This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

The Resolution Matrix in Tomographic Multiplexing: Optimization of Inter-parameter Cross-talk, Relative Quantitation and Localization

Steven S. Hou, Brian J. Bacskai, and Anand T. N. Kumar

Abstract*—*Objective: **We use a resolution matrix-based Bayesian framework to compare inversion methods for tomographic fluorescence lifetime multiplexing in a diffuse medium, such as biological tissue.** ***Methods:*** **We consider three inversion methods; an asymptotic time domain (ATD) approach, based on a multi-exponential analysis of time domain data, a direct time domain (DTD) approach, which is a minimum error solution, and a cross-talk constrained time domain (CCTD) inversion, which is a solution to an optimization problem that minimizes both error and cross-talk. We compare these methods using Monte Carlo simulations and time domain fluorescence measurements with tissue-mimicking phantoms.** ***Results:*** **The ATD approach provides high accuracy of relative quantitation and spatial localization of two fluorophores embedded in a 18-mm thick turbid medium, with concentration ratios of up to 1:4.25. DTD leads to significant errors in relative quantitation and localization. CCTD provides improved quantitation accuracy over DTD, and better spatial resolution compared to ATD. We present a rigorous theoretical basis for these results and provide a complete derivation of the CCTD estimator. The Bayesian analysis also leads to a formula for rapid computation of the DTD inverse operator for large-scale tomography measurements.** ***Conclusion:*** **The ATD and CCTD inversion methods provide significant advantages over DTD for accurately estimating multiple overlapping fluorophores.** ***Significance:*** **Time domain fluorescence tomography, using zero cross-talk estimators, can serve as a powerful tool for quantifying multiple fluorescently labeled biological processes. The Bayesian framework presented here can be applied to general multi-parameter inverse problems for the quantitative estimation of multiple overlapping parameters.**

***Index Terms*—Molecular imaging, fluorescence tomography, inverse problems, time-resolved imaging, lifetime multiplexing**

I. Introduction

The simultaneous, non-invasive detection of multiple biological components or processes (such as receptor expression, tumor growth, angiogenesis) through multiplexing can deepen our understanding about how these processes interact in a living complex biological system. Optical imaging offers the unique opportunity for multiplexing using a wide range of near infrared (NIR) fluorophores that can be conjugated with disease-specific molecular markers. Several NIR probes exhibit distinct fluorescence lifetimes and absorption and emission spectra, thereby allowing the capability for multiplexing using either spectral or fluorescence lifetime contrast.

Tomographic lifetime multiplexing (TFLM) belongs to a larger class of linear multi-parameter inverse problems (MPIPs) and can be solved with standard inversion techniques, such as the Tikhonov method. The Tikhonov approach produces a least squares solution with a constraint on the energy of the solution. In the more general Bayesian approach, prior information about the unknowns is described using a probability distribution. Through application of Bayes' rule, optimal solutions can be derived based on the posterior probability distribution for the unknowns [1].

While standard inversion techniques can minimize reconstruction error, they are not ideal for multi-parameter problems, since they do not directly account for inter-parameter cross-talk, an important source of error specific to MPIPs. Cross-talk refers to the interference among different parameters. Its effect can be illustrated most easily in TFLM by considering a medium where fluorophores with different lifetimes are separated spatially. When image reconstruction is performed, spatial regions which should only reconstruct for a single lifetime component will reconstruct for other lifetimes as well. The amount of cross-talk varies depending on the inherent non-uniqueness of the data with respect to different parameter distributions. For example, early time points on the time-domain (TD) fluorescence profile result in significantly higher cross-talk than late time points [2], [3], since early TD data are less sensitive to typical fluorescence lifetimes than late TD data. Similarly, continuous-wave (CW) fluorescence reconstructions result in 100% crosstalk since CW data cannot distinguish fluorescence lifetime from concentration [3], [4].

Cross-talk is important for TFLM since it directly leads to errors in localization and relative quantitation. Previously, we have shown using simulations [5] that the asymptotic time domain (ATD) approach, which is based on a multi-exponential analysis of the asymptotic (or decay) portion of TD data, provides minimal cross-talk and superior localization compared to the standard inversion of TD data (DTD), which is a minimum error solution. We also presented a crosstalk constrained TD (CCTD) approach for TFLM as a minimum error solution with zero cross-talk constraints on the resolution matrix. The CCTD estimator was shown to be identical in form to ATD, except for a diagonal covariance matrix of decay amplitudes. In this paper, we present a full analytical

This work was supported by National Institutes of Health grants R01 CA 211084 and R01 EB 000768.

S. S. Hou and B. J. Bacskai is with the Alzheimer's Disease Research Unit, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts 02129, USA. e-mail: shou@nmr.mgh.harvard.edu.

A. T. N. Kumar is with the Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts 02129, USA. e-mail: ankumar@nmr.mgh.harvard.edu.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

derivation of the CCTD estimator. We also present the first experimental comparison of the relative quantitation performance of the DTD, CCTD and ATD methods. The results demonstrate that the ATD approach provides the best relative quantitation performance among the three methods, and can accurately recover concentration ratios of up to 1:4.25. ATD can accurately localize inclusions as close as 1.5 mm, which is several-fold smaller than DTD. DTD results in large errors in quantitation and localization, but provides better spatial resolution. The CCTD method provides improved relative quantitation compared to DTD, while also providing better spatial resolution than the ATD method.

We show that a resolution matrix-based formalism provides a rigorous explanation for these experimental results. The off-diagonal terms of the resolution matrix provide a complete description of cross-talk between multiple lifetime components. The ATD approach produces a resolution matrix that is zero for off-diagonal terms and is ideal when relative quantitation is of importance, while CCTD is more appropriate when spatial resolution and localization are important. We also present a statistical, Bayesian interpretation of the ATD and DTD estimators, by expressing them in alternate forms. A by-product of this analysis is a formula for DTD inversion that allows rapid computation of the DTD solution, which is currently infeasible to compute for more than a few time gates and large tomographic data sets. The formula allows comparison of the DTD and ATD inversion problems for an arbitrary number of time gates and a dense set of measurements.

II. Background

A. Linear MMSE Estimator

A general multi-parameter linear inverse problem can be described by the matrix equation:

$$y = Wx + n, \tag{1}$$

where y is a $(M \times 1)$ data vector, W is a $(M \times NP)$ matrix representing the forward model, x is a $(NP \times 1)$ model vector containing P physical parameters, each evaluated at N spatial locations and n is a $(M \times 1)$ additive noise vector. A common approach for solving this inverse problem is by adopting a Bayesian approach where both the noise and model vectors terms are assumed to be random variables with known first and second order moments: $E[n] = 0$, $cov[n] = C_n$ and $E[x] = \mu_x$, $cov[x] = C_x$. It is further assumed that the model and noise vectors are uncorrelated: $E[nx^T] = 0$. When estimators are restricted to be linear with respect to the data, the estimated model vector $\widehat{x}$ can be written in the form:

$$\widehat{x} = \widehat{W}y + b, \tag{2}$$

where $\widehat{W}$ is a $(NP \times M)$ matrix operator and b is a $(NP \times 1)$ vector.

Among the class of linear estimators, the linear minimum mean square error (LMMSE) estimator seeks to find the W and b which minimize the mean square error (MSE) given by:

$$\Omega = E\left[\|x - \widehat{x}\|^2\right]. \tag{3}$$

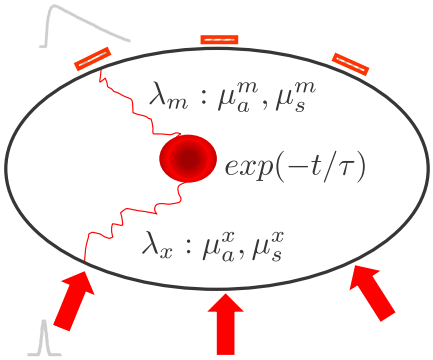

Fig. 1. Schematic showing the three components of fluorescent light transport in a scattering medium: propagation of the excitation light at λ_x, response of the fluorophore to excitation ($exp(-t/\tau)$) and propagation of the fluorescence light at λ_m. Also shown is the broadening of the temporal response due to the medium compared to the excitation pulse.

By setting $\partial\Omega/\partial W = 0$ and $\partial\Omega/\partial b = 0$ and solving for both $\widehat{W}$ and b, we find the LMMSE solution is given by:

$$\widehat{x} = \mu_x + \widehat{W}_{LMMSE}(y - W\mu_x) \tag{4}$$

$$\widehat{W}_{LMMSE} = C_x W^T (W C_x W^T + C_n)^{-1}. \tag{5}$$

In the case when x and n can be modeled as jointly Gaussian distributed and independent then the LMMSE estimator is equivalent to the mean of the posterior probability distribution $p(x|y)$ [1]. The solution in (4) and (5) corresponds to the commonly used generalized least squares solution in diffuse optical tomography image reconstruction [6] and is equivalent to the Wiener filter solution when $\mu_x = 0$ [1].

B. The Forward Problem for TFLM

Light transport in a turbid medium filled with an arbitrary distribution of fluorophores can be modeled (Fig. 1) as the sequential propagation of excitation light from source locations (r_s) on the boundary of the medium to fluorophores within the medium volume, followed by fluorescence emission and the propagation of the emitted fluorescence from the fluorophore to detectors (r_d) on the boundary [7], [8]. The propagation of light through tissue is most rigorously described by coupled radiative transport equations (RTEs) at the excitation and emission wavelengths [8]. Using the Green's functions of the coupled RTEs and neglecting the re-emission of fluorescence (first order approximation), the fluorescence intensity at r_d due to a source at r_s can be expressed as a weighted sum of fluorophore-dependent weight function for N fluorophores:

$$y(r_s, r_d, t) = \sum_{n=1}^{N} \int_{\Omega} W_n(r_s, r_d, r, t)\eta_n(r) d^3r, \tag{6}$$

where $W_n = G^x(r_s, r, t) * exp(-t/\tau_n) * G^m(r, r_d, t)$ is a temporal-double convolution of transport Green's functions at the excitation (G^x) and emission (G^m) wavelengths and an exponential decay with a time constant equal to the fluorescence lifetime, τ_n. The fluorescence yield of the n'th fluorophore is expressed as $\eta_n(r) = Q_n \epsilon_n c_n(r)$ where Q_n, ϵ_n and c_n are the quantum yield, extinction coefficient and concentration of the fluorophore, respectively. The integral equation in (6) can

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

be discretized into a linear matrix equation for L time gates and M source-detector pairs:

$$y = W\eta \tag{7}$$

where y is a $ML \times 1$ vector of the measured fluorescence data, $W = [W_1, ...W_N]$ is a $(ML \times VN)$ matrix of TD sensitivity functions, and $\eta = [\eta_1, ...\eta_N]^T$ is a $VN \times 1$ vector of yield distributions corresponding to lifetimes $\tau_1, \tau_2, ...\tau_N$.

C. The Inverse Problem for TFLM

The goal of tomography is to recover the 3D yield distributions, $\eta_n(r)$ for all distinct lifetimes τ_n present in the medium from the set of tomographic measurements y as expressed in (7). We assume that the τ_n's are known *a priori* from independent measurements [9]. The standard method [10], [11] for solving for η from TD measurements, y, is to directly invert the entire TD weight matrix, W, using Tikhonov regularization. In this approach, which we label as the direct TD (DTD) approach, the reconstructed yield distribution $\widehat{\eta}_{DTD}$ is given by:

$$\widehat{\eta}_{DTD} = \widehat{W}_{DTD} y, \tag{8}$$

where $\widehat{W}_{DTD}$ is the Tikhonov inverse operator, given by:

$$\widehat{W}_{DTD} = W^T (WW^T + \lambda I)^{-1}. \tag{9}$$

The Tikhonov solution can be made equivalent to the LMMSE solution in the Bayesian formulation of the TFLM inverse problem when the commonly made statistical assumptions that y and η are white Gaussian distributed are made and for appropriate choice of λ. For the remainder of this paper it is assumed that these assumptions are met and that the DTD solution and LMMSE solution can be used interchangeably. The DTD approach in the form of (9) is computationally intensive for more than a few time gates (L) and does not optimally exploit the information content in the TD data. Further, we will show below that DTD results in significant cross-talk between yield distributions η_n.

An alternate approach is made possible for tomographic lifetime multiplexing that uses time points in the "asymptotic" regime (defined by $t \gg \tau_D$), under the condition that the intrinsic lifetimes are greater than the diffusive time scale, viz., $\tau_n > \tau_D$. Under these two widely applicable conditions [12], we have shown that the TD weight matrix in (7) factorizes into temporal and spatial matrices as [2]:

$$W \overset{t \gg \tau_D}{=} A\overline{W}, \tag{10}$$

where $A = \begin{bmatrix} exp(-t/\tau_1) \otimes I & \dots & exp(-t/\tau_N) \otimes I \end{bmatrix}$ is a $ML \times MN$ dimensional basis matrix containing Kronecker products of exponential decay terms and $M \times M$ identitiy matrices, I and $\overline{W} = diag(\overline{W}_1, \dots, \overline{W}_N)$ is a $(MN \times VN)$ time-independent block diagonal matrix containing $(M \times M)$ CW weight matrices $\overline{W}_n$, which are evaluated using a reduced absorption of $\mu_a(r) - \Gamma_n/v$ (see [3] for a detailed derivation). The factorization of W from (10) allows a two-step approach for lifetime tomography, as can be seen by writing the forward problem in (7) as:

$$y = A\overline{W}\eta. \tag{11}$$

In the first step, the well conditioned matrix A is inverted using its Moore-Penrose pseudoinverse and applied on the measurements y:

$$A^\dagger y (= a) = \overline{W}\eta. \tag{12}$$

Note that the first equality in (12) is essentially the normal equations in linear regression [13], which in the present case is a multi-exponential fit of the decay portion of the TD data and leads to decay amplitudes a. In the second step, the weight matrix $\overline{W}$ is inverted using the standard Tikhonov regularization (as in (9)) and applied to the decay amplitudes. The two-step procedure can be written as a single linear inverse operator by combining (12) with the Tikhonov inversion of $\overline{W}$. The resulting asymptotic TD (ATD) solution is then given by:

$$\widehat{\eta}_{ATD} = \overline{W}^{-1} A^\dagger y = \widehat{W}_{ATD} y, \tag{13}$$

where $\overline{W}^{-1}$ is the Tikhonov inverse of $\overline{W}$ and $\widehat{W}_{ATD}$ is the ATD inverse operator incorporating both steps and given by:

$$\widehat{W}_{ATD} = \overline{W}^T (\overline{W}\overline{W}^T + \lambda I)^{-1} A^\dagger. \tag{14}$$

Due to the block diagonal form of $\overline{W}$, the yield distributions for each lifetime can be solved separately in the ATD approach. As we have shown previously, this property of ATD provides a much lower cross-talk compared to the DTD approach, thereby significantly improving localization accuracy when resolving multiple yield distributions with distinct lifetimes [2]. The separation of inverse problems also allows us to individually normalize the CW weight matrices ($\overline{W}_n$) (divide $\overline{W}_n$ by its maximum matrix entry) while keeping the regularization, λ the same for all n. We have found such a scheme can result in some minor improvements in relative quantitation as lifetime separation is increased.

While it can be inferred from the separation of the inverse problem for different lifetime components in (14) that ATD provides reduced crosstalk between lifetime components, the fundamental distinction between the ATD and DTD methods, in terms of cross-talk performance, is not yet apparent. It is clear that DTD results in higher cross-talk. However DTD can provide a MMSE solution (for proper selection of λ), while ATD will lead to higher reconstruction error. A natural question, therefore, is whether there exist more general linear estimators that can provide the improved cross-talk performance of ATD while providing minimum reconstruction error. In the next sections, we precisely address this question from the point of view of general multi-parameter inverse problems. We also recast the ATD and DTD inverse problems in a form that provides a clear statistical interpretation (in a Bayesian sense) of the performance advantages of ATD. In order to do this we next provide a formal definition of cross-talk in terms of the model resolution matrix of a linear inverse operator.

D. Model Resolution Matrix

A useful tool to evaluate the performance of a linear inverse operator, $\widehat{W}$, is the model resolution matrix, R defined as:

$$R = \widehat{W}W. \tag{15}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

By substitution of (1) into (4) and assuming for simplicity $b = 0$ ($\mu_x = 0$), the optimal solution can be written in terms of R as:

$$\widehat{x} = Rx + \widehat{W}n. \tag{16}$$

This solution can be divided into two parts: the noise amplification term $\widehat{W}n$ represents the effect of the noise on the solution while the bias term Rx can be interpreted as the model resolution matrix acting as an averaging filter between the estimated model vector $\widehat{x}$ and the true model vector x. The columns of R have simple interpretations in the case of a single parameter inverse problem [14]. The c'th column of R contains the estimate at every voxel due to a unit inclusion at the c'th voxel. Hence, each column contains a point spread function of the imaging process. While the model resolution matrix for a single parameter is straightforward to interpret, we will see in the next section that for multi-parameter inverse problems, the model resolution matrix contains the complete information about the cross-talk between parameters.

III. Results

A. Resolution Matrix Definition of Inter-parameter Cross-talk

The resolution matrix provides a transparent way to analyze inter-parameter cross-talk when dealing with multiple parameter problems. For simplicity, we consider inverse problems with two parameter types, although the statistical analysis and proposed methods can be readily generalized to any number of parameters. We first write the true (x) and estimated ($\widehat{x}$) model vectors as:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \widehat{x} = \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{bmatrix}. \tag{17}$$

The forward matrix, W and inverse operator, $\widehat{W}$ can also be divided into sub-matrices corresponding to each parameter:

$$W = \begin{bmatrix} W_1 & W_2 \end{bmatrix}, \widehat{W} = \begin{bmatrix} \widehat{W}_1 \\ \widehat{W}_2 \end{bmatrix}. \tag{18}$$

The model resolution matrix from (15) for the two-parameter case, $R^{(2)}$, thus takes the form:

$$R^{(2)} = \begin{bmatrix} \widehat{W}_1 W_1 & \widehat{W}_1 W_2 \\ \widehat{W}_2 W_1 & \widehat{W}_2 W_2 \end{bmatrix} \equiv \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}. \tag{19}$$

$R^{(2)}$ can be divided into four quadrants. The diagonal quadrants (1, 1) and (2, 2) (corresponding to the terms R_{11} and R_{22} in (19)) contain the individual point spread functions for parameters x_1 and x_2, respectively. However, the off-diagonal quadrants represent a different type of point spread function involving the cross-talk between the parameters. Specifically, columns in quadrant (1, 2) (corresponding to $\widehat{W}_1 W_2$) represent the cross-talk at all voxels for parameter x_1 due to a point inclusion for parameter x_2, and vice versa for quadrant (2, 1). The off-diagonal blocks of the model resolution matrix, R_{12} and R_{21}, allow direct quantification of cross-talk for general linear MPIPs, and provide a systematic and complete measure of cross-talk for all possible combinations of true and image voxels. Since the terms in R are independent of the data y and only depend on W and the *a priori* assumptions about the noise and model, (19) can be used to design and evaluate the performance of a linear inverse operator without generating data from the forward problem. In the following sections we will derive an estimator for TFLM based on an LMMSE estimator with a direct constraint placed on the off diagonal blocks of $R^{(2)}$.

B. Bayesian Interpretation of Direct and Asymptotic TD Methods

Before we derive the optimal estimator for TFLM based on a cross-talk constraint, we present alternate formulations for the DTD and ATD inverse operators that can provide insights into the statistical properties of the solutions from a Bayesian perspective, and also reveal the connection between the two approaches. The ATD inverse operator from (14) can be expressed as follows by using the factorization in (10) (derivation is provided in supplemental materials):

$$\widehat{W}_{ATD} = W^T (WW^T + \lambda AA^T)^{\dagger}. \tag{20}$$

Next, the DTD inverse operator from (9) can also be written in an alternate form in terms of the block diagonal CW weight matrix, $\overline{W}$ (see supplemental materials) as follows :

$$\widehat{W}_{DTD} = \overline{W}^T (\overline{W}\overline{W}^T + \lambda (A^T A)^{-1})^{-1} A^{\dagger}. \tag{21}$$

These substitutions can be used to understand the key differences between the ATD over the DTD (MMSE) solution. We first recall that the DTD solution is the optimal solution in a Bayesian sense, assuming $\lambda = (\sigma_n/\sigma_\eta)^2$, where σ_n and σ_η are the variances of the measurement noise and model, which are assumed to obey white Gaussian distributions, i.e., $\eta \sim N(0, \sigma_\eta^2 I)$ and $n \sim N(0, \sigma_n^2 I)$. The alternate forms of ATD (20) and DTD (21) show that the ATD approach is equivalent to the DTD approach (MMSE) under two conditions:

1) Comparing (20) with (9), we see that the key difference between ATD and DTD is that the ATD inverse includes a measurement covariance of the form AA^T in the under-determined form of the DTD inverse operator. Thus, the ATD method can be viewed as a Tikhonov technique where the regularization parameter is varied based on index of the fluorescence data. This is analogous to other Tikhonov techniques such as previously described spatially variant regularization (SVR) [15], in which the regularization parameter is based on spatial location inside the medium. While in SVR, the voxel-based regularization matrix ($\lambda J^T J$) is to provide spatially uniform resolution and contrast, with the ATD approach, the data-based regularization matrix (λAA^T) is to effect a reduction in cross-talk.
2) Comparing (21) with (14), it is clear that DTD includes a covariance matrix $(A^T A)^{-1}$ in the inverse operator, whereas in ATD, this term is set to an identity matrix I. The term $(A^T A)^{-1}$ is recognized from linear regression theory [13] as the covariance matrix for the parameters, or unknowns (which in the present case are the decay amplitudes) in a multi-exponential analysis of the raw TD data. The ATD solution thus directly excludes the influence of inter-amplitude covariance (resulting from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

multi-exponential fitting) during the inversion for the yield distributions. The alternate form of the DTD inverse in (21) shows that this approach implicitly contains the inter-amplitude covariance, thereby explaining the increased cross talk in the DTD yield reconstructions.

The alternate form for DTD derived in (21) allows us to compare the resolution matrices for DTD and ATD in a clear fashion. Using (21), (14), and the definition of R in (15), the resolution matrices for the DTD (R_{DTD}) and ATD (R_{ATD}) inverse operators take the form:

$$R_{DTD} = \overline{W}^T(\overline{W}\overline{W}^T + \lambda(A^TA)^{-1})^{-1}\overline{W} \quad (22)$$

$$R_{ATD} = \overline{W}^T(\overline{W}\overline{W}^T + \lambda I)^{-1}\overline{W} \quad (23)$$

It can be seen that (22) and (23) differ by the presence of the decay amplitude covariance matrix $(A^TA)^{-1}$ in the DTD resolution matrix. Since every term in the RHS of (23) is block diagonal, the resolution matrix R_{ATD} is also a block diagonal matrix. Therefore, ATD reconstructions contain zero cross-talk. On the other hand, R_{DTD} is generally not block diagonal since the covariance term $(A^TA)^{-1}$ is not block-diagonal. From these observations it is clear that although the DTD solution provides a MMSE solution, it suffers from inter-parameter cross talk when fitting multiple parameters as in the lifetime multiplexing case. The ATD reconstruction provides reduced cross-talk, which leads to improved localization of closely separated fluorescent targets [2], [3]. We will numerically illustrate the resolution matrices in Section IV.

C. Cross talk-constrained-MMSE Estimator for TFLM

We next address the question of whether optimal estimators exist that provide MMSE solutions while also providing zero cross-talk, thereby providing better error performance than the ATD approach. To address this question, we propose a novel, Bayesian inversion algorithm for cross-talk reduction in MPIPs, by seeking an MMSE solution with an imposed zero cross talk constraint. The basic idea of our new estimator is to provide optimal separation between parameters at the expense of higher total MSE than the DTD approach, while providing better MSE than the ATD approach. We begin by assuming that both $E[x] = E[n] = 0$. In our approach, the mean square error cost function in (3) is minimized while linear constraints are placed on the cross-talk matrices defined in (19). We refer to the new estimator as a cross-talk constrained-minimum mean square error estimator (CCMMSE). With the inverse operator defined as in (2) (with $b = 0$), the optimization problem takes the form:

$$\widehat{W}_{CCMMSE} = \arg\min_{\widehat{W}} E\left[\|x - \widehat{x}\|^2\right] \quad (24)$$

with the constraints:

$$R_{12} = 0 \text{ and } R_{21} = 0. \quad (25)$$

Note that the constraint is on the model resolution matrix itself as opposed to other constrained MMSE approaches which have focused on the moments of $\widehat{x}$ [16]. The optimization problem can be described as finding the estimator with the lowest error rate among all estimators that produce resolution matrices of a particular form (block diagonal in the present case). The minimization in (24) and (25) is an example of a quadratic programming problem and can be solved by standard methods [17]. An expression for the CCMMSE estimator is derived in Appendix A. Briefly, we transform the quadratic programming problem with linear constraints into an unconstrained optimization problem using a change of variables [18]. The final solution for the optimal estimator is given by:

$$\widehat{W}_1 = C_{x_{11}}W_1^TN_1(N_1^T(W_1C_{x_{11}}W_1^T + C_n)N_1)^{-1}N_1^T \quad (26)$$

$$\widehat{W}_2 = C_{x_{22}}W_2^TN_2(N_2^T(W_2C_{x_{22}}W_2^T + C_n)N_2)^{-1}N_2^T \quad (27)$$

where $N_1 = null(W_2^T)$, $N_2 = null(W_1^T)$ and $C_x = \begin{bmatrix} C_{x_{11}} & C_{x_{12}} \\ C_{x_{12}} & C_{x_{22}} \end{bmatrix}$. It is clear that the constraints in (25) can only result in a nontrivial solution for the estimator when $nullity(W_1^T) > 0$ and $nullity(W_2^T) > 0$. By the rank-nullity theorem, this is only satisfied when W_1 and W_2 do not have full row rank. While this condition is generally satisfied for over-determined systems, it only holds for certain under-determined systems. Due to redundancy from the spatial-temporal factorization, the TFLM forward matrix does satisfy this condition and does not have full row rank, even for the under-determined case [5]. For other multi-parameter inverse problems, this condition will need to be evaluated to determine if a nontrivial solution can be found. We discuss less restrictive constraints for cross-talk reduction which allow nontrivial solutions even in the full row rank case in Section VI.

We now apply this new CCMMSE estimator to the TFLM inverse problem. We use the same implicit assumptions as Tikhonov regularization for the model and data covariance matrices ($C_x = \sigma_\eta^2 I$ and $C_n = \sigma_n^2 I$). We then substitute the TFLM weight matrix and covariance matrices in (26) and (27). The full derivation is presented in Appendix B. The optimal inverse operator with constraints on cross-talk, which we term the CCTD estimator, was found to be:

$$\widehat{W}_{CCTD} = \overline{W}^T(\overline{W}\overline{W}^T + \lambda DIAG\left((A^TA)^{-1}\right))^\dagger A^\dagger, \quad (28)$$

where $DIAG(X)$ converts the matrix X to block diagonal form by setting all elements of the off-diagonal block to 0 and keeping all other elements the same.

Using the definition in (17), the corresponding resolution matrix for CCTD takes the form:

$$R_{CCTD} = \overline{W}^T(\overline{W}\overline{W}^T + \lambda DIAG\left((A^TA)^{-1}\right))^\dagger \overline{W}. \quad (29)$$

The inverse operator shown in (28) provides the optimal CCMMSE estimator for lifetime multiplexing. Remarkably, (28), which was derived by solving an optimization problem, is essentially the same as the ATD inverse operator, which is based on a two step inversion of the factorized TD weight matrix (10) in the asymptotic region. The difference between ATD and CCTD is the replacement of the identity matrix, λI in (14) in ATD with a block diagonal matrix, $\lambda DIAG(A^TA)^{-1}$ in CCTD. This substitution shows that the optimal solution with zero crosstalk is obtained by simply setting the off-diagonal matrix elements of the covariance matrix in the DTD inverse operator to zero, to maintain the off-diagonal structure of the resulting model resolution matrix. The uncertainties in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

the individual amplitudes are still retained in the on-diagonal blocks. The DTD inverse operator from (21) implicitly retains the full amplitude covariance matrix into the regularization, thereby explaining the higher cross talk between multiple yield distributions in the DTD approach.

IV. Simulations

We illustrate the key aspects of the above theoretical results using Monte Carlo (MC) simulations, performed using tMCimg [19], a MC computing software package. A simulation medium was set up with dimensions 50 mm $\times$ 30 mm $\times$ 20 mm. The optical absorption (μ_a) and scattering (μ_s) were set to 0.1 cm^{-1} and 10 cm^{-1}, respectively, and the anisotropy (g) was set to 0.01. All MC simulations were performed for 10^9 photons per source. The simulations used 21 equally spaced (5 mm) sources and detectors in a transmission geometry. The MC software also generated the source and detector Green's functions (or "2-pt functions"), G^x and G^m, which were used to calculate the weight functions W, $\overline{W}_n$ and $\overline{W}$. The DTD, CCTD and ATD inversions were performed using (9), (14) and (28), with 40 time gates in the asymptotic regime, which was chosen to start 1.2 ns from the peak of the TD fluorescence data. 2% shot noise was added to the simulated TD data.

A. Resolution Matrix

Figure 2 shows the resolution matrices for DTD, CCTD and ATD, R_{DTD}, R_{CCTD} and R_{ATD}, respectively, computed using (22), (29) and (23). To aid with visualization, the rows and columns of the matrices were binned down by a factor of 300 and normalized to the maximum value of the matrix. It can be seen that the off-diagonal blocks of R_{DTD} contain non-zero terms, while all the elements in the off-diagonal blocks of R_{CCTD} and R_{ATD} are identically zero. The resolution matrix therefore shows in the most general form, that CCTD and ATD inversions have zero cross-talk across all voxels of the imaging medium. It should also be noted that corresponding elements within the two block matrices on the main diagonal (R_{11} and R_{22} in (19)) of R_{DTD} are significantly different. The two diagonal blocks represent the point-spread functions for the two lifetime components. As such, this difference contributes to the inaccurate recovery of the relative concentrations of the two lifetimes using DTD. In contrast, the relative magnitudes of corresponding elements on the diagonal blocks of both R_{CCTD} and R_{ATD} are much closer, with corresponding elements for R_{ATD} being nearly identical. This implies improved relative quantitation for both CCTD and ATD. These results will be confirmed using simulations and experiments below.

B. Relative Quantitation

To compare the relative quantitation performance of the DTD, CCTD and ATD inversions, we performed simulations with two overlapping fluorophores (lifetimes of τ_1 = 0.75 ns, and τ_2 = 1 ns) located at the same 1 mm^3 voxel at a height of 10 mm in the slab medium. Reconstructions were performed for five ratios of the yields, η_1 and η_2, of the two lifetime components, namely $\eta_1/\eta_2 = 1:1, 2:1, 3:1, 4:1$ and $5:1$. To calculate the standard deviation of the yield estimates, simulations were repeated with 100 different realizations of shot noise (2%). Figure 3 shows the recovered yield ratios (obtained as the ratio of maximum of the reconstructed yields) using DTD (blue), CCTD (green) and ATD (red) along with the true ratio in black, with the standard deviations shown as error bars. It is clear that ATD recovers the true yield ratios accurately for all ratios considered, while the DTD results in a significant error, vastly underestimating the short lifetime component. CCTD also underestimates the short lifetime component but results in improved relative quantitation as compared to DTD. It can be noted that the standard deviation for the reconstructed ratio was highest for ATD followed by CCTD and DTD. This disparity in standard deviation among the methods was greater for higher values of true ratios and can at least partly be attributed to the increased uncertainty inherent for larger recovered ratios. The smaller uncertainty for DTD could also be the result of the MMSE nature of the DTD solution which provides improved noise sensitivity.

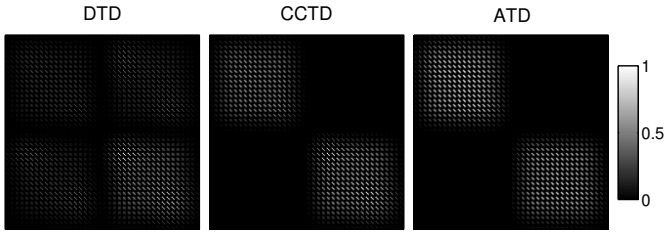

Fig. 2. Resolution matrices, R_{DTD}, R_{CCTD} and R_{ATD} for tomographic inversion of two fluorophores with distinct lifetimes of 0.75 ns and 1 ns in a turbid slab medium (parameters discussed in text). The matrices are binned down by factor of 300 and normalized to their maximum values.

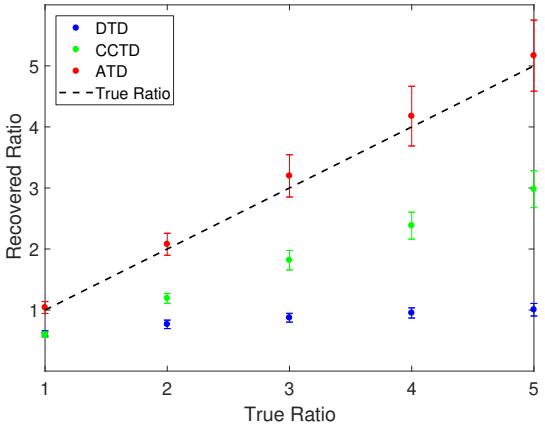

Fig. 3. Relative quantitation performance of DTD (blue), CCTD (green) and ATD (red) tomographic inversions of simulated data. Shown are the tomographically recovered yield ratios, η_1/η_2, for two fluorophores with distinct lifetimes (τ_1 = 0.75 ns, and τ_2 = 1 ns) at the center of a 2 cm thick slab phantom ($\mu_a = 0.1$ cm^{-1} and $\mu_s' = 10$ cm^{-1}). Reconstructions were performed for five ratios of the yields, namely $\eta_1/\eta_2 = 1:1, 2:1, 3:1, 4:1$ and $5:1$. Vertical error bars indicate yield estimates obtained from simulations with 100 different realizations of shot noise (2%). Dashed lines indicate the true ratios.

V. Experimental Validation

We next validated the above simulation results using phantom measurements with a TD fluorescence tomography system

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

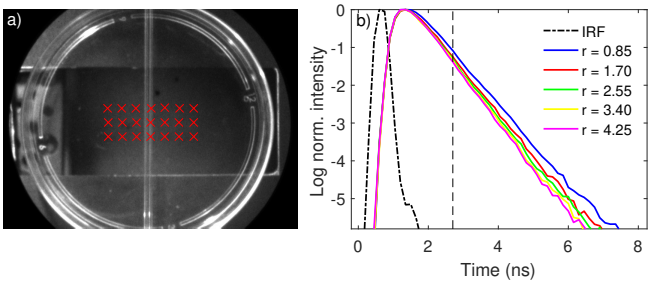

Fig. 4. a) Dish phantom with capillary tubes separated by 1.5 mm. During measurements, the dish is filled with Intralipid and nigrosin solutions, the right tube is filled with IR-806 ($\tau = 0.74$ ns, 2 μM) and the left tube with Alexa Fluor 750 ($\tau = 1.01$ ns, 0.2-1 μM). Also shown are the source and detector positions (labeled with red 'x'). b) Representative log-normalized fluorescence intensity curves for a source-detector pair at the center of the dish, for each yield ratio $r = \eta_1/\eta_2$. The dash dotted black line shows the instrument response function (IRF) while the dashed black vertical line indicates the start of the asymptotic region.

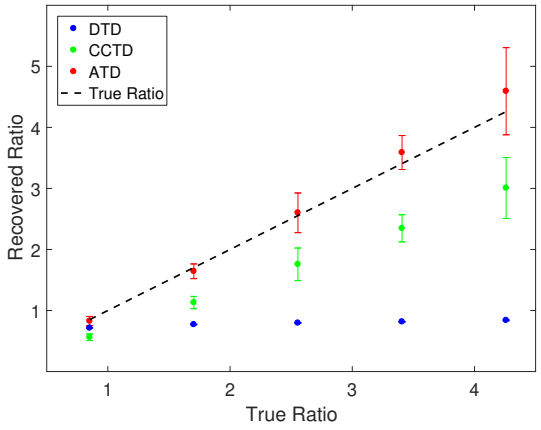

Fig. 5. Experimental comparison of the relative quantitation performance of DTD, CCTD and ATD tomographic inversions of TD measurements. Reconstructions were performed on TD tomogaphic data with a dish phantom (1.8 cm height, $\mu_a = 0.1$ cm^{-1} and $\mu_s' = 10$ cm^{-1}) with two capillary tubes (I.D.: 0.90 mm, O.D.: 1.20 mm) located at a depth of 9 mm and separated by 1.5 mm center-to-center. The reconstructed yield ratio η_1/η_2 is shown for DTD (blue), CCTD (green) and ATD (red) along with the true ratio of the yields (black dashed line).

[4]. Liquid phantoms were created using polystyrene cell culture dishes (100 mm diameter, Corning). The phantom was filled with an Intralipid and nigrosin mixture to a height of 18 mm, resulting in optical properties of $\mu_a = 0.1$ cm^{-1} and $\mu_s' = 10$ cm^{-1}. Two capillary tubes (I.D.: 0.90 mm, O.D.: 1.20 mm) were located at a depth of 9 mm with a 1.5 mm center-to-center separation. The choice of closely spaced but non-overlapping inclusions for the experiment was to avoid chemical interactions between the fluorophores that could occur in a mixture. In addition, this allows comparison of the ability of the three methods to localize closely separated targets. The tubes were filled with IR-806 (Sigma-Aldrich, 2 uM in ethanol. $\tau_1 = 0.74$ ns) and Alexa Fluor 750 (Invitrogen, 1 uM in ethanol. $\tau_2 = 1.01$ ns). The ratio of the fluorescence yields of IR-806 (η_1) and Alexa Fluor 750 (η_2) was varied from $\eta_1/\eta_2 = 0.85$ to 4.25 in 5 steps, using serial dilutions of the Alexa Fluor 750 solution. The true (*in vitro*) yield ratio of the two tubes for each combination was estimated by first directly measuring the fluorescence intensity from the tubes in the phantom (without Intralipid) for the smallest yield ratio (0.85), and calculating the subsequent ratios based on the dilutions used for Alexa Fluor 750.

For tomography, 21 sources (3 rows of 7 sources across the tubes) were used at the bottom of the dish, with detectors assigned as camera pixels directly above the sources (4 $\times$ 4 hardware binning) on top of the dish (See Fig. 4). The full TD tomographic data was collected using a TD imaging system with a Ti:Sapphire laser ($\sim$ 100 fs pulses) and a gated intensified CCD camera (ICCD) (LaVision:Picostar, 500 ps gatewidth, 560 V gain, 20-25 ms CCD integration time, 150 ps time step). For each ratio of the yields, η_1/η_2, 10 tomographic data sets were collected to allow estimation of the uncertainty in the reconstructed ratios. The decay amplitudes for each source detector pair were obtained using the Moore-Penrose pseudoinverse ($A^\dagger y$, see (12)) of the raw TD measurements y, using 55 times gates in the asymptotic region, which was chosen to start at 1.2 ns from the peak of the TD fluorescence signal. The instrument response function (IRF), measured using a white paper placed on the imaging plate, was directly incorporated in the linear bi-exponential fit using a forward convolution with exponentials, which formed the basis functions for the fit.

Figure 5 shows the reconstructed fluorescence yield ratios from the experimental data for a range of true ratios, using the DTD (9), CCTD (28) and ATD (14) methods. Vertical error bars indicate variations across the 10 measurements for a given ratio. Figure 6 shows the 1-D line profiles, taken along the x-axis through the maximum of the yields, η_1 and η_2, for DTD, CCTD and ATD, reconstructed from the averaged data over all 10 trials. Also shown are the true ratios (solid vertical rectangles), normalized to η_1 for convenience. The ATD approach provides excellent accuracy in both relative yield ratio and the localization, resulting in a range of error in relative quantitation of 1.8%-7.8% across all ratios. DTD results in significantly higher error in relative quantitation (16.1%-80.3% across all ratios), vastly underestimating the yield of the short lifetime dye (IR-806), and is unable to localize the inclusions as spatially separate. CCTD is able to accurately localize the tubes while providing a reduced relative quantitation error over DTD of 29.4%-33.9% across all ratios. In addition, CCTD provides a 11.0% reduction in the full width at half maximum (FWHM) of the spatial distribution for η_2 compared to ATD, while the η_1 distribution remains the same for both methods. These observations are in agreement with the simulations (Fig. 3). It is also noteworthy that while DTD results in large errors in quantitation and localization, it provides a reduction in FWHM of 23.7% for η_1 and 12.7% for η_2, compared to CCTD. Thus DTD is ideal for applications where only a single lifetime is present, when quantitative multiplexing is not of interest. It should be noted that a center-to-center separation of 1.5 mm is more than two-fold smaller than the previously reported center-to-center spatial separation using CW fluorescence tomography ($\sim$4.7 mm center-to-center) [20], illustrating the powerful advantage of ATD and CCTD for accurate quantitative multiplexing.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

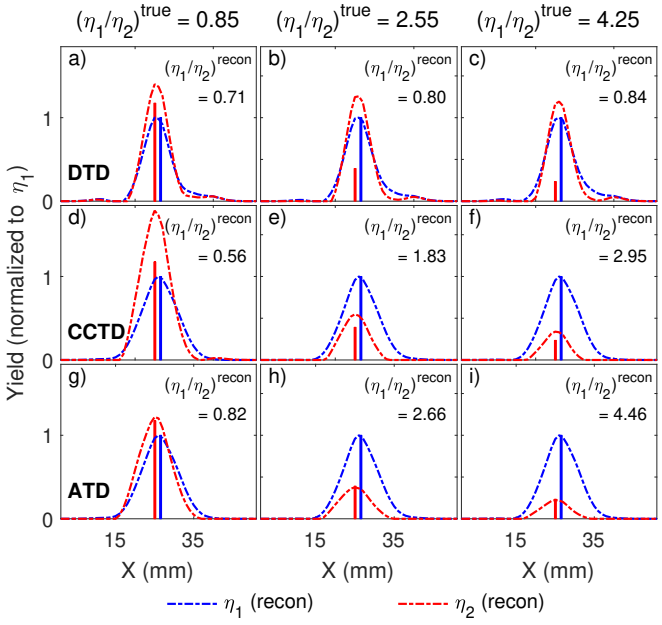

Fig. 6. 1-D line profiles of the DTD (a-c), CCTD (d-f) and ATD (g-i) yield reconstructions, η_1 (blue dashed-dot line) and η_2 (red dashed-dot line), corresponding to the relative quantitation results shown in Fig. 5. The line profiles are taken along the x-axis through the maximum of the reconstructed yields. Also shown are the true ratios (solid rectangles), normalized to η_1.

VI. Discussion

The resolution matrix-based framework for quantifying cross-talk in tomographic lifetime multiplexing shows that MMSE solutions, such as the DTD approach, may not be ideal for multiplexing problems, since they do not constrain inter-parameter cross-talk. Using simulations and experiments, we showed that the high cross-talk of DTD translates to a significant error in relative quantitation and spatial localization. On the other hand, the ATD inversion, which is not an MMSE solution but is rather based on a two-step approach using multi-exponential analysis of TD data, provides zero cross talk and excellent accuracy for relative quantitation for a wide range of concentration ratios. The ATD reconstructions result in broader spatial point spread functions (PSF), implying poorer spatial resolution than DTD. Intermediate between ATD and DTD solutions is the CCTD estimator, which is an MMSE solution with a constraint on the off-diagonal blocks of the resolution matrix, which represent cross talk. We showed experimentally that CCTD provides improved relative quantitation compared to DTD while also providing narrowed spatial PSF compared to ATD. These results suggest that ATD is the method of choice for optimal relative quantitation and CCTD is the method for optimal spatial resolution while maintaining zero cross-talk. It is conceivable that ATD could be combined with CCTD in some manner to incorporate the improved spatial resolution of the CCTD approach into ATD. The spatial resolution of both the CCTD and ATD methods can also be further improved by combining with early photons, as in the hybrid TD (HTD) approach [2].

Our approach for deriving the CCTD estimator was based on shaping the model resolution matrix to minimize cross-talk. It should be noted that in this derivation, no assumptions were made regarding the form of the inverse operator. The solution for the optimization problem was obtained analytically, and was nearly identical in form to the ATD inverse operator. The key difference between CCTD and ATD is that while the regularization matrix for ATD is an identity matrix, the CCTD regularization matrix has unequal diagonal elements, and is related to the amplitude covariances, which have the effect of lowering reconstruction error while maintaining zero cross-talk. This derivation therefore rigorously establishes the relationship of the ATD solution to the minimum error solution on the spectrum of zero crosstalk estimators. We also presented alternate forms of DTD and ATD inversions that revealed the distinction between the cross-talk performance of DTD and ATD as simply related to distinct regularization matrices in a Bayesian interpretation. In addition, the alternate form of the DTD inverse operator allows a computationally efficient method of performing reconstruction of asymptotic TD data with an arbitrary number of time gates.

The framework for quantifying and reducing cross-talk in TFLM could also be extended to other types of computational imaging modalities. An important example in optical imaging is diffuse optical tomography (DOT) where absorption and scattering are the parameters to be separated and cross-talk has been shown to be a confounding factor. Applying our constrained optimization approach to DOT would likely be more difficult than TFLM since the factorization, which occurs naturally in TFLM and plays a key role in the derivation, does not occur in DOT. However, the approach employed here in deriving the CCTD estimator, based on sacrificing reconstruction error performance for better crosstalk performance, should prove useful in DOT and other inverse problems. Furthermore, even though cross-talk has been the primary focus of the present study, the general procedure of constraining the model resolution matrix while minimizing reconstruction error could be extended to other performance measures. For example, this approach could help equalize the spatial resolution, which is represented by columns in the model resolution matrix, throughout the imaging medium. The spatial resolution for standard reconstruction algorithms typically exhibits strong depth dependence, thereby precluding comparisons of quantitation across spatial locations in the medium.

An important issue to consider when using a constrained optimization approach is the increase in error due to the imposing of constraints. For the TFLM inverse problem, the equality constraints in (25) still produced sufficiently regularized solutions at practical levels of the noise. However, for other multi-parameter problems, the equality constraints may cause the noise amplification error to dominate the solution (the solution becomes under-regularized). While this error can be decreased by assuming higher noise covariance (increasing regularization) the effect of noise amplification can also be mitigated by considering less restrictive constraints. In the case of cross-talk the constraint could be placed on a subset of terms in the off diagonal blocks instead of the entire off diagonal block. Other types of constraints could also be considered such as inequality constraints instead of equality constraints. However, these are numerically more difficult to solve.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

Our experimental results highlight the ability of zero cross-talk approaches such as ATD and to a lesser extent CCTD to achieve highly accurate relative quantitation between multiple parameters. In addition, we observed that the recovered relative quantitation for ATD and CCTD was robust, and was minimally affected by the choice of regularization parameter. Regularization had a bigger effect on the DTD recovered ratios, with higher regularization resulting in further underestimation of the ratios. Although absolute quantitation is difficult to achieve both in TFLM and optical tomography due to partial volume effects inherent in ill-posed inverse problems, our paper demonstrates that high accuracy for relative quantitation is possible over a wide range of fluorophore concentration ratios.

VII. Conclusion

We have presented the first experimental comparison of the ATD, DTD and CCTD methods for quantitative multiplexing in turbid media. The ATD approach provides superior relative quantitation and localization accuracy for recovering two fluorophores in a turbid medium, with true yield ratios of up to 1:4.25. While the standard DTD approach, which is a minimum error solution, provides improved spatial resolution compared to the ATD approach, it results in significant errors in relative quantitation. The CCTD approach provides improved quantitation compared to DTD and better spatial resolution compared to ATD. The ability to recover accurate relative quantitation of multiple fluorophores has important applications for biomedical imaging, including the extension of microscopy based fluorescence lifetime imaging applications [21]–[23] to tomographic whole body imaging. One application that holds particular promise is the detection of molecular interactions using fluorescence resonance energy transfer (FRET) [11], [24]. The ATD approach should prove to be a powerful approach to quantify the ratio of the donor to acceptor concentrations (and hence the FRET efficiency) in deep tissue. Another potentially important application where accurate relative quantitation is critical is the imaging of activatable probes. These probes have the property that their lifetime shifts upon activation by disease-specific enzymes such as proteases [25]. The ATD approach can allow accurate determination of the ratio of probes in an activated state to probes in the inactivated state, thereby accurately quantifying disease-specific molecular expression *in vivo*. Future work will be focused on applying ATD and CCTD for quantitative *in vivo* multiplexing applications.

Appendix A
CCMMSE Estimator

We seek to find an expression for W_{CCMMSE} which solves the quadratic programming problem defined in (24) and (25). The MSE cost function $f(\widehat{W})$ is first separated into terms which depend on $\widehat{W}_1$ and $\widehat{W}_2$, $f(\widehat{W}) = E\left[\|x-\widehat{x}\|^2\right] = f_1(\widehat{W}_1) + f_2(\widehat{W}_2)$,

$$f_1(\widehat{W}_1) = tr\big(-2\widehat{W}_1 W C_{x_1} + \widehat{W}_1 W C_x W^T \widehat{W}_1^T + \widehat{W}_1 C_n \widehat{W}_1^T + C_{x_{11}}\big) \quad (30)$$

$$f_2(\widehat{W}_2) = tr\big(-2\widehat{W}_2 W C_{x_2} + \widehat{W}_2 W C_x W^T \widehat{W}_2^T + \widehat{W}_2 C_n \widehat{W}_2^T + C_{x_{22}}\big) \quad (31)$$

where $C_x = \begin{bmatrix} C_{x_{11}} & C_{x_{12}} \\ C_{x_{12}} & C_{x_{22}} \end{bmatrix}$, $C_{x_1} = \begin{bmatrix} C_{x_{11}} \\ C_{x_{12}} \end{bmatrix}$, $C_{x_2} = \begin{bmatrix} C_{x_{12}} \\ C_{x_{22}} \end{bmatrix}$

We perform a change of variables to combine the equality constraints in (25) and the cost function in (24) into a single unconstrained cost function using the following procedure. We begin by defining the null spaces for W_1^T and W_2^T as:

$$N_1 = null(W_2^T) \quad (32)$$

$$N_2 = null(W_1^T) \quad (33)$$

and introducing two new variables Z_1 and Z_2:

$$\widehat{W}_1^T = N_1 Z_1 \quad (34)$$

$$\widehat{W}_2^T = N_2 Z_2. \quad (35)$$

We focus on solving for $\widehat{W}_1$ by substituting (34) into cost function (30) to get:

$$f_1(Z_1) = tr(-2Z_1^T N_1^T W C_{x_1} + Z_1^T N_1^T (W C_x W^T + C_n) N_1 Z_1 + C_{x_{11}}).$$

The problem of minimizing $f_1(Z_1)$ for Z_1 (an unconstrained minimization problem) is equivalent to the constrained optimization problem of (24) and (25) since for any Z_1 the constraint $\widehat{W}_1 W_2 = 0$ is satisfied. To minimize $f_1(Z_1)$, we take the derivative with respect to Z_1:

$$\frac{\partial f_1(Z_1)}{\partial Z_1} = -2N_1^T W C_{x_1} + 2N_1^T (W C_x W^T + C_n) N_1 Z_1.$$

We now set the derivative to 0 and solve for Z_1 to get:

$$Z_1 = (N_1^T (W C_x W^T + C_n) N_1)^{-1} N_1^T W C_{x_1}.$$

Substituting Z_1 back into (34) and using $W_2^T N_1 = 0$ (from (32)) we get:

$$\widehat{W}_1 = C_{x_{11}} W_1^T N_1 (N_1^T (W_1 C_{x_{11}} W_1^T + C_n) N_1)^{-1} N_1^T. \quad (36)$$

$\widehat{W}_2$ can be solved similarly to give:

$$\widehat{W}_2 = C_{x_{22}} W_2^T N_2 (N_2^T (W_2 C_{x_{22}} W_2^T + C_n) N_2)^{-1} N_2^T. \quad (37)$$

Appendix B
CCMSE Approach for TFLM

Our derivation will focus on the inverse operator for the first parameter, τ_1 as the derivation for τ_2 proceeds in a similar manner. We begin by listing some properties of orthogonal projection matrices which will be useful for the subsequent derivation. The orthogonal projection matrix for W_2 is defined by:

$$P_{W_2}^{\perp} = I - W_2 (W_2^T W_2)^{\dagger} W_2^T \quad (38)$$

and can be related to the null space of its transpose (W_2^T) using:

$$P_{W_2}^{\perp} = N_1 N_1^T. \quad (39)$$

It is also straightforward to show by substitution into (38) that:

$$P_{A_2}^{\perp} = P_{W_2}^{\perp}. \quad (40)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2018.2889043, IEEE Transactions on Biomedical Engineering

Using the formula for pseudoinverse of a block matrix [26], we divide the pseudoinverse of the exponential basis matrix A from (10) into components corresponding to each lifetime:

$$A^\dagger = \begin{bmatrix} A_1 & A_2 \end{bmatrix}^\dagger = \begin{bmatrix} (P^\perp_{A_2} A_1)^\dagger \\ (P^\perp_{A_1} A_2)^\dagger \end{bmatrix} = \begin{bmatrix} a_1^\dagger \\ a_2^\dagger \end{bmatrix}. \tag{41}$$

From (41) and using (39) and (40), we can obtain an expression for a_1:

$$a_1 = N_1 N_1^T A_1. \tag{42}$$

Assuming $C_x = \sigma_\eta^2 I$ and $C_n = \sigma_n^2 I$, the CCMMSE estimator from (36) for τ_1 becomes ($N_1^T N_1 = I$):

$$\widehat{W}_1 = W_1^T N_1 (N_1^T W_1 W_1^T N_1 + \lambda I)^{-1} N_1^T, \tag{43}$$

where $\lambda = (\sigma_n/\sigma_\eta)^2$. We next obtain an equivalent overdetermined form for $\widehat{W}_1$ by premultiplying (43) by $(W_1^T N_1 N_1^T W_1 + \lambda I)$ and solving for $\widehat{W}_1$ to get

$$\widehat{W}_1 = (W_1^T N_1 N_1^T W_1 + \lambda I)^{-1} W_1^T N_1 N_1^T. \tag{44}$$

We proceed by substituting the rightmost W_1^T term in (44) with $W_1 = A_1 \overline{W_1}$ and using (42)

$$\begin{aligned} \widehat{W}_1 &= (W_1^T N_1 N_1^T W_1 + \lambda I)^{-1} \overline{W_1}^T a_1^T \\ &= (W_1^T N_1 N_1^T W_1 + \lambda I)^{-1} \overline{W_1}^T a_1^T a_1 a_1^\dagger \\ &= (W_1^T N_1 N_1^T W_1 + \lambda I)^{-1} W_1^T N_1 N_1^T A_1 a_1^\dagger, \end{aligned} \tag{45}$$

where (45) is due to the identity $a_1^T = a_1^T a_1 a_1^\dagger$. Terms corresponding to the overdetermined form of $\widehat{W}_1$ (44) are then replaced with terms corresponding to the undetermined form (43):

$$\begin{aligned} \widehat{W}_1 &= W_1^T N_1 (N_1^T W_1 W_1^T N_1 + \lambda I)^{-1} N_1^T A_1 a_1^\dagger \\ &= \overline{W_1}^T (N_1^T A_1)^T (N_1^T W_1 W_1^T N_1 + \lambda I)^{-1} (N_1^T A_1) a_1^\dagger \\ &= \overline{W_1}^T (((N_1^T A_1)^T)^\dagger)^\dagger ((N_1^T A_1) \overline{W_1 W_1}^T (N_1^T A_1)^T \\ &\quad + \lambda I)^\dagger ((N_1^T A_1)^\dagger)^\dagger a_1^\dagger \\ &= \overline{W_1}^T (\overline{W_1 W_1}^T + \lambda (N_1^T A_1)^\dagger ((N_1^T A_1)^T)^\dagger))^{-1} a_1^\dagger \\ &= \overline{W_1}^T (\overline{W_1 W_1}^T + \lambda ((N_1^T A_1)^T N_1^T A_1)^\dagger)^{-1} a_1^\dagger \\ &= \overline{W_1}^T (\overline{W_1 W_1}^T + \lambda (A_1^T P^\perp_{A_2} A_1)^{-1})^{-1} a_1^\dagger, \end{aligned} \tag{46}$$

where (46) follows from the generalized inverse product rule $(ABC)^\dagger = C^\dagger B^\dagger A^\dagger$ [27].

Using a similar procedure for τ_2. the optimal estimator with equality constraints for both lifetimes can be written as:

$$\widehat{W}_1 = \overline{W_1}^T (\overline{W_1 W_1}^T + \lambda (A_1^T P^\perp_{A_2} A_1)^{-1})^{-1} a_1^\dagger \tag{47}$$

$$\widehat{W}_2 = \overline{W_2}^T (\overline{W_2 W_2}^T + \lambda (A_2^T P^\perp_{A_1} A_2)^{-1})^{-1} a_2^\dagger. \tag{48}$$

Finally, comparing (47) and (48) with the formula for inverse of the block matrix $A^T A$

$$(A^T A)^{-1} = \begin{bmatrix} (A_1^T P^\perp_{A_2} A_1)^{-1} & f(A_1, A_2) \\ g(A_1, A_2) & (A_2^T P^\perp_{A_1} A_2)^{-1} \end{bmatrix}$$

we see that the optimal estimator $\widehat{W}$ can be compactly written as:

$$\widehat{W} = \overline{W}^T (\overline{W W}^T + \lambda DIAG\left((A^T A)^{-1}\right))^\dagger A^\dagger. \tag{49}$$

REFERENCES

[1] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*. CRC press, 1998.

[2] S. S. Hou *et al*., "Tomographic lifetime imaging using combined early- and late-arriving photons," *Optics letters*, vol. 39, no. 5, pp. 1165–1168, 2014.

[3] A. T. Kumar *et al*., "Time resolved fluorescence tomography of turbid media based on lifetime contrast," *Optics express*, vol. 14, no. 25, pp. 12 255–12 270, 2006.

[4] A. Kumar *et al*., "A time domain fluorescence tomography system for small animal imaging," *IEEE transactions on medical imaging*, vol. 27, no. 8, pp. 1152–1163, 2008.

[5] S. S. Hou *et al*., "Optimal estimator for tomographic fluorescence lifetime multiplexing," *Optics letters*, vol. 41, no. 7, pp. 1352–1355, 2016.

[6] P. K. Yalavarthy *et al*., "Weight-matrix structured regularization provides optimal generalized least-squares estimate in diffuse optical tomography," *Medical physics*, vol. 34, no. 6, pp. 2085–2098, 2007.

[7] A. D. Klose, "Radiative transfer of luminescence light in biological tissue," in *Light Scattering Reviews 4*. Springer, 2009, pp. 293–345.

[8] S. R. Arridge and J. C. Schotland, "Optical tomography: forward and inverse problems," *Inverse Problems*, vol. 25, no. 12, p. 123010, 2009.

[9] S. B. Raymond *et al*., "Lifetime-based tomographic multiplexing," *Journal of biomedical optics*, vol. 15, no. 4, p. 046011, 2010.

[10] R. W. Holt *et al*., "Multiple-gate time domain diffuse fluorescence tomography allows more sparse tissue sampling without compromising image quality," *Optics letters*, vol. 37, no. 13, pp. 2559–2561, 2012.

[11] V. Venugopal *et al*., "Quantitative tomographic imaging of intermolecular fret in small animals," *Biomedical optics express*, vol. 3, no. 12, pp. 3161–3175, 2012.

[12] A. T. Kumar *et al*., "Fluorescence-lifetime-based tomography for turbid media," *Optics letters*, vol. 30, no. 24, pp. 3347–3349, 2005.

[13] W. H. Press, *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[14] W. Menke, *Geophysical data analysis: Discrete inverse theory*. Academic press, 2012.

[15] B. W. Pogue *et al*., "Spatially variant regularization improves diffuse optical tomography," *Applied optics*, vol. 38, no. 13, pp. 2950–2961, 1999.

[16] D. Lederman and J. Tabrikian, "Constrained mmse estimator for distribution mismatch compensation," in *Fourth IEEE Workshop on Sensor Array and Multichannel Processing*. IEEE, 2006, pp. 439–443.

[17] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[18] N. I. Gould *et al*., "On the solution of equality constrained quadratic programming problems arising in optimization," *SIAM Journal on Scientific Computing*, vol. 23, no. 4, pp. 1376–1395, 2001.

[19] D. A. Boas *et al*., "Three dimensional monte carlo code for photon migration through complex heterogeneous media including the adult human head," *Optics express*, vol. 10, no. 3, pp. 159–170, 2002.

[20] E. E. Graves *et al*., "A submillimeter resolution fluorescence molecular imaging system for small animal imaging," *Medical physics*, vol. 30, no. 5, pp. 901–911, 2003.

[21] O. Berezovska *et al*., "Amyloid precursor protein associates with a nicastrin-dependent docking site on the presenilin 1–γ-secretase complex in cells demonstrated by fluorescence lifetime imaging," *The Journal of neuroscience*, vol. 23, no. 11, pp. 4560–4566, 2003.

[22] M. Y. Berezin and S. Achilefu, "Fluorescence lifetime measurements and biological imaging," *Chemical reviews*, vol. 110, no. 5, pp. 2641–2684, 2010.

[23] W. Becker, "Fluorescence lifetime imaging–techniques and applications," *Journal of microscopy*, vol. 247, no. 2, pp. 119–136, 2012.

[24] A. Pietraszewska-Bogiel and T. Gadella, "Fret microscopy: from principle to routine technology in cell biology," *Journal of microscopy*, vol. 241, no. 2, pp. 111–118, 2011.

[25] C. J. Goergen *et al*., "In vivo fluorescence lifetime detection of an activatable probe in infarcted myocardium," *Journal of biomedical optics*, vol. 17, no. 5, pp. 0 560 011–0 560 016, 2012.

[26] J. K. Baksalary and O. M. Baksalary, "Particular formulae for the moore–penrose inverse of a columnwise partitioned matrix," *Linear algebra and its applications*, vol. 421, no. 1, pp. 16–23, 2007.

[27] R. Hartwig, "The reverse order law revisited," *Linear algebra and its applications*, vol. 76, pp. 241–246, 1986.