMS (and mean) total DGD for a fiber of fixed length L decreases with the number of sections N, if $\Delta\beta$ is held constant. Hence, $\Delta\beta$ is made proportional to $\sqrt{N}$ in order to attain a mean/RMS DGD that is independent of N for a fixed fiber length L. The ratio of the maximum DGD that can be obtained from the model to the RMS value is simply given by

$$\frac{\max \mathrm{DGD}_{\mathrm{full}}}{\sqrt{\mathrm{E}\left[\mathrm{DGD}_{\mathrm{full}}^2\right]}} = \sqrt{N}. \tag{85}$$

When mean (RMS) DGD is held constant as a function of N as discussed above, the maximum DGD increases when more sections are used in the model. In the limit when $N \to \infty$, the DGD distribution asymptotes to a Maxwellian distribution with no bound on maximum DGD. As shown in [33], the distribution of DGD obtained with the full model can be approximated well with a Maxwellian distribution even for finite values of N, but only for values of DGD below the maximum value in (83). As argued in [33] and [34], there is a lower limit on the number of sections N that is needed if one would like to use the model as a PMD emulator in estimating outage probabilities smaller than 10^{-5}.

D. Pole Positions of the Reduced Model

In Section IV-D, we have shown that the reduced transfer function matrix for PMD can be factorized in a very similar manner as the full transfer function. Hence, one would be tempted to form a random parameterization for the reduced model in exactly the same manner. However, the question remains on how to choose or parameterize the complex poles of the reduced transfer function. Based on the results presented in Section V-E, we have observed that the positions of the complex poles of the reduced transfer functions are in proximity to the poles of a rational transfer function obtained as the multipoint Pade approximation (having the same dimension n as the reduced transfer function) for a pure delay system with delay equal to the total maximum DGD of the full model. Even though the positions of the complex poles for the reduced transfer function of a fiber vary from realization to realization, the variation is relatively small and the poles stay in proximity to the poles of the Pade approximation mentioned above, and, hence, we choose to use the complex poles of this Pade approximation as the fixed poles for the sections in our reduced

model. At first glance, it may seem that, by fixing the complex pole positions of the reduced model at the pole locations of the Pade approximation, we are essentially making the reduced and full models equivalent. However, even if there is a structural similarity between the full and reduced models, there are three key differences between them.

- The poles of the full model are all real, but the ones for the reduced model are complex.
- The poles for the sections in the full model are all equal to each other, the ones for the reduced model are different for the sections.
- The fixed real pole of the full model is obtained by forming a first-order rational Pade approximation for the DGD of one section. In the case of the reduced model, the complex pole positions are obtained by forming a rational Pade approximation for the total maximum DGD obtainable from the full model, with Pade approximation order equal to the number of sections n in the reduced model.

These differences are in fact the reasons behind the effectiveness and, in some sense, the optimality of the reduced model when compared with the full model. By keeping the same structure as cascaded degree-one sections in going from the full model to the reduced model but at the same time allowing the poles for the sections to be complex and different from each other, we are in some sense maximizing the expressiveness of the PMD model and enabling the use of much fewer sections for emulation and inversion purposes.

E. Statistical Properties of the Reduced Model

Next, we investigate the statistical properties of the reduced model with section poles fixed at the pole locations obtained from a Pade approximation. The issue to be addressed here and in the next section is how to form a random parameterization of $\mathbf{u}_k$ in (64) so that the reduced model not only matches the full model (in the sense of matching the transfer function for a frequency range of interest) for a given realization of the random parameters in the full model, but it also reproduces the same statistical behavior over many realizations in the probability space.

First, we observe that no matter what kind of a random parameterization is chosen for $\mathbf{u}_k$ in (65), the maximum DGD that can be obtained from a reduced model with dimension n can be computed by simply adding the DGD's of all the sections using (80) with the poles obtained from the Pade approximation discussed above. This corresponds to setting $\mathbf{U}$ in (65) for all of the sections to the identity matrix, essentially equivalent to aligning the directions of the PMD vectors of the sections. In this case, even though the DGD's of the individual sections as given by (80) are frequency dependent, the total maximum DGD obtainable from the reduced model is frequency independent and equal to the maximum DGD obtainable from the full model. This should not be surprising, because the poles of the reduced order model are obtained from a Pade approximation of a continuous-time delay equal to the total maximum DGD obtainable from the full model. In fact, when all $\mathbf{U}$ in (65) are set to the identity matrix and the transfer function matrices for n sections are cascaded to form the overall transfer function, one obtains a diagonal matrix, with the first diagonal entry equal to 1 and the second diagonal entry equal to a rational function that is equal to the nth order Pade approximation of total maximum DGD. In Fig. 8, the frequency dependent section DGD's are shown along with the total maximum DGD (153.5 ps) obtained for ROM 6 in Table I with $n = 16$, over a frequency range of 40 GHz. In this figure, the total maximum DGD for the full model with $N = 200$ sections is also shown (overlaps with the total DGD curve for the reduced model), computed by simply multiplying the section DGD in (82) with N. As seen in Fig. 8, the DGD's of individual sections exhibit significant variation as a function of frequency, but the total maximum DGD is frequency independent. In contrast with the reduced model, the section DGD's for the full model are frequency independent and identical.

As mentioned before, one is tempted to form a random parameterization for the reduced model in exactly the same manner, i.e., by randomizing $\mathbf{u}_k$ in (65) as in (81) and by employing uniformly distributed and independent parameters γ_k, θ_k, and ϕ_k. The distribution and statistical properties of DGD in this case can be evaluated again based on a random flight model in three dimensions. However, one cannot use the results in [33], since the section DGD's in the reduced model are not identical. In this case, the results obtained in [34] may be used, where the authors consider the problem of computing the distribution of the length of a vector in three dimensions that is obtained by adding a number of vectors of varying (but still nonrandom) lengths that have uniformly random directions. For the reduced model, the distribution and statistical properties of total DGD needs to be evaluated separately at all frequencies of interest. The mean, standard deviation and PDF of DGD as a function of frequency are shown in Figs. 9 and 10, which have been computed using the formulas derived in [34]. As seen in these figures, the mean, standard deviation and PDF of total DGD exhibit variations as a function of frequency. Even though the shape of the PDF is same at all frequencies, the mean and standard deviation is smaller at lower frequencies and larger at higher frequencies. This frequency dependent statistical behavior we observe here for DGD is undesirable. It is believed that DGD of a fiber possesses an ergodicity property, i.e., the distribution of DGD obtained with statistical data collected at only one frequency is in fact equal to the DGD distribution one would obtain by collecting only one data point at every frequency point but at many points over a wide frequency range. This in fact means that DGD is a stationary stochastic process when it is considered as a function of frequency. There is a second issue with the results we obtained on the statistical properties of the reduced model. The mean of total DGD shown in Fig. 9 is not only frequency dependent (unlike in the case of the full model), but it is also larger than the mean DGD one obtains from the corresponding full model (10 ps in this case).

F. Random Parameterization of the Reduced Model

The results we have presented in Section V showed that the reduced complexity PMD models we produce are accurate in the sense of matching/inverting the Jones transfer function matrix of the full model (for a frequency range of interest) over the whole statistical PMD space. This seems to be in conflict with the statistical properties of the reduced models we have obtained just

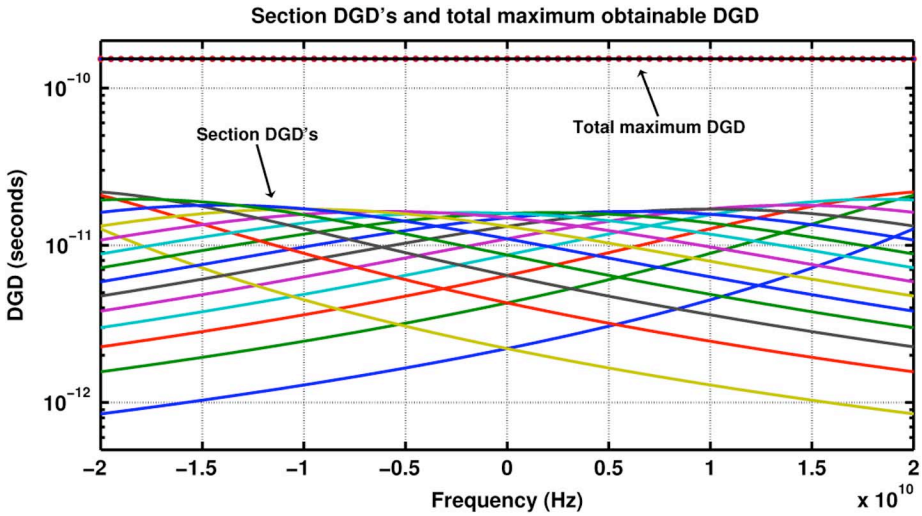

Fig. 8. Section DGD's and total maximum DGD for the reduced model.

above. The statistical properties of the full model are in fact frequency independent as we have shown in Section VI-C. If the reduced model is accurate in the sense of matching the full model, then it should not only exhibit frequency independent statistical behavior but also produce the same mean DGD as in the full model. The resolution of this conflict lies in reconsidering the manner we have performed the random parameterization of the reduced model. Based on the above results, we may conclude that the random parameterization of the reduced model has to be performed in a different way in order for it to exhibit frequency independent statistical behavior and produce the same mean DGD as the full model. In order to gain insight into how one should proceed, we have done the following experiment. We have obtained reduced models from many randomly generated full models using our model reduction technique. Then each reduced model was factorized using the technique described in Section IV-D in order to compute the $\mathbf{u}_k$ in (65) and the parameters γ_k, θ_k, and ϕ_k based on (81) for the sections. Next, statistical properties of these parameters were obtained based on computing averages over many reduced models. The results of this experiment showed that γ_k, θ_k, and ϕ_k are uniformly distributed as in the case of the full model, but they are *correlated* with each other. In the random parameterization of the full model, these parameters are generated independently. Thus, in order for the reduced model to exhibit the same frequency independent statistical properties as the full one, its random parameterization has to be done more carefully by determining the correlation structure among the random parameters γ_k, θ_k and ϕ_k and their joint probability distribution. A "little bit" of tinkering reveals that the determination of the joint PDF for these correlated parameters is a nontrivial problem, and it can be solved only in some very special cases. Fortunately, one can draw samples from this joint PDF, without knowing it, by using Markov chain Monte Carlo (MCMC) sampling and the Metropolis algorithm [35]–[37]. Even though the joint PDF for the random parameters of the reduced model is unknown, the desired frequency independent PDF for DGD (dictated by the full model) is known. This desired DGD PDF can be prescribed as the limiting distribution for the Markov chain in MCMC sampling by properly formulating the Metropolis rejection rule [36], [37]. Since MCMC and related random sampling techniques such as the multicanonical Monte Carlo method have been used and extensively discussed in the PMD literature, we provide only a brief description of how we perform MCMC sampling on our reduced complexity PMD models. We start at a randomly chosen point in the random parameter space of the reduced model, i.e., a set of values for γ_k, θ_k, and ϕ_k for all of the sections. A Metropolis step involves proposing a candidate sample with $\hat{\gamma}_k = \gamma_k + \delta\gamma_k$, $\hat{\theta}_k = \theta_k + \delta\theta_k$ and $\hat{\phi}_k = \phi_k + \delta\phi_k$ where $\delta\gamma_k$ is uniformly distributed in the range $[-0.1, 0.1]$, and $\delta\theta_k$ and $\delta\phi_k$ are uniformly distributed in the range $[-\pi/20, \pi/20]$ [35], [37]. The candidate sample is then accepted as the next sample with probability $\min\{1, p(\hat{d})/p(d)\}$, otherwise the next sample is set to the previous sample. Here, d and $\hat{d}$ are the DGD values at the current and candidate sample points, and $p(\cdot)$ is the desired PDF for the DGD. The Metropolis algorithm generates samples from the desired DGD distribution by exploring the probability space using a random walk. It can be shown that the desired distribution $p(\cdot)$ is in fact the limiting, steady-state probability distribution of the underlying Markov chain implied by the random walk process. The form of the Metropolis rejection rule above is appropriate for obtaining samples with the desired DGD distribution at a single, chosen frequency. In order to attain the same desired DGD distribution over a range of frequencies, we modify the acceptance probability as

$$\min\left\{1, \frac{\prod_{i=1}^{F} p(\hat{d}(f_i))}{\prod_{i=1}^{F} p(d(f_i))}\right\} \tag{86}$$

where f_i's are a set of frequencies evenly distributed in the frequency range of interest, well separated from each other so that DGD at two consecutive frequency points is approximately uncorrelated.

We now present results we have obtained by using the MCMC sampling technique on our reduced models. The mean and standard deviation of DGD obtained with the MCMC method as a function of frequency are shown in Fig. 11. These

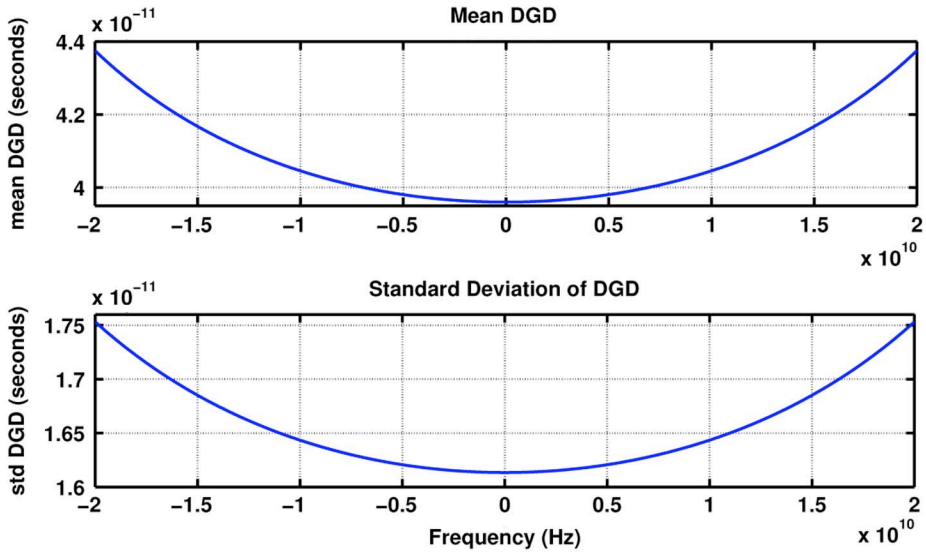

Fig. 9. Mean and standard deviation of DGD for the reduced model.

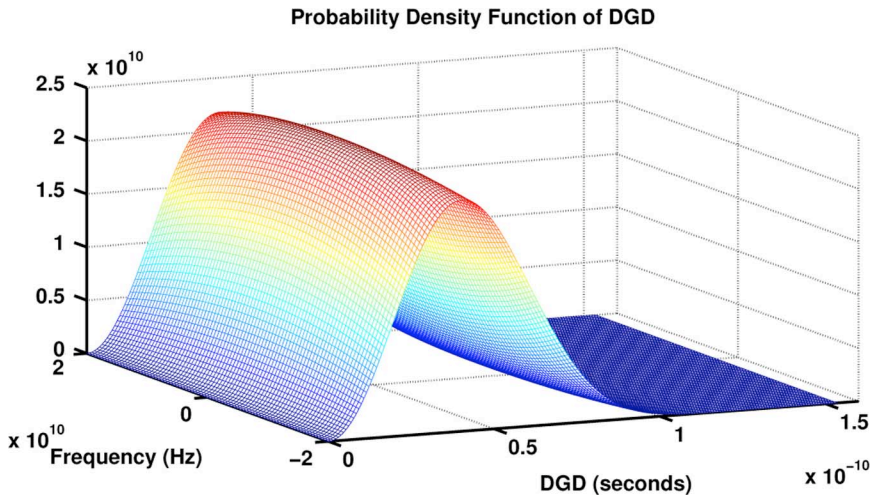

Fig. 10. Probability density function of DGD for the reduced model.

two graphs also contain the mean and standard deviation of DGD obtained using a standard random parameterization for comparison, the same curves from Fig. 9. MCMC simulation was performed with 2010000 samples, and the first 10000 samples were discarded in order to remove the initial transient from the Markov chain. For the results in Fig. 11, three frequencies ($f = 0, 10, 20$ GHz) were used in computing the acceptance probability in (86) for the rejection rule in the Metropolis algorithm. The desired DGD distribution $p(\cdot)$ was chosen as the DGD distribution of the full model. The MCMC sampling technique (with the specific rejection rule described above) enables us not only to generate any desirable DGD statistics from the reduced model, but also make DGD statistics frequency independent. For comparison, in Fig. 12, we also provide results on DGD statistics that we have obtained with the MCMC technique when only one frequency ($f = 10$ GHz) is used in computing the acceptance probability in (86). This figure demonstrates the effectiveness of the rejection rule based on (86) with a set of frequencies evenly distributed in the frequency range of interest in obtaining the desired frequency independent DGD statistics.

One key feature of the reduced models produced by our methodology is that the maximum DGD that can be obtained from the reduced model is the same as the maximum DGD obtainable from the full model. Moreover, one can obtain the same DGD statistics as in the full model using the MCMC technique as described above. As such, the ratio of maximum DGD to mean DGD is preserved during model reduction. This is important, because this ratio (larger is better) determines the capability of a PMD model in predicting very small outage probabilities [33], [34]. In contrast, if one generates a small model by simply reducing the number of birefringent sections in the usual PMD model that is widely used, the maximum to mean DGD ratio is not preserved, and in fact, it is reduced since this ratio is proportional to the square root of the number of sections used in the model, as given by (85). The model reduction methodology proposed in this paper removes this important restriction, and makes it possible to predict very

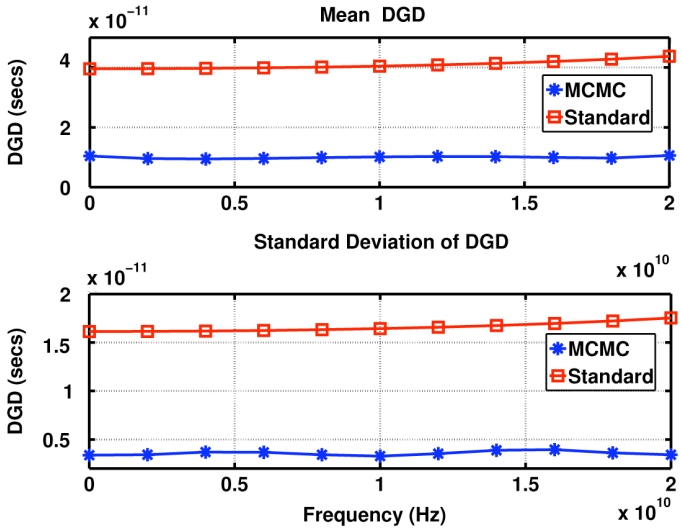

Fig. 11. DGD statistics with MCMC: rejection rule with three frequencies.

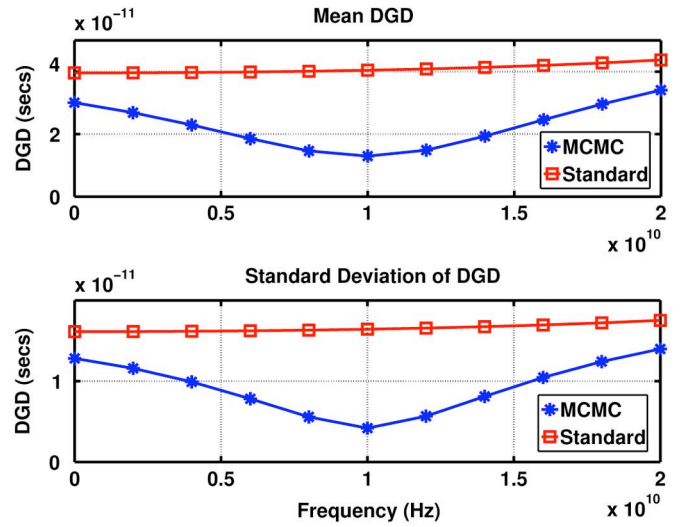

Fig. 12. DGD statistics with MCMC: rejection rule with one frequency.

small outage probabilities even with PMD models composed of a small number of sections.

VII. Summary and Conclusion

We have developed a formalism and methodology for producing compact, low complexity PMD models, by adapting well established model reduction techniques that have been successfully used for large electronic systems. The compact PMD models produced by this methodology can be implemented in software for computational studies. They can also be implemented in electronic hardware as a digital or continuous-time lumped filter for the emulation and compensation of PMD. The contributions of the work described in this paper are two-pronged.

- *A systematic, rigorous model complexity reduction methodology* for PMD that is based on obtaining multipoint Pade approximations of the Jones transfer matrix of the fiber in a transparent manner. The flexibility offered by the choice of the expansion points for the multipoint Pade approximation is a key feature of the method proposed, enabling one to obtain the compactest PMD model that is accurate over a specifiable frequency bandwidth.
- *A novel, compact PMD model* in the form of a continuous-time lumped filter with complex poles that come from the Pade approximation of a desirable total worst-case DGD. This compact model can generate the same statistical behavior as in a full traditional PMD model composed of many birefringent fiber sections, including the same worst-case to mean DGD ratio. The model complexity reduction methodology developed in this paper was the key enabler in discovering this new model.

Finally, the new results we have obtained and the techniques we have developed in this paper on preserving losslessness in model dimension reduction and the factorization of lossless transfer functions are general and not specific to PMD models.

References

[1] G. P. Agrawal, *Fiber-Optic Communication Systems*. New York: Wiley, 2002.
[2] C. D. Poole and R. E. Wagner, "Phenomenological approach to polarisation dispersion in long single-mode fibres," *Electron. Lett.*, vol. 22, no. 19, pp. 1029–1030, Sep. 1986.
[3] C. D. Poole and J. A. Nagel, "Polarization effects in lightwave systems," in *Optical Fiber Telecommunications IIIA*, I. P. Kaminov and T. L. Koch, Eds. New York: Academic, 1997, pp. 114–161.
[4] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," in *Proc. Nat. Acad. Sci. USA*, Apr. 2000, vol. 97, no. 9, pp. 4541–4550.
[5] H. Kogelnik, L. E. Nelson, and R. M. Jopson, "Polarization mode dispersion," in *Optical Fiber Telecommunications IVB*, I. P. Kaminov and T. Li, Eds. New York: Academic, 2002, pp. 725–861.
[6] H. Kogelnik, L. E. Nelson, and J. P. Gordon, "Emulation and inversion of polarization-mode dispersion," *J. Lightw. Technol.*, vol. 21, no. 2, pp. 482–495, Feb. 2003.
[7] R. W. Freund, "Krylov-subspace methods for reduced-order modeling in circuit simulation," *J. Comput. Appl. Math.*, vol. 123, no. 1–2, pp. 395–421, Nov. 2000.
[8] R. Marz, *Integrated Optics: Design and Modeling*. Norwell, MA: Artech House, 1994.
[9] M. R. Wohlers, *Lumped and Distributive Passive Networks*. New York: Academic, 1994.
[10] D. Yevick, M. Chanachowicz, M. Reimer, M. O'Sullivan, W. Huang, and T. Lu, "Chebyshev and Taylor approximation of polarization mode dispersion for improved compensation bandwidth," *J. Opt. Soc. Amer. A*, vol. 22, no. 8, pp. 1662–1667, Aug. 2005.
[11] G. A. Baker, Jr. and P. Graves-Morris, *Pade Approximants*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1996.
[12] E. J. Grimme, "Krylov projection methods for model reduction," Ph.D. dissertation, Univ. Illinois at Urbana-Champaign, Urbana, IL, 1994.
[13] S. J. Savory and F. P. Payne, "Pulse progation in fibers with polarization mode dispersion," *J. Lightw. Technol.*, vol. 19, no. 3, pp. 350–357, Mar. 2001.
[14] G. P. Agrawal, *Nonlinear Fiber Optics*. New York: Academic, 1995.
[15] N. K. Bose and A. Fettweis, "Skew symmetry and orthogonality in the equivalent representation of a time-varying multiport inductor," *IEEE Trans. Circuits Syst.—I: Fundam. Theory Appl.*, vol. 51, no. 7, pp. 1321–1329, Jul. 2004.
[16] B. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Control*, vol. 26, no. 1, pp. 17–32, Feb. 1981.
[17] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds," *Int. J. Control*, vol. 39, no. 6, pp. 1115–1193, Jun. 1984.
[18] L. Sirovich, "Turbulence and the dynamics of coherent structures. I—Coherent structures. II—Symmetries and transformations. III—Dynamics and scaling," *Quart. Appl. Math.*, vol. 45, pp. 561–571, 573–590, Oct. 1987.
[19] L. T. Pillage and R. A. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 9, no. 4, pp. 352–366, Apr. 1990.

[20] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by Pade approximation via the lanczos process," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 14, no. 5, pp. 639–649, May 1995.
[21] J. R. Li, "Model reduction of large linear systems via low rank system gramians," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, 2000.
[22] A. C. Antoulas, "Summary of recent results on the decay rate of Hankel singular values," in *Proc. 40th IEEE Conf. Decision and Control*, Dec. 2001, vol. 4, pp. 3655–3656.
[23] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*. Philadelphia, PA: SIAM, 1997.
[24] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*. Manchester, U.K.: Manchester Univ. Press, 1992.
[25] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. Philadelphia, PA: SIAM, 2003.
[26] R. W. Freund and F. Jarre, "An extension of the positive real lemma to descriptor systems," *Optim. Meth. Softw.*, vol. 19, no. 1, pp. 69–87, Feb. 2004.
[27] A. C. Antoulas, "A new result on passivity preserving model reduction," *Syst. Control Lett.*, vol. 54, pp. 361–374, 2005.
[28] V. Belevitch, *Classical Network Theory*. New York: Holden-Day, 1968.
[29] B. D. O. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis: A Modern Systems Theory Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
[30] T. Genin, P. Van Dooren, T. Kailath, J.-M. Delosme, and M. Morf, "Sigma-lossless transfer functions and related questions," *Linear Algebra Appl.*, vol. 50, pp. 251–275, 1983.
[31] B. N. Bond and L. Daniel, "Stabilizing schemes for piecewise-linear reduced order models via projection and weighting functions," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design*, Nov. 2007, pp. 860–866.
[32] G. H. Golub and C. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[33] M. Karlsson, "Probability density functions of the differential group delay in optical fiber communication systems," *J. Lightw. Technol.*, vol. 19, no. 3, pp. 324–331, Mar. 2001.
[34] C. Antonelli and A. Mecozzi, "Statistics of the DGD in PMD emulators," *IEEE Photon. Technol. Lett.*, vol. 16, no. 8, pp. 1840–1842, Aug. 2004.
[35] D. Yevick, "Multicanonical communication system modeling—Application to PMD statistics," *IEEE Photon. Technol. Lett.*, vol. 14, no. 11, pp. 1512–1514, Nov. 2002.
[36] M. Secondini and E. Forestieri, "All-order PMD outage probability evaluations by Markov chain Monte Carlo simulations," *IEEE Photon. Technol. Lett.*, vol. 17, no. 7, pp. 1417–1419, Jul. 2005.
[37] A. Bilenca, A. Desjardins, B. E. Bouma, and G. J. Tearney, "Multicanonical Monte-Carlo simulations of light propagation in biological media," *Opt. Exp.*, vol. 13, no. 24, pp. 9822–9833, Nov. 2005.

Alper Demir (S'93–M'96) received the B.S. degree in electrical engineering from Bilkent University, Ankara, Turkey, in 1991, and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, in 1994 and 1997, respectively.

In the summer of 1995, he was at Motorola, in the summer of 1996 at Cadence Design Systems, from 1997 to 2000 at Bell Laboratories Research, from 2000 to 2002 at CeLight (start-up in optical communications), and, in the summer of 2002 and in August 2005, at the Research Laboratory for Electronics at the Massachusetts Institute of Technology. He was with the Department of Electrical and Electronics Engineering, Koc University, Sariyer-Istanbul, Turkey, as an Assistant Professor between February 2002 and December 2007, and he has been an Asssociate Professor since January 2008. His research interests are in computational prototyping of electronic and optoelectronic systems, numerical modeling and analysis, stochastic dynamical systems, and noise in nonlinear electronic, optical, communication, and biological systems. The work he did at Bell Laboratories and CeLight is the subject of six patents. He has coauthored two books in the areas of nonlinear-noise analysis and analog-design methodologies and published around 40 articles in journals and conferences.

Dr. Demir was the recipient of several best paper awards: the 2002 Best of International Conference on Computer-Aided Design (ICCAD) Award: 20 years of excellence in CAD, the 2003 IEEE/Association for Computing Machinery William J. McCalla ICCAD Best Paper Award, and the 2004 IEEE Circuits and Systems Society Guillemin-Cauer Best Paper Award. In 1991, he was the recipient of the Regents Fellowship and the Eugene-Mona Fay Gee Scholarship from the University of California, Berkeley, and was selected to be an Honorary Fellow of the Scientific and Technological Research Council of Turkey (TUBITAK). In 2003, he was selected by the Turkish Academy of Sciences to receive the Distinguished Young Scientist Award, in 2005, he won a TUBITAK Career Award, and in 2007, he received the TUBITAK Young Scientist Award.

Alper T. Erdogan (M'00) was born in Ankara, Turkey, in 1971. He received his B.S. degree from the Middle East Technical University, Ankara, in 1993, and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1995 and 1999, respectively.

He was a Principal Research Engineer with Globespan-Virata Corporation (formerly Excess Bandwidth and Virata Corporations) from September 1999 to November 2001. He joined Electrical and Electronics Engineering Department, Koc University, Sariyer-Istanbul, Turkey, in January 2002, where he is currently an Associate Professor.

Dr. Erdogan is the recipient of the TUBITAK Career Award (2005), the Werner Von Siemens Excellence Award (2007), and the Turkish Academy of Sciences Distinguished Young Scientist Award (2008). His research interests include wireless, fiber, and wireline communications, adaptive signal processing, optimization, system theory and control, and information theory.