Fluorescence optical diffusion tomography using multiple-frequency data

Adam B. Milstein

School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907-2035

Jonathan J. Stott

Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts 02129

Seungseok Oh

School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907-2035

David A. Boas

Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts 02129

R. P. Millane

Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand

Charles A. Bouman and Kevin J. Webb

School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907-2035

Received September 2, 2003; revised manuscript received December 11, 2003; accepted January 15, 2004

A method is presented for fluorescence optical diffusion tomography in turbid media using multiple-frequency data. The method uses a frequency-domain diffusion equation model to reconstruct the fluorescent yield and lifetime by means of a Bayesian framework and an efficient, nonlinear optimizer. The method is demonstrated by using simulations and laboratory experiments to show that reconstruction quality can be improved in certain problems through the use of more than one frequency. A broadly applicable mutual information performance metric is also presented and used to investigate the advantages of using multiple modulation frequencies compared with using only one. © 2004 Optical Society of America

OCIS codes: 170.6280, 290.7050, 100.3010, 100.3190, 100.6950, 170.3010, 290.3200.

1. INTRODUCTION

Optical diffusion tomography (ODT) or, equivalently, diffuse optical tomography, has attracted considerable recent interest.[1,2] In ODT, volumetric images of the absorption and scattering properties of the tissue are computed from near-infrared boundary measurements. The propagation of light is modeled by use of a partial differential equation known as the diffusion equation. ODT has great potential in a variety of medical imaging applications including tumor imaging and functional brain imaging. In tumor imaging, the specificity and the contrast of ODT can be enhanced through the use of fluorescent agents.[3–12] Injected fluorophores accumulate in diseased tissue because of increased vascular density[7] or selective targeting.[6,8,11,12] The fluorescent lifetime is also used to image cancerous tissue,[5] independently of fluorophore concentration.

In frequency-domain fluorescence optical diffusion tomography (FODT), light at the fluorophore's excitation wavelength is launched into the tissue. The light can be sinusoidally modulated or continuous wave (CW, meaning unmodulated). The fluorophore absorbs the incident light and then decays to its ground state with some characteristic time constant, emitting some of the light at a longer wavelength. The emitted photons are then measured by an array of detection devices. From the data, one can reconstruct images of the fluorescent yield (a measure of the fluorescent efficiency) and the fluorescent lifetime (the fluorescent decay parameter). Multiple photon scattering in tissue must be properly accounted for in the reconstruction.[13,14]

Imaging of absorption, scattering, and fluorescence in diffuse media, however, is computationally intensive. The inversion process is also ill-posed and, frequently, underdetermined. Photon scattering is inherently three dimensional, necessitating full three-dimensional models and reconstructions. Because the optical absorption and scattering properties of the tissue are represented as co-

efficients in the diffusion equation, their reconstruction is a nonlinear optimization problem. If the scattering and the absorption are known, the reconstruction of the fluorescent yield and lifetime is linear after an invertible transformation, provided that only one modulation frequency is used. In frequency-domain imaging with multiple modulation frequencies, though, the inversion is necessarily nonlinear because of coupling between the unknown parameters and the frequency terms. As a result of these difficulties, the inverse problems must be solved iteratively, with robust modeling and optimization techniques.

Previously, several groups addressed the frequency-domain FODT problem by using CW light or light modulated with a single frequency (typically in the range 70–100 MHz).[15–22] In simulation studies, the fluorescent yield and lifetime, given known absorption and scattering, have been simultaneously reconstructed.[15–18] In addition, the fluorescent yield has been reconstructed from laboratory data.[19,20] Shives *et al.*[22] have reconstructed the spatially dependent fluorescent lifetime of an oxygen-sensitive tissue phantom. Recently, we introduced a nonlinear Bayesian inversion method for reconstructing the absorption, scattering, and fluorescence parameters.[21] Building on our earlier work in nonfluorescent image reconstruction,[23–26] our method uses iterative coordinate descent (ICD) optimization and a generalized Gaussian Markov random field (GGMRF) prior model for regularization. We have applied our method to reconstruct the absorption and the fluorescent yield of a tissue phantom composed of Intralipid solution (which scatters the light and has optical properties similar to those of tissue) and a glass sphere containing a mixture of Intralipid solution and indocyanine green (ICG), a fluorescent diagnostic agent.[21]

Several authors have suggested that it should be possible to improve the resolution of ODT reconstructions by using either time-resolved or multiple-frequency data.[27,28] Intuitively, one might expect that additional modulation frequencies would provide additional information that is useful in the reconstruction. To date, however, no one has addressed the question of how to rigorously reconstruct the fluorescent yield and lifetime by use of multiple modulation frequencies or whether multiple-frequency data can improve FODT reconstruction quality. There has been little done on the use of performance metrics for evaluating the impact of multiple frequencies. Here, we present a method for reconstructing the fluorescent yield and lifetime by using multiple modulation frequencies. We show, using numerical simulations, that the proposed method of incorporating multiple-frequency components can indeed improve image quality when reconstructing well-localized objects. We also validate the reconstruction algorithm by reconstructing fluorescent yield, fluorescent lifetime, and absorption of measurements from an experimental tissue phantom containing ICG embedded within a lipid suspension.

Based on the simulation and the experimental results, it appears that the advantage of using multiple modulation frequencies is most apparent in reconstructing well-localized objects and less apparent in reconstructing targets with broader features. To investigate this claim, we present a performance metric based on information theory[29,30] for evaluating an experimental configuration. The performance metric incorporates statistical models of both the unknown image and the measurement device. Hence it provides insight relevant to entire classes of problems, rather than only a few anecdotal examples. Previously, information-theory-based performance metrics have been used to evaluate computed tomography[31] and magnetic resonance imaging.[32] Our performance metric is closely related to previous work by Shao *et al.*, who used mutual information to evaluate different aperture designs for single-photon emission tomography.[33,34] While information theory has not been applied to the ODT problem, several groups have used singular-value analysis to evaluate source/detector arrangements[35,36] or data types.[37] However, none of these investigations incorporated statistical models of the unknown image's properties. Mutual information measures the degree to which the measurement apparatus is matched to the statistical model of the unknown fluorescence image. Using results of rate distortion theory,[38] it provides a lower bound for the mean square error (MSE) for any estimator of the unknown image. Application of the performance metric to various image priors and measurement models agrees well with the reconstruction results in predicting improvement due to use of multiple frequencies.

2. FLUORESCENCE OPTICAL DIFFUSION TOMOGRAPHY MODELS

A. Diffusion Model

The transport of light that is amplitude modulated at a frequency ω through a scattering medium can be modeled with the photon transport equation.[39–41] A common simplification, the photon diffusion equation, is reasonably accurate for imaging applications where there is significant scatter.[1] For $\exp(j\omega t)$ time variation and point source excitation, it is given by

$$\nabla \cdot [D(r)\nabla\phi(r,\omega)] - [\mu_a(r) + j\omega/c]\phi(r,\omega) = -\delta(r - r_{s_k}), \quad (1)$$

where $\phi(r,\omega)\,(\mathrm{W/cm^2})$ is the complex modulation envelope of the photon flux, $\delta(r)$ is the Dirac function, and r_{s_k} is the point source location. The diffusion coefficient $D(r)\,(\mathrm{cm})$ is inversely related to the scattering coefficient, and $\mu_a(r)\,(\mathrm{cm^{-1}})$ is the absorption coefficient.

In a fluorescent scattering medium, the fluorophore is excited with light at wavelength λ_x and emits light at a longer wavelength λ_m (assuming a single emission wavelength, for simplicity). In general, the scattering and absorption properties of the medium, and thus the solution to the diffusion equation, differ between λ_x and λ_m. Hence two diffusion equations must be used to describe a fluorescence measurement: the first, to represent excitation at λ_x, and the second, to represent the propagation of the emitted λ_m photons to the detectors. This coupled diffusion model[13,14,42] is

$$\nabla \cdot [D_x(r)\nabla\phi_x(r,\omega)] - [\mu_{a_x}(r) + j\omega/c]\phi_x(r,\omega) = -\delta(r - r_{s_k}), \quad (2)$$

$$\nabla \cdot [D_m(r)\nabla\phi_m(r,\omega)] - [\mu_{a_m}(r) + j\omega/c]\phi_m(r,\omega) = -\phi_x(r,\omega)\,\eta\mu_{a_f}(r)\frac{1 - j\omega\tau(r)}{1 + [\omega\tau(r)]^2}, \quad (3)$$

where the subscripts x and m denote excitation and emission wavelengths λ_x and λ_m, respectively. We also assume single exponential decay in this model. The fluorescence parameters are the lifetime $\tau(r)$ (s) and the yield $\eta\mu_{a_f}(r)$ (cm^{-1}). The fluorescent yield $\eta\mu_{a_f}$ incorporates the fluorophore's quantum efficiency η (which depends on the type of fluorophore and the chemical environment) and its absorption coefficient μ_{a_f} (which depends on the fluorophore concentration). The fluorescent lifetime $\tau(r)$ is the exponential decay constant. The coupling of the two equations takes place through the right-hand side of Eq. (3).

B. Tomography Problem

In general, one does not know μ_{a_x}, μ_{a_m}, D_x, D_m, τ, and $\eta\mu_{a_f}$ *a priori* in a practical experiment. To obtain the fluorescence parameters, one would first need to perform appropriate measurements and use them to reconstruct μ_{a_x}, D_x, μ_{a_m}, and D_m, because of their appearance in the coupled diffusion equations. Previously, we have presented a suitable measurement and reconstruction approach,[21] depicted schematically in Fig. 1. Reconstructions of D_x and μ_{a_x} may be obtained by using data from sources and detectors at the excitation wavelength λ_x. Similarly, D_m and μ_{a_m} may be obtained by using data from sources and detectors at the emission wavelength λ_m. Finally, measurements with sources at λ_x and detectors filtered at λ_m are used to reconstruct τ and $\eta\mu_{a_f}$. In some practical situations, sufficiently accurate reconstructions of τ and $\eta\mu_{a_f}$ might be obtainable without performing rigorous reconstructions of all of the remaining unknowns, under reasonable simplifying assumptions.[20,21]

We discretize the domain into N voxels of equal size (although one can generalize to irregular meshes). Let r_i denote the position of the ith voxel centroid, and let η denote $\eta\mu_{a_f}$, for brevity. We define three image vectors, with each corresponding to a particular measurement:

$$x_x = [x_{xA}^{\mathrm{T}}, x_{xB}^{\mathrm{T}}]^{\mathrm{T}} = [\mu_{a_x}(r_1),\ldots,\mu_{a_x}(r_N), D_x(r_1),\ldots,D_x(r_N)]^{\mathrm{T}}, \quad (4)$$

$$x_m = [x_{mA}^{\mathrm{T}}, x_{mB}^{\mathrm{T}}]^{\mathrm{T}} = [\mu_{a_m}(r_1),\ldots,\mu_{a_m}(r_N), D_m(r_1),\ldots,D_m(r_N)]^{\mathrm{T}}, \quad (5)$$

$$x_f = [x_{fA}^{\mathrm{T}}, x_{fB}^{\mathrm{T}}]^{\mathrm{T}} = [\eta(r_1),\ldots,\eta(r_N), \tau(r_1),\ldots,\tau(r_N)]^{\mathrm{T}}, \quad (6)$$

where the subscript f denotes the fluorescence image and the superscript T denotes the transpose operation. Note that the three image vectors are each of size $2N$, consisting of two unknown parameter vectors of size N. The photon flux measurement vectors corresponding to the above image vectors may be defined, respectively, as y_x, y_m, and y_f.

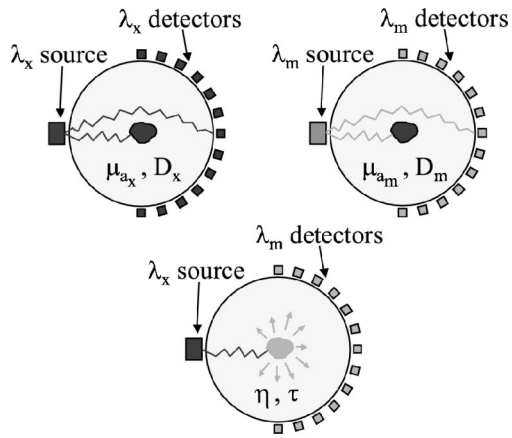


Fig. 1. Measurement approach for reconstructing all unknowns, showing appropriate source and detector wavelengths for reconstructing $[\mu_{a_x}, D_x]$, $[\mu_{a_m}, D_m]$, and $[\eta, \tau]$.

3. INVERSE PROBLEM

A. Bayesian Framework

The estimation of $\{x_x, x_m, x_f\}$ from $\{y_x, y_m, y_f\}$ is an ill-posed inverse problem. To address this, we formulate the inversion in a Bayesian framework,[21,23–26] as we have presented previously. Let x denote one of the images $\{x_x, x_m, x_f\}$, and let y denote its corresponding data set. We compute the maximum *a posteriori* (MAP) estimate by using Bayes' rule:

$$\hat{x}_{\mathrm{MAP}} = \arg\max_{x\geq 0}[p_{X|Y}(x|y)] = \arg\max_{x\geq 0}[\log p_{Y|X}(y|x) + \log p_X(x)], \quad (7)$$

where $p_{X|Y}(x|y)$ is the posterior density, $p_{Y|X}(y|x)$ is the data likelihood, and $p_X(x)$ is the prior density for the image. The circumflex denotes estimated quantities. For the data likelihood, we assume an independent Gaussian model

$$p_{Y|X}(y|x) = \frac{1}{(\pi\alpha)^P|\Lambda|^{-1}}\exp\left(-\frac{\|y - f(x)\|_\Lambda^2}{\alpha}\right), \quad (8)$$

where P is the number of measurements, α is a scalar parameter that scales the noise variance, $\|w\|_\Lambda^2 = w^{\mathrm{H}}\Lambda w$ (where H denotes Hermitian transpose), and $\alpha\Lambda^{-1}/2$ is the covariance matrix.[23] The appropriate forward operator f is computed by solving the diffusion equations for the photon flux at the detector positions due to the point sources. We use the same approximate shot-noise model as that of Ye *et al.*[23,24] and adopt the same notation:

$$\frac{\alpha}{2}\Lambda^{-1} = \frac{\alpha}{2}\mathrm{diag}(|y_1|, |y_2|,\ldots,|y_P|). \quad (9)$$

For $p_X(x)$, we use the generalized Gaussian Markov random field (GGMRF) model, which enforces smoothness in the solution while preserving sharp edge transitions.[23,43] We define a three-dimensional neighborhood system around each node (voxel) from the 26 adjacent nodes. Defining $x^{\mathrm{T}} = [x_A^{\mathrm{T}}, x_B^{\mathrm{T}}]$, as in Eqs. (4)–(6), we assume that x_A and x_B are independent:

$$p_X(x) = p_{X_A}(x_A)p_{X_B}(x_B) \tag{10}$$

$$= \frac{1}{\sigma_A^N \zeta(\rho_A)} \times \exp\left(-\frac{1}{\rho_A \sigma_A^{\rho_A}} \sum_{\{i,j\}\in\mathcal{N}_A} b_{i-j}|x_i - x_j|^{\rho_A} \right) \times \frac{1}{\sigma_B^N \zeta(\rho_B)} \times \exp\left(-\frac{1}{\rho_B \sigma_B^{\rho_B}} \sum_{\{i,j\}\in\mathcal{N}_B} b_{i-j}|x_i - x_j|^{\rho_B} \right), \tag{11}$$

where A and B correspond to Eqs. (4)–(6), x_i is the ith node of x, $\mathcal{N}$ consists of all pairs of neighboring (adjacent) nodes, and b_{i-j} is the weighting coefficient corresponding to the ith and jth nodes. We assign the coefficients b_{i-j} to be inversely proportional to the node separation in a cube-shaped node layout, requiring that $\Sigma_j b_{i-j} = 1$. The hyperparameters ρ and σ control the shape and the scale of the distribution, and the partition function $\zeta(\rho)$ normalizes the density function.

We consider α to be an unknown instrument parameter, and we incorporate it into the inverse problem. Previously, we have found that this tends to improve the robustness and the speed of convergence.[24] The FODT estimation problem becomes

$$\hat{x}_x = \arg\max_{x_x\geq 0,\alpha_x\geq 0}[p_{X_x|Y_x}(x_x|y_x, \alpha_x)], \tag{12}$$

$$\hat{x}_m = \arg\max_{x_m\geq 0,\alpha_m\geq 0}[p_{X_m|Y_m}(x_m|y_m, \alpha_m)], \tag{13}$$

$$\hat{x}_f = \arg\max_{x_f\geq 0,\alpha_f\geq 0}[p_{X_f|Y_f}(x_f|y_f, \alpha_f, \hat{x}_x, \hat{x}_m)]. \tag{14}$$

The estimations of x_x and x_m are performed initially, and these estimates are incorporated into the coupled diffusion equations (2) and (3) to estimate x_f. To solve the above joint MAP estimation problems, we follow Ye *et al.*[24] by maximizing the log posterior probability $l(x)$:

$$l(x) = -P\log\|y - f(x)\|_\Lambda^2 - \frac{1}{\rho_A \sigma_A^{\rho_A}} \sum_{\{i,j\}\in\mathcal{N}_A} b_{i-j}|x_i - x_j|^{\rho_A} - \frac{1}{\rho_B \sigma_B^{\rho_B}} \sum_{\{i,j\}\in\mathcal{N}_B} b_{i-j}|x_i - x_j|^{\rho_B}. \tag{15}$$

Equation (15) can be implemented by alternating closed-form updates of $\hat{\alpha}$ with updates of $\hat{x}$ (Ref. 24):

$$\hat{\alpha} \leftarrow \frac{1}{P}\|y - f(\hat{x})\|_\Lambda^2, \tag{16}$$

$$\hat{x} \leftarrow \arg\operatorname{update}_{x\geq 0}[\log p_{Y|X}(y|x, \hat{\alpha}) + \log p_X(x|\hat{\alpha})], \tag{17}$$

where "←" denotes assignment and "arg update" denotes an iteration of some optimizer. To implement expression (17), we form objective functions from Eqs. (8), (11), and (15):

$$c(x_x, \hat{\alpha}_x) = \frac{1}{\hat{\alpha}_x}\|y_x - f_x(x_x)\|_{\Lambda_x}^2 + \frac{1}{\rho_{xA}\sigma_{xA}^{\rho_{xA}}} \sum_{\{i,j\}\in\mathcal{N}_{xA}} b_{i-j}|x_{xA_i} - x_{xA_j}|^{\rho_{xA}} + \frac{1}{\rho_{xB}\sigma_{xB}^{\rho_{xB}}} \sum_{\{i,j\}\in\mathcal{N}_{xB}} b_{i-j}|x_{xB_i} - x_{xB_j}|^{\rho_{xB}}, \tag{18}$$

$$c(x_m, \hat{\alpha}_m) = \frac{1}{\hat{\alpha}_m}\|y_m - f_m(x_m)\|_{\Lambda_m}^2 + \frac{1}{\rho_{mA}\sigma_{mA}^{\rho_{mA}}} \sum_{\{i,j\}\in\mathcal{N}_{mA}} b_{i-j}|x_{mA_i} - x_{mA_j}|^{\rho_{mA}} + \frac{1}{\rho_{mB}\sigma_{mB}^{\rho_{mB}}} \sum_{\{i,j\}\in\mathcal{N}_{mB}} b_{i-j}|x_{mB_i} - x_{mB_j}|^{\rho_{mB}}, \tag{19}$$

$$c(x_f, \hat{x}_x, \hat{x}_m, \hat{\alpha}_f) = \frac{1}{\hat{\alpha}_f}\|y_f - f_f(x_f, \hat{x}_x, \hat{x}_m)\|_{\Lambda_f}^2 + \frac{1}{\rho_{fA}\sigma_{fA}^{\rho_{fA}}} \sum_{\{i,j\}\in\mathcal{N}_{fA}} b_{i-j}|x_{fA_i} - x_{fA_j}|^{\rho_{fA}} + \frac{1}{\rho_{fB}\sigma_{fB}^{\rho_{fB}}} \sum_{\{i,j\}\in\mathcal{N}_{fB}} b_{i-j}|x_{fB_i} - x_{fB_j}|^{\rho_{fB}}, \tag{20}$$

where the subscripts have the same meaning as those in Eqs. (4)–(6).

B. Definitions

Previously,[21,25] we have shown how to reconstruct x_x and x_m in a framework that can incorporate multiple frequencies. Hence, suppose that x_x and x_m have been reconstructed or are known by some other means. The remaining unknowns, η and τ, appear on the right-hand side of Eq. (3) as part of the source term. Let

$$h(x_f, r, \omega) = \eta(r)\frac{1 - j\omega\tau(r)}{1 + [\omega\tau(r)]^2}. \tag{21}$$

Also, let $g(r_{s_k}, r_{d_{m'}}; \omega, x)$ be the diffusion equation Green's function for the problem domain computed by using the image vector x and a numerical forward solver, with r_{s_k} as the source location, $r_{d_{m'}}$ as the observation point, and modulation frequency ω. More specifically, let $g_x(r_{s_k}, r_{d_{m'}}; \omega, x_x)$ be the Green's function for wavelength λ_x, and let $g_m(r_{s_k}, r_{d_{m'}}; \omega, x_m)$ be the Green's function at λ_m. Set

$$\phi_f(r_{s_k}, r_{d_{m'}}; \omega, x_f) = \int h(x_f, r, \omega) g_x(r_{s_k}, r; \omega, x_x) \times g_m(r, r_{d_{m'}}; \omega, x_m) \mathrm{d}^3 r, \quad (22)$$

where we omit the dependence on x_x and x_m for simplicity. From Eq. (3),

$$h(x_f, r, \omega) g_x(r_{s_k}, r; \omega, x_x) = \phi_x(r, \omega) \frac{\eta(r)[1 - j\omega\tau(r)]}{1 + [\omega\tau(r)]^2}. \quad (23)$$

Suppose that we have K sources and M detectors at a modulation frequency ω. Let $f_\omega(x_f)$ be the forward model describing the expected value of the data given x_f. Then

$$f_\omega(x_f) = \begin{pmatrix} \phi_f(r_{s_1}, r_{d_1}; \omega, x_f) \\ \phi_f(r_{s_1}, r_{d_2}; \omega, x_f) \\ \vdots \\ \phi_f(r_{s_1}, r_{d_M}; \omega, x_f) \\ \phi_f(r_{s_2}, r_{d_1}; \omega, x_f) \\ \vdots \\ \phi_f(r_{s_K}, r_{d_M}; \omega, x_f) \end{pmatrix}. \quad (24)$$

Let Q be the number of modulation frequencies used, and let

$$f(x_f) = [[f_{\omega_1}(x_f)]^{\mathrm{T}}, [f_{\omega_2}(x_f)]^{\mathrm{T}}, \ldots, [f_{\omega_Q}(x_f)]^{\mathrm{T}}]^{\mathrm{T}}. \quad (25)$$

Similarly, we define the measurement vector y as:

$$y = [y_{\omega_1}^{\mathrm{T}}, y_{\omega_2}^{\mathrm{T}}, \ldots, y_{\omega_Q}^{\mathrm{T}}]^{\mathrm{T}}, \quad (26)$$

corresponding to the same order as that used in Eq. (25). Note that $g(r_{s_k}, r_{d_{m'}}, \cdot, \cdot) = g(r_{d_{m'}}, r_{s_k}, \cdot, \cdot)$ at λ_x and at λ_m, as a result of the reciprocity theorem.[41]

For the discretized problem, we define matrix operators to approximate the integration in Eq. (22):

$$G^x(\omega) = \begin{bmatrix} g_x(r_{s_1}, r_1; \omega, x_x) & \cdots & g_x(r_{s_1}, r_N; \omega, x_x) \\ \vdots & \ddots & \vdots \\ g_x(r_{s_K}, r_1; \omega, x_x) & \cdots & g_x(r_{s_K}, r_N; \omega, x_x) \end{bmatrix}, \quad (27)$$

$$G^m(\omega) = \begin{bmatrix} g_m(r_{d_1}, r_1; \omega, x_m) & \cdots & g_m(r_{d_1}, r_N; \omega, x_m) \\ \vdots & \ddots & \vdots \\ g_m(r_{d_M}, r_1; \omega, x_m) & \cdots & g_m(r_{d_M}, r_N; \omega, x_m) \end{bmatrix}, \quad (28)$$

$$J_\omega = V \begin{bmatrix} G^x_{1,1}(\omega) G^m_{1,1}(\omega) & \cdots & G^x_{1,N}(\omega) G^m_{1,N}(\omega) \\ \vdots & \ddots & \vdots \\ G^x_{1,1}(\omega) G^m_{M,1}(\omega) & \cdots & G^x_{1,N}(\omega) G^m_{M,N}(\omega) \\ G^x_{2,1}(\omega) G^m_{1,1}(\omega) & \cdots & G^x_{2,N}(\omega) G^m_{1,N}(\omega) \\ \vdots & \ddots & \vdots \\ G^x_{K,1}(\omega) G^m_{M,1}(\omega) & \cdots & G^x_{K,N}(\omega) G^m_{M,N}(\omega) \end{bmatrix}, \quad (29)$$

where V is the volume of a voxel. Let

$$h_\omega(x_f) = [h(x_f, r_1, \omega), \ldots, h(x_f, r_N, \omega)]^{\mathrm{T}}. \quad (30)$$

Neglecting discretization error, we have

$$f_\omega(x_f) = J_\omega h_\omega(x_f). \quad (31)$$

Suppressing the $\hat{x}_x$ and $\hat{x}_m$ arguments for brevity, we can rewrite Eq. (20) as

$$c(x_f, \hat{\alpha}_f) = \frac{1}{\hat{\alpha}_f} \sum_{q=1}^{Q} \| y_{f\omega_q} - J_{\omega_q} h_{\omega_q}(x_f) \|^2_{\Lambda_{f\omega_q}} + \frac{1}{\rho_{fA} \sigma_{fA}^{\rho_{fA}}} \sum_{\{i,j\} \in \mathcal{N}_{fA}} b_{i-j} |x_{fA_i} - x_{fA_j}|^{\rho_{fA}} + \frac{1}{\rho_{fB} \sigma_{fB}^{\rho_{fB}}} \sum_{\{i,j\} \in \mathcal{N}_{fB}} b_{i-j} |x_{fB_i} - x_{fB_j}|^{\rho_{fB}}. \quad (32)$$

In the single-frequency case, we can take advantage of the linearity of $f_\omega(x_f)$ with respect to $h_\omega(x_f)$ by reparameterizing $\{\eta, \tau\}$ to create a quadratic optimization problem.[21] However, since the resulting parameters must contain frequency terms in their denominators, they are inappropriate for the multiple-frequency case. Hence we perform a nonlinear optimization of Eq. (32) directly over $\{\eta, \tau\}$.

C. Iterative Coordinate Descent

As in our previous work,[21,23–26] we use the iterative coordinate descent (ICD) algorithm,[44] a Gauss–Seidel approach, to optimize Eq. (32). The voxels are scanned in random order, and the cost function is optimized with respect to each individual voxel. Previously,[21] we have shown how to optimize Eqs. (18) and (19). Hence we focus on Eq. (20) and omit the f subscript here for simplicity. In one update scan for $\hat{x}$, we update all of the N voxels with respect to $x_A = \eta$ and subsequently update all of the voxels with respect to $x_B = \tau$. Let the scalar x_i denote the ith element of x. With all other image elements fixed, the ICD update for the estimate $\hat{x}_i$ is given by

$$\hat{x}_i \leftarrow \arg\min_{x_i \geq 0} \left\{ \frac{1}{\hat{\alpha}} \sum_{q=1}^{Q} \| y_{\omega_q} - [J_{\omega_q}]_{*(i)} h(x, r_i, \omega_q) \|^2_{\Lambda_{\omega_q}} + \frac{1}{\rho \sigma^{\rho}} \sum_{j \in \mathcal{N}_i} b_{i-j} |x_i - \hat{x}_j|^{\rho} \right\}, \quad (33)$$

where $\mathcal{N}_i$ is the set of nodes neighboring node i, and ρ and σ are chosen appropriately from $\{\rho_A, \rho_B\}$ and $\{\sigma_A, \sigma_B\}$. $[J_{\omega_q}]_{*(i)}$ denotes the ith column of J_{ω_q}. Suppose that we have an initial guess $\tilde{x}$ and let $z_{\omega_q} = y_{\omega_q} - f_{\omega_q}(\tilde{x})$. Then relation (33) is equivalent to

$$\hat{x}_i \leftarrow \arg\min_{x_i \geq 0} \left\{ \frac{1}{\hat{\alpha}} \sum_{q=1}^{Q} \| z_{\omega_q} - [J_{\omega_q}]_{*(i)} [h(x, r_i, \omega_q) - h(\tilde{x}, r_i, \omega_q)] \|^2_{\Lambda_{\omega_q}} + \frac{1}{\rho \sigma^{\rho}} \sum_{j \in \mathcal{N}_i} b_{i-j} |x_i - \hat{x}_j|^{\rho} \right\},$$

$$= \arg\min_{x_i \geq 0} \left(\frac{1}{\hat{\alpha}} \sum_{q=1}^{Q} \left\{ \theta_{1,\omega_q} [h(x, r_i, \omega_q) - h(\tilde{x}, r_i, \omega_q)] + \frac{\theta_{2,\omega_q}}{2} [h(x, r_i, \omega_q) - h(\tilde{x}, r_i, \omega_q)]^2 \right\} + \frac{1}{\rho \sigma^{\rho}} \sum_{j \in \mathcal{N}_i} b_{i-j} |x_i - \tilde{x}_j|^{\rho} \right), \quad (34)$$

where

$$\theta_{1,\omega_q} = -2 \, \mathrm{Re}([J_{\omega_q}]^H_{*i} \Lambda_{\omega_q} z_{\omega_q}), \quad (35)$$

$$\theta_{2,\omega_q} = 2 [J_{\omega_q}]^H_{*i} \Lambda_{\omega_q} [J_{\omega_q}]_{*i}. \quad (36)$$

In relation (34), θ_{1,ω_q} and θ_{2,ω_q} are not functions of x_i and thus do not need to be recomputed during the nonlinear, one-dimensional line search over x_i. This property permits significant computational savings, as repeated computations of θ_{1,ω_q} and θ_{2,ω_q} would require numerous complex multiplications. We perform the minimization over x_i by use of a Golden Section search.[45]

We implement the joint estimation of α and x iteratively. One iteration consists of a closed-form update of $\hat{\alpha}$ with the use of relation (16), followed by ICD scans to update all of the voxels in $\hat{x}_A$ and $\hat{x}_B$. Appendix A summarizes the ICD optimization algorithm in pseudocode form.

4. SIMULATION

We performed a simulation incorporating multiple modulation frequencies by use of the proposed algorithm. Figure 2(a) shows an $8 \times 8 \times 5.7$ cm cubic tissue phantom, with five sources on one side and five detectors on the other. Two small η heterogeneities were placed near each other, slightly off center. The other parameters were constant, with $\mu_{a_{x,m}} = 0.047\,\mathrm{cm}^{-1}$, $D_{x,m} = 0.27$ cm, and $\tau = 0$ s. Three data sets were computed numerically, using multigrid finite differences on a $33 \times 33 \times 17$ grid: 78.4-MHz data, 314-MHz data, and a combination of both. We used the extrapolated zero-flux boundary condition[40] with internal sources interpolated among the nearest grid nodes. The sources were placed 0.08 cm inside the physical boundary, corresponding approximately to one transport mean free path. The voxels were $0.26 \times 0.26 \times 0.38$ cm in size. Gaussian noise was added, using the shot-noise model presented by Ye *et al.*[23] and assuming equal input source power for each modulation frequency. The average signal-to-noise ratio was 21.2 dB for the 78.4-MHz data and 14.7 dB for the 314-MHz data. With the same $33 \times 33 \times 17$ grid, the ICD algorithm was used to reconstruct η, with a constant initial guess of 0 cm^{-1}. The shape and scale hyperparameters σ and ρ were set to 0.021 cm^{-1} and 2, respectively,

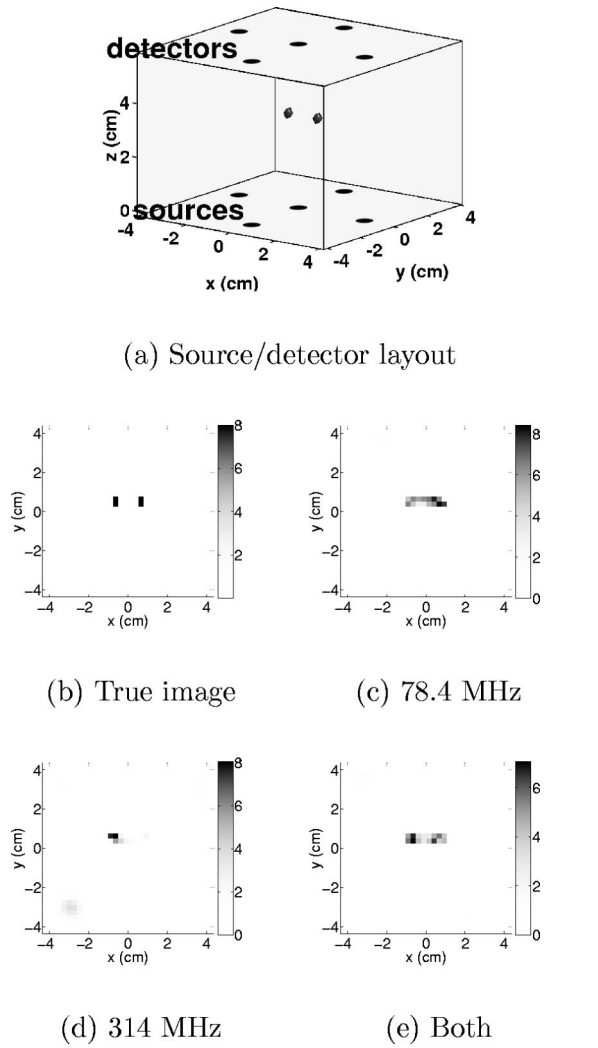


Fig. 2. Reconstruction of fluorescent yield using simulated data, showing the improvement due to use of multiple modulation frequencies: (a) source/detector geometry, (b) true image cross section, (c) reconstruction using 78.4-MHz data, (d) reconstruction using 314-MHz data, (e) reconstruction using 78.4- and 314-MHz data.

for all three reconstructions. For each reconstruction, using $\sigma = 0.021\,\mathrm{cm}^{-1}$ gave the best results. The algorithm was run to convergence, with subsequent iterations changing the cost function negligibly. To avoid singularities near the sources and the detectors, we did not update within a 2-voxel border of the computational boundary. Figure 2(b) shows the true image cross section at $z = 2.85$ cm. Figures 2(c)–2(e) show the results. In the single-frequency reconstructions, the two objects are not clearly distinct. However, in the multiple-frequency reconstruction, the two objects are clearly distinguishable. The result suggests that the proposed method for making use of multiple-frequency components can improve reconstruction quality in some cases.

5. EXPERIMENT

A. Design and Procedure

To test the method, we prepared a time-domain fluorescence and absorption imaging experiment, shown sche-

matically in Fig. 3(a). The tissue phantom comprised a 2-cm glass sphere containing ICG and 1% Intralipid (a fat emulsion) embedded within a 1% Intralipid suspension. As in earlier work,[21] a measurement box contained the phantom and fiber-optic connectors. The box had inner dimensions of 17×17 cm, with adjustable height. A grid of 16×16 fiber-optic connectors, with 1-cm spacing, was used to position source fibers on the bottom of the box. We selected 24 source positions, as shown in Fig. 3. A tunable, mode-locked Spectra-Physics MaiTai Ti:sapphire pulsed laser was used as the source. The pulse width was less than 100 fs, according to manufacturer's data. The 0.8-W average output of the laser was split by using the reflection off a glass microscope cover slip, and 20 mW was coupled into the source fibers. A galvanometer-based optical scanner from Nutfield Technologies was used to sequentially couple the beam into the source fibers. The top of the phantom box was opened, and a LaVision PicoStar image-intensified CCD camera imaged the exposed top surface of the Intralipid. Effective detector positions were sampled from the CCD images by integrating over 24 square regions, each 4 mm wide. The effective detector positions were in the same arrangement as that of the sources in Fig. 3.

A 1.0% solution of Intralipid and distilled water was added to the box to a depth of 5.7 cm, and to the sphere. The sphere was measured to be vertically centered at $z \simeq 2.8$ cm. However, the sphere was suspended only by flexible rubber tubes, which added an uncertainty of roughly 2–3 mm to the estimated vertical position. The rubber tubes were used for titrating the fluorophore from a separate reservoir of Intralipid solution.

Initially, baseline measurements were performed on the essentially homogeneous slab, without the fluorophore present. These baseline measurements were incorporated into a calibration scheme that we have used previously.[21] For the purpose of reconstructing μ_{a_x}, impulse responses were measured with the laser tuned to 780 nm and the intensifier voltage set to 350 V. Each sample of the temporal response was recorded by gating the intensifier for 1 ns and integrating for 75 ms, with the gating triggered at the laser's repetition rate of 80 MHz. Using the intensifier's adjustable triggering delay, we recorded samples at 250-ps-delay intervals. Because of the 80-MHz repetition rate, the impulse responses were measured over a range of 12.5 ns, which was sufficient to encompass the entire impulse response duration. Similarly, for the μ_{a_m} reconstruction baseline data, the laser was tuned to 830 nm, the ICG emission wavelength, and the above measurement procedure was repeated. No baseline data were recorded for reconstructing η and τ, as no Intralipid fluorescence was observed. After collection of the baseline data, ICG was titrated into the sphere at a concentration of 0.125 μM, and impulse responses were recorded as above. To measure the fluorescence, we tuned the laser to 780 nm, and an 830-nm bandpass filter with 10-nm FWHM (CVI Laser Corporation) was placed in front of the camera. Because the fluorescence measurements were dimmer than the previous measurements, the intensifier voltage was increased to 550 V, the samples were collected at 500-ps intervals (thus decreasing the temporal resolution and the collection time), and the integration time per sample was increased to 1 s, while the gate width remained at 1 ns. We did not need to account for the finite gate width in the reconstruction, as the normalizations performed in the calibrations cancel any windowing effects.

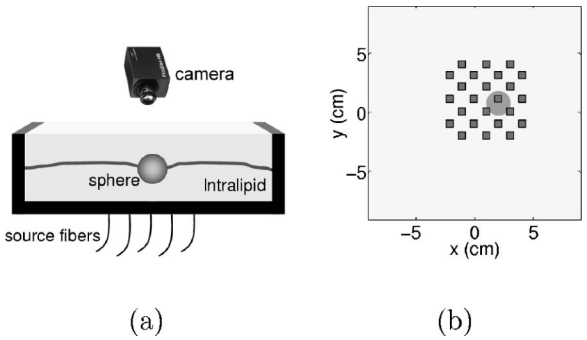


Fig. 3. (a) Schematic of the experimental setup, showing the box and the tissue phantom and a glass sphere filled with ICG/Intralipid, rubber tubes, and Intralipid suspension. (b) Source fiber positions. The same positions were selected as detection regions from the camera images.

Some processing of the data was necessary before proceeding with the reconstructions. To remap the CCD camera images into two-dimensional uniform grid coordinates, we estimated, in a least-squares sense, the projection transformation between the points of maximum brightness (assumed to be directly above the sources) and the known grid locations of their corresponding source fibers. This transformation was applied to convert between detector grid locations and image pixel coordinates. For both λ_x and λ_m, the background μ_a and D and the unknown initial time offset were determined by fitting one of the point-spread functions to the analytical Green's function for a homogeneous slab.[46] The point-spread functions were Fourier transformed, and the 78-, 314-, and 627-MHz components were selected. Calibration factors were obtained by computing simulated measurements on a homogeneous domain discretized into $33 \times 33 \times 17$ voxels ($0.51 \times 0.51 \times 0.38$ cm in size) and dividing these results by the corresponding baseline measurements. These factors were used to calibrate the data with the procedure of our previous work.[21]

B. Reconstructions

To reconstruct μ_{a_x} and μ_{a_m} from the measured data, we used the ICD algorithm presented previously[21,25] to optimize Eqs. (18) and (19). The domain was discretized into $33 \times 33 \times 17$ voxels, $0.51 \times 0.51 \times 0.38$ cm in size. The reconstructions were initialized with the constant background μ_a and D values estimated from the baseline impulse responses recorded without the fluorophore. Both D_x and D_m were assumed to be constant and not modified from the initial values. Although automatic estimation of the GGMRF hyperparameters ρ and σ is possible in principle,[47] we follow our previous work[21,23] and use parameter values that empirically give good results. For the μ_{a_x} and μ_{a_m} reconstructions, the algorithm was run to convergence (i.e., until ICD iterations produce negligible change in the cost function's value). We used model parameters $\sigma = 0.005\,\mathrm{cm}^{-1}$ and $\rho = 2$. The estimates $\hat{\mu}_{a_x}$

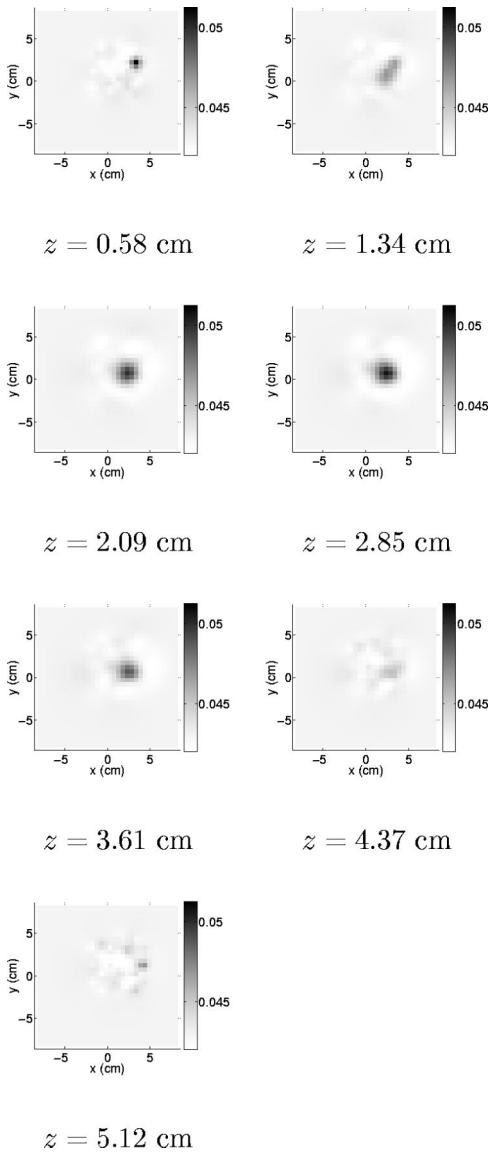


Fig. 4. Reconstruction of $\mu_{a_x}(\mathrm{cm}^{-1})$, obtained by using 78-, 314-, and 627-MHz data.

and $\hat{\mu}_{a_m}$ were incorporated into the coupled diffusion equations, and η and τ were reconstructed by using the proposed algorithm.

Figures 4 and 5 show the reconstructions of μ_{a_x} and μ_{a_m}, respectively, obtained by using the 78-, 314-, and 627-MHz data. The peak value of the μ_{a_x} deviation from the background is smaller than the value of 0.016 cm^{-1} that one would obtain by dividing a previously reported[14] 1-μM ICG absorption value by 8 to account for the concentration difference. (We note, however, that an ICG solution's absorption and fluorescence may not be linear with concentration, nor constant over time.[48,49]) The μ_{a_m} peak value is consistent with our previous results.[21] Aside from a significant artifact near one of the sources, the images are also reasonable. The reconstructed sphere's center is approximately 4 mm below the center of the box (where the true sphere lies with an uncertainty of approximately 2–3 mm), and its diameter is close to the true diameter of 2 cm.

The fluorescence properties were also reconstructed by using 78-, 314-, and 627-MHz data. Figure 6 shows the reconstruction of η, obtained by using $\sigma = 2.5 \times 10^{-5}\,\mathrm{cm}^{-1}$ and $\rho = 2$. The effects of changing the intensifier voltage and inserting the bandpass filter were accounted for in the calibration by using manufacturer's data. The image is accurate, with few artifacts and a position similar to that of the reconstructed absorber. Dividing the peak fluorescent yield by the peak absorption of the heterogeneity at λ_x, we obtain a quantum efficiency of approximately 0.018. This result is slightly higher than the value of 0.016 obtained by Sevick-Muraca *et al.*,[14] who measured a micromolar aqueous solution with a spectrofluorometer. Figure 7 shows the reconstructed τ, obtained by using $\sigma = 0.1\,\mathrm{ns}$ and $\rho = 2$. Quantitatively, it is similar to a literature-reported lifetime of 0.56 ns.[17] The image appears spread out, owing to the fact that τ can be nonzero in regions where $\eta \simeq 0$, with small effect (as τ is multiplied by η in the diffusion equation).

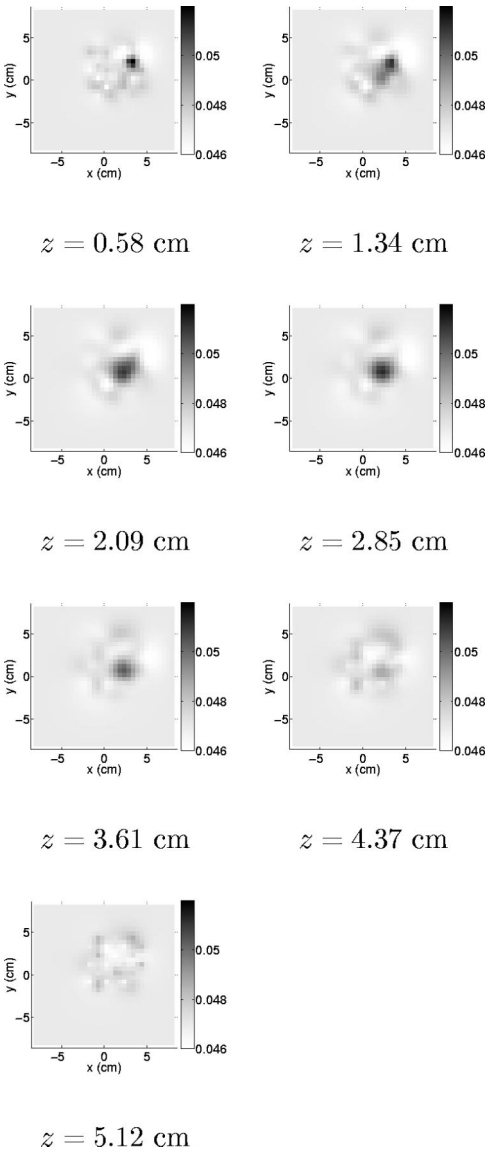


Fig. 5. Reconstruction of $\mu_{a_m}(\mathrm{cm}^{-1})$, obtained by using 78-, 314-, and 627-MHz data.

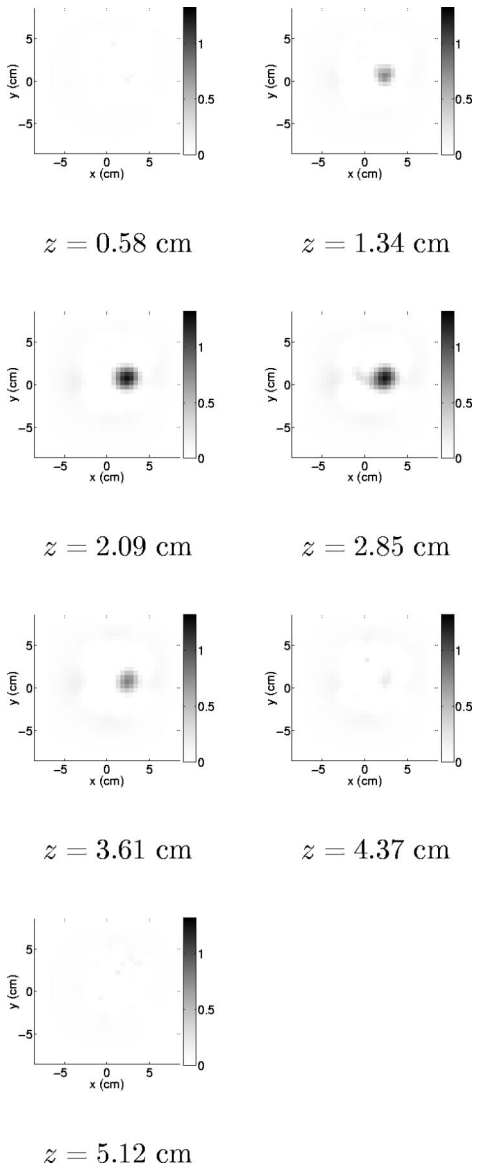


Fig. 6. Reconstruction of η (in $10^{-4}\,\mathrm{cm}^{-1}$), obtained by using 78-, 314-, and 627-MHz data.

Figures 8 and 9 show reconstructions of η and τ obtained by using only 78-MHz data. The results are very similar to the reconstructions obtained by using 78-, 314-, and 627-MHz data together. Hence, for this experiment, the additional modulation frequencies (perhaps surprisingly) do not appear to contribute information that is useful for reconstructing the sphere.

C. Effects of Regularization

Figures 10 and 11 show reconstructions of η and τ, respectively, obtained over a range of σ values. The η reconstructions in Fig. 10 show a progression from overregularization to underregularization with increasing σ (where the τ prior parameters were fixed to $\rho = 2$ and $\sigma = 0.1$ ns, as used above). Figure 10(a), which we deem to be overregularized, appears overly broadened, with blurred edges. In contrast, Fig. 10(e) is underregularized, with sharp peaks appearing in the image. The τ images in Fig. 11 were also computed over a range of σ values, with the η prior parameters fixed to $\rho = 2$ and $\sigma = 2.5 \times 10^{-4}\,\mathrm{cm}^{-1}$. For the lifetime images, discerning underregularized and overregularized images is more complicated. Because τ is multiplied by η in the diffusion equations, overregularized τ images may appear reasonable and not excessively broadened. Figures 11(d) and 11(e), which are deemed to be underregularized, have much of their structure outside the voxels containing the fluorophore in Fig. 6 rather than within it.

One method of selecting σ for the fluorescence lifetime problem might be to choose the largest one that gives physically acceptable images. Less regularization im-

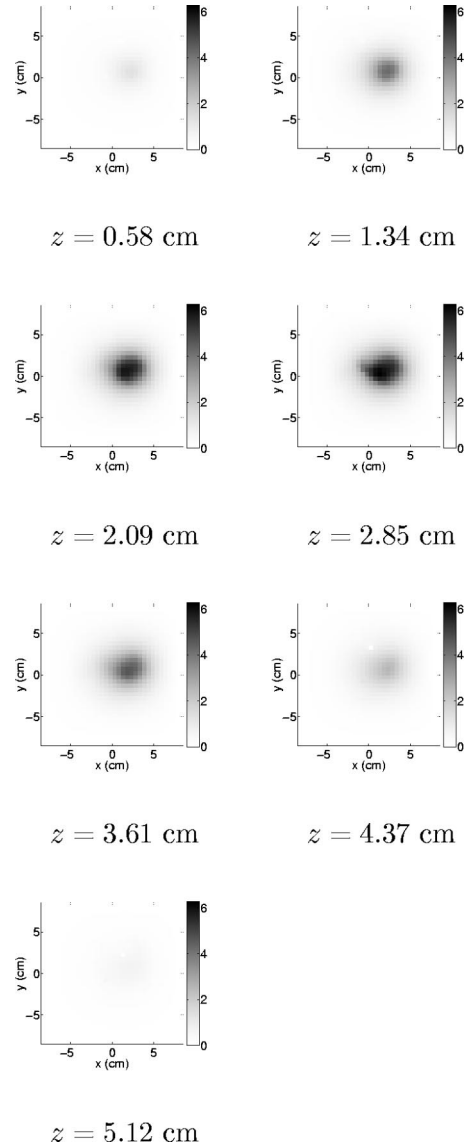


Fig. 7. Reconstruction of τ (in 10^{-10} s), obtained by using 78-, 314-, and 627-MHz data.

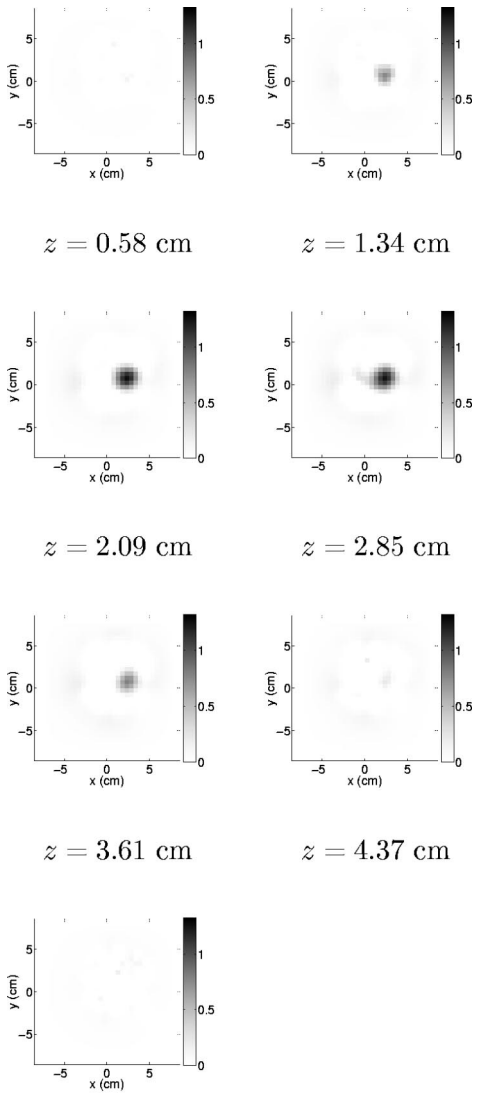


Fig. 8. Reconstruction of η (in $10^{-4}\,\mathrm{cm}^{-1}$), obtained by using only 78-MHz data.

plies a more accurate fit to the data irrespective of image quality. As in our previous work,[21] we define the weighted average lifetime:

$$\hat{\tau}_{\mathrm{avg}} = \frac{\sum_{i=1}^{N} \hat{\eta}(r_i)\hat{\tau}(r_i)}{\sum_{i=1}^{N} \hat{\eta}(r_i)}. \tag{37}$$

This weighted average accounts for the possibility that the reconstruction could be significant in spurious regions where $\hat{\eta} \approx 0$. It is also related to the weighting that occurs in the source term of Eq. (3) and thus represents the effect of τ on the data y_f. Figure 11(f) shows $\hat{\tau}_{\mathrm{avg}}$ as a function of σ. As σ increases, $\hat{\tau}_{\mathrm{avg}}$ asymptotically approaches a constant value of approximately 0.49 ns, just under a literature-reported value of 0.56 ns.[17] The stability of $\hat{\tau}_{\mathrm{avg}}$ with increasing σ suggests that selecting the largest σ that produces a physically acceptable image is reasonable. However, despite the stability of $\hat{\tau}_{\mathrm{avg}}$, the image becomes more distorted (and less physically meaningful) with increasing σ because of the inherent trade-off between regularization and accurate fitting of the data. We selected $\sigma = 0.1$ ns, which gave $\hat{\tau}_{\mathrm{avg}} = 0.205$ ns, as the best compromise with the lifetime distributed evenly throughout the fluorophore. This result is shown in Fig. 11(c).

6. MUTUAL INFORMATION PERFORMANCE MEASURE

A. Motivation and Mathematics

The results of Sections 4 and 5 suggest that the additional modulation frequencies provide useful information for some problems but not for others. Here, we use the mutual information, defined in information theory,[29,30,38] as

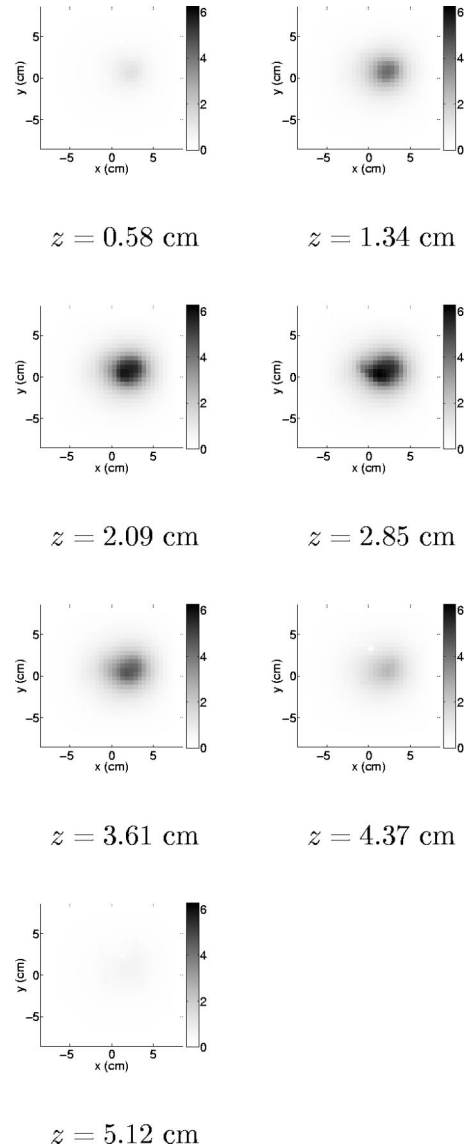
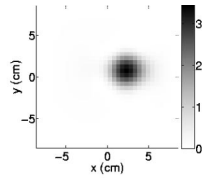
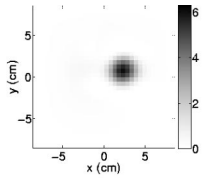


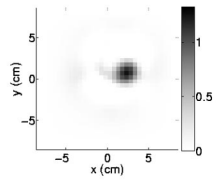
Fig. 9. Reconstruction of τ (in 10^{-10} s), obtained by using only 78-MHz data.



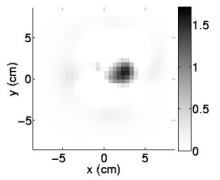
(a) $\sigma = 2.5 \times 10^{-6}$ cm^{-1}



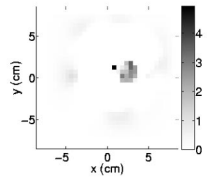
(b) $\sigma = 5.0 \times 10^{-6}$ cm^{-1}



(c) $\sigma = 2.5 \times 10^{-5}$ cm^{-1}



(d) $\sigma = 5.0 \times 10^{-5}$ cm^{-1}



(e) $\sigma = 2.5 \times 10^{-4}$ cm^{-1}

Fig. 10. Reconstructions of η (in 10^{-4} cm^{-1}) using various values of σ, showing a progression from overregularization to underregularization. The $z = 2.85$ cm cross sections are shown. The τ model used $\sigma = 1 \times 10^{-10}$ s in all cases.

an indicator of the reconstruction fidelity that can be achieved with a particular measurement scenario.

We first define the mutual information for our problem. Let X be the random vector corresponding to the fluorescent yield image, and let Y be the random vector corresponding to the measured data. Then the mutual information $I(X; Y)$ is defined as[30]

$$I(X; Y) = H(Y) - H(Y|X), \tag{38}$$

where $H(Y)$ and $H(Y|X)$ are the differential entropy and the conditional differential entropy[30] defined by

$$H(Y) = E[-\log p_Y(Y)], \tag{39}$$

$$H(Y|X) = E[-\log p_{Y|X}(Y|X)], \tag{40}$$

where $p_Y(y)$ and $p_{Y|X}(y|x)$ are the densities of X and $Y|X$, respectively, and $E[\cdot]$ is the expected value. Informally, the mutual information measures the information that the data Y contains about the unknown image X. If base 2 logarithms are used, then $I(X; Y)$ has units of bits, whereas a natural logarithm results in units of "nats." We use the natural logarithm.

For our problem, the mutual information may be easily computed. For simplicity, assume that $\tau = 0$ and consider the information content for reconstructing η. To model the image statistics, we use the Gaussian Markov random field (GMRF) prior model, which is equivalent to the GGMRF with $\rho = 2$. In this case, the density of the image X is given by

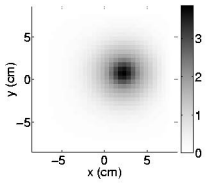
$$p_X(x) = \frac{1}{\sqrt{(2\pi)^N \sigma^2}} |C|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} x^{\mathrm{H}} C x \right), \tag{41}$$

where the matrix C has elements

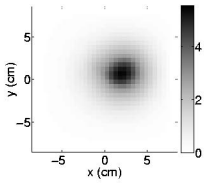
$$C_{i,j} = \begin{cases} 2 & \text{if } i = j \\ 2b_{i-j} & \text{if } i \neq j \end{cases},$$

and b_{i-j} and σ are the same as in Eq. (11). As in previous sections, we let

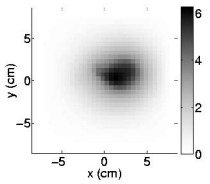
$$E[Y|X] = JX, \tag{42}$$



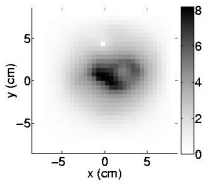
(a) $\sigma = 1 \times 10^{-11}$ s



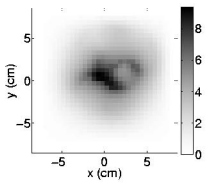
(b) $\sigma = 5 \times 10^{-11}$ s



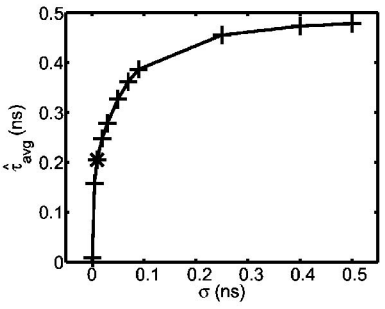
(c) $\sigma = 1 \times 10^{-10}$ s



(d) $\sigma = 3 \times 10^{-10}$ s



(e) $\sigma = 5 \times 10^{-10}$ s



(f)

Fig. 11. (a)–(e) $z = 2.85$ cm cross sections of reconstructed τ (in 10^{-10} s) for various σ, showing a progression from overregularization to underregularization. (f) $\hat{\tau}_{\mathrm{avg}}$ as a function of σ for the τ reconstruction. The $\times$ symbol represents the value of σ that was used in generating the data of Fig. 7. The η model used $\sigma = 2.5 \times 10^{-5}$ cm^{-1} in all cases.

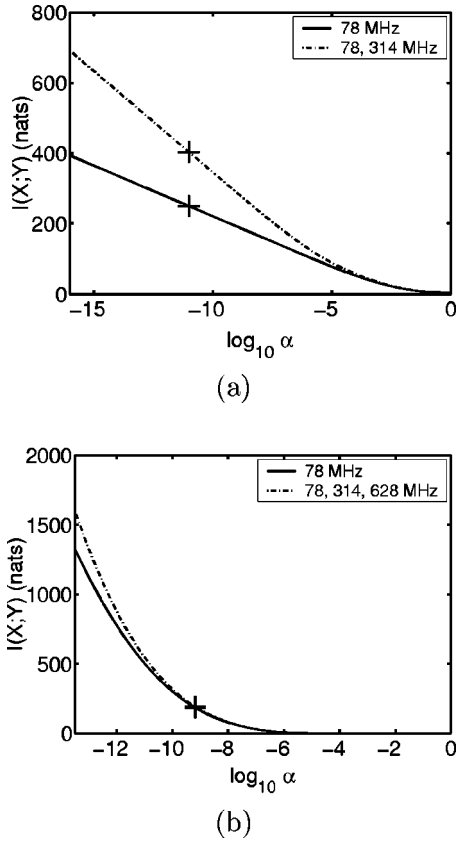


Fig. 12. Mutual information versus α for (a) the simulation model and (b) the experiment model. In (a), the + symbols mark the results for the true value of α used in the simulation. In (b), the + symbol marks the result for the estimated value of α in the experiment. The units of information are nats rather than bits, as the base e logarithm was used.

where J is the matrix representing the linear forward operator, and $\alpha\Lambda^{-1}/2$ is the measurement noise convariance matrix as defined in Eq. (9). Using these definitions, we show in Appendix B that the mutual information is given by

$$I(X;\,Y) = \frac{1}{2}\log\left|I + \frac{2\sigma^2}{\alpha}\Lambda J C^{-1} J^{\mathrm{H}}\right|, \tag{43}$$

where I is the identity matrix.

While Eq. (43) provides a straightforward expression for the mutual information, the question remains as to why one would expect it to be a good predictor of reconstruction quality. In fact, as the mutual information increases, a lower bound on the reconstruction distortion, or MSE, is reduced. The minimum MSE estimate of X given Y is given by the conditional expectation $\hat{X} = E[X|Y]$. Furthermore, for each random vector X, there exists a distortion rate function $D_X(\,\cdot\,)$ that lower bounds the achievable distortion at a specified rate.[38] In communication applications, the rate is the amount of information content per unit time that is sent over a channel and decoded in order to reconstruct the original signal. Here, the rate refers to the information obtained by the measurement device per acquisition for reconstructing the fluorescence image. Importantly, the distortion rate function depends only on X and the choice of the MSE distortion metric.[38] For any particular choice of Y, we then know that[38]

$$E[\|X - \hat{X}\|^2] \geq D_X(I(X;\,Y)). \tag{44}$$

In addition, it is known that all distortion rate functions are convex, and monotone decreasing with $D_X(0) = E[\|X - E[X]\|^2]$.[38] Therefore, as the mutual information increases, this lower bound on MSE decreases. Alternatively, when mutual information is low, then the MSE is necessarily large.

B. Application to Simulation and Experiment Models

Figure 12(a) shows the mutual information as a function of α for the same geometry and statistical models as those assumed in the simulation study of Section 4. Curves are plotted for both the single-frequency (78-MHz) case and the multiple-frequency (78- and 314-MHz) case. The + symbols mark the curve points corresponding to the true α used to generate the synthetic data. The mutual information values differ significantly at these points. The result suggests that the additional (multiple-frequency) data provide significant information for the set of problems modeled by the same image and measurement statistics as those in the simulation.

In Fig. 12(b), mutual information versus α is plotted for the geometry and statistical models used in the reconstructions from experimental data. With the true α for this instrument unknown, the (overlapping) + symbols mark the curve points corresponding to the estimated α. Previously, we have found that the estimation of α yields accurate values.[24] In contrast to Fig. 12(a), the mutual information values are virtually the same. Hence there is no significant information gain from the multiple-frequency data in this case. In this case of a single, relatively large inhomogeneity, additional frequencies, even with a small number of sources and detectors, are not beneficial.

7. CONCLUSIONS

The algorithm described is a general framework for reconstructing fluorescent yield, fluorescent lifetime, absorption, and scattering from frequency-domain data. Simulation results for reconstructing two small objects indicate improvement due to the incorporation of multiple-frequency data. However, similar improvements are not noted in the experiment, which considers a comparatively large and low-resolution spherical target.

The simulation used a relatively modest number of spatial measurements. The limited-data case is of interest because it is more difficult, possibly making it easier to observe the relative advantages due to increased modulation diversity. However, even when the number of spatial measurements used in the experiment was reduced dramatically, the reconstructions of fluorescent yield and lifetime were substantially similar, showing little improvement with additional frequencies. Evidently, the physical properties of the unknown image play an important role in determining the relative improvement due to multiple frequencies. We suppose that the additional, higher-resolution image components observable by the

use of the additional modulation frequencies may be largely absent from the true sphere image. Simulations with single, spherical images seem to confirm that the sphere is fully resolved using just one frequency, in agreement with the experiment. The frequencies considered in the simulation and the experiment were chosen to be typical of those used in practical measurement devices. Reconstructions from simulated and experimental data showed similar behavior over a wide range of frequencies.

The mutual information performance metric provides a theoretical foundation for our observations. A significant increase in the information corresponds with improved reconstruction results in the simulation. Also, a lack of information gain corresponds with lack of improvement in the reconstruction results for the experiment. Hence it appears that information theory can accurately predict any improvement in reconstruction results due to changes in the experimental configuration. If accurate statistical models of the unknown image and measurement devices are known, this property is useful in the design of an experiment. Theoretically, a variety of other important design parameters, such as the source/detector locations and the measurement data types, may also be optimized by application of the mutual information performance metric. In principle, for Gaussian problems, it is possible to use other, related performance metrics such as the reconstruction MSE. However, the reconstruction MSE is quite difficult to compute,[50] making it less computationally tractable than the mutual information.

As in a previous study,[21] we note that the baseline calibration procedure simplifies the experimental reconstruction. In practice, baseline data may be available before administration of a contrast agent, but collecting it still may be problematic. Previous investigations, such as the normalized Born approximation used by Ntziachristos and Weissleder,[20] suggest that the baseline calibration step can be circumvented under certain simplifying assumptions about the absorption problem. Using higher-order transport equations[51] to improve modeling accuracy or estimating source/detector and boundary condition coefficients numerically[26,52–54] may also alleviate the need for baseline calibrations.

APPENDIX A: PSEUDOCODE FOR INVERSION ALGORITHM

```
main {
    1. Form G_x^(s) and G_m^(d)
    2. Repeat until converged: {
       (a) α̂_f ← (1/P_f)‖y_f − f_f(x̂_f, x̂_x, x̂_m)‖²_{Λ_f}
       (b) x̂_f ← ICD_update(x̂_f, α̂_f, G_x^(s), G_m^(d))
       }
    }
x̂ ← ICD_update(x̂, α̂, G^(s), G^(d); x) {
    1. For q = 1,...,Q {
       (a) z_{ω_q} ← y_{ω_q} − J_{ω_q} h_{ω_q}(x̂)
       }
    2. For i = 1,...,N (in random order), {
       (a) x̃_i ← x̂_i
       (b) For q = 1,...,Q {
          i.  Compute [J_{ω_q}]_{*(i)} by taking the ith column in Eq. (29)
          ii. θ_{1,ω_q} ← −2 Re([J_{ω_q}]^H_{*i} Λ_{ω_q} z_{ω_q})
          iii. θ_{2,ω_q} ← 2[J_{ω_q}]^H_{*i} Λ_{ω_q} [J_{ω_q}]_{*i}
          }
       (c) x̂_i ← arg min_{x_i≥0}((1/α̂) Σ_{q=1}^{Q} {θ_{1,ω_q}[h(x, r_i, ω_q)
          − h(x̃, r_i, ω_q)]
          + (θ_{2,ω_q}/2)[h(x, r_i, ω_q) − h(x̃, r_i, ω_q)]²}
          + [1/(ρ_A σ_A^{ρ_A})] Σ_{j∈N_i} b_{i−j} |x_i − x̃_j|^{ρ_A})
       (d) For q = 1,...,Q {
             z_{ω_q} ← z_{ω_q} + [J_{ω_q}]_{*i}[h(x̂, r_i, ω_q)
             − h(x̃, r_i, ω_q)]
          }
       }
    3. For i = N + 1,...,2N (in random order), {
       (a) x̃_i ← x̂_i
       (b) For q = 1,...,Q {
          i.  Compute [J_{ω_q}]_{*(i−N)} by taking the (i − N)th column in Eq. (29)
          ii. θ_{1,ω_q} ← −2 Re([J_{ω_q}]^H_{*(i−N)} Λ_{ω_q} z_{ω_q})
          iii. θ_{2,ω_q} ← 2[J_{ω_q}]^H_{*(i−N)} Λ_{ω_q} [J_{ω_q}]_{*(i−N)}
          }
       (c) x̂_i ←
          arg min_{x_i≥0}((1/α̂) Σ_{q=1}^{Q} {θ_{1,ω_q}[h(x, r_{i−N}, ω_q)
          − h(x̃, r_{i−N}, ω_q)]
          + (θ_{2,ω_q}/2)[h(x, r_{i−N}, ω_q) − h(x̃, r_{i−N}, ω_q)]²}
          + [1/(ρ_B σ_B^{ρ_B})] Σ_{j∈N_i} b_{i−j} |x_i − x̃_j|^{ρ_B})
       (d) For q = 1,...,Q {
             z_{ω_q} ← z_{ω_q} + [J_{ω_q}]_{*(i−N)}[h(x̂, r_{i−N}, ω_q)
             − h(x̃, r_{i−N}, ω_q)]
          }
       }
    }
```

APPENDIX B: MUTUAL INFORMATION DERIVATION

Here, we derive the expression for the mutual information that is given in Eq. (43). The mutual information given by Eq. (38) depends on $H(Y)$ and $H(Y|X)$. We obtain the expressions for $H(Y)$ and $H(Y|X)$ and use them to compute $I(X; Y)$.

Let $p_X(x)$, $p_{Y|X}(y|x)$, σ, α, the linear forward operator J, and the matrices C and Λ be as in Eqs. (8), (11), (41), and (42). Define the shot-noise process $Z = Y - JX$, and assume that Z is independent of X. Then $Y = JX + Z$ is Gaussian, and

$$E[Y] = E[E[Y|X]] = E[JX] = JE[X] = 0. \quad \text{(B1)}$$

In addition,

$$E[YY^{\mathrm{H}}] = E[(JX + Z)(JX + Z)^{\mathrm{H}}] \quad \text{(B2)}$$

$$= E[ZZ^{\mathrm{H}}] + E[(JX)Z^{\mathrm{H}}] + E[Z(JX)^{\mathrm{H}}] + E[JXX^{\mathrm{H}}J^{\mathrm{H}}] \quad \text{(B3)}$$

$$= E[ZZ^{\mathrm{H}}] + JE[XX^{\mathrm{H}}]J^{\mathrm{H}} \quad \text{(B4)}$$

$$= \frac{\alpha}{2}\Lambda^{-1} + \sigma^2 JC^{-1}J^{\mathrm{H}}. \quad \text{(B5)}$$

Hence

$$p_Y(y) = \frac{1}{[(2\pi)^P|\Upsilon|]^{1/2}} \exp\left\{-\frac{1}{2}\|y\|^2_{\Upsilon^{-1}}\right\}, \quad \text{(B6)}$$

where

$$\Upsilon = \frac{\alpha}{2}\Lambda^{-1} + \sigma^2 J C^{-1} J^{\mathrm{H}}. \quad \text{(B7)}$$

With Eqs. (39) and (B6), $H(Y)$ is given by[30]

$$H(Y) = \frac{1}{2}\log[(2\pi)^P] + \frac{P}{2} + \frac{1}{2}\log|\Upsilon|. \quad \text{(B8)}$$

Similarly, using Eq. (40), we obtain

$$H(Y|X) = \frac{1}{2}\log[(2\pi)^P] + \frac{P}{2} + \frac{1}{2}\log\left|\frac{\alpha}{2}\Lambda^{-1}\right|. \quad \text{(B9)}$$

Substituting Eqs. (B8) and (B9) into Eq. (38) yields

$$\begin{aligned} I(X; Y) &= H(Y) - H(Y|X) \\ &= \frac{1}{2}\log\frac{|\Upsilon|}{|\frac{\alpha}{2}\Lambda^{-1}|} \\ &= \frac{1}{2}\log\frac{|\frac{\alpha}{2}\Lambda^{-1} + \sigma^2 J C^{-1} J^{\mathrm{H}}|}{|\frac{\alpha}{2}\Lambda^{-1}|} \\ &= \frac{1}{2}\log\left|I + \frac{2\sigma^2}{\alpha}\Lambda J C^{-1} J^{\mathrm{H}}\right|, \end{aligned} \quad \text{(B10)}$$

where I is the identity matrix, and we have used the determinant identity $|A||B| = |AB|$.

ACKNOWLEDGMENTS

This work was funded by the National Science Foundation under contract CCR-0073357. Jonathan J. Stott and David A. Boas acknowledge support from Advanced Research Technologies Inc.

Send correspondence to Kevin J. Webb, School of Electrical and Computer Engineering, Purdue University, 465 Northwestern Avenue, West Lafayette, Indiana 47907-2035. E-mail, webb@purdue.edu; phone, 765-494-3373; fax, 765-494-2706.

REFERENCES

1. S. R. Arridge, "Optical tomography in medical imaging," Inverse Probl. **15**, R41–R93 (1999).
2. D. A. Boas, D. H. Brooks, E. L. Miller, C. A. DiMarzio, M. Kilmer, R. J. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," IEEE Signal Process. Mag. **18**, 57–75 (2001).
3. A. Pèlegrin, S. Folli, F. Buchegger, J. Mach, G. Wagnières, and H. van den Bergh, "Antibody-fluorescein conjugates for photoimmunodiagnosis of human colon carcinoma in nude mice," Cancer **67**, 2529–2537 (1991).
4. B. Ballou, G. W. Fisher, T. R. Hakala, and D. L. Farkas, "Tumor detection and visualization using cyanine fluorochrome-labeled antibodies," Biotechnol. Prog. **13**, 649–658 (1997).
5. R. Cubeddu, G. Canti, A. Pifferi, P. Taroni, and G. Valentini, "Fluorescence lifetime imaging of experimental tumors in hematoporphyrin derivative-sensitized mice," Photochem. Photobiol. **66**, 229–236 (1997).
6. J. A. Reddy and P. S. Low, "Folate-mediated targeting of therapeutic and imaging agents to cancers," Crit. Rev. Ther. Drug Carrier Syst. **15**, 587–627 (1998).
7. J. S. Reynolds, T. L. Troy, R. H. Mayer, A. B. Thompson, D. J. Waters, K. K. Cornell, P. W. Snyder, and E. M. Sevick-Muraca, "Imaging of spontaneous canine mammary tumors using fluorescent contrast agents," Photochem. Photobiol. **70**, 87–94 (1999).
8. U. Mahmood, C. Tung, J. A. Bogdanov, and R. Weissleder, "Near-infrared optical imaging of protease activity for tumor detection," Radiology **213**, 866–870 (1999).
9. K. Licha, B. Riefke, V. Ntziachristos, A. Becker, B. Chance, and W. Semmler, "Hydrophilic cyanine dyes as contrast agents for near-infrared tumor imaging: synthesis, photophysical properties and spectroscopic *in vivo* characterization," Photochem. Photobiol. **72**, 392–398 (2000).
10. V. Ntziachristos, A. G. Yodh, M. Schnall, and B. Chance, "Concurrent MRI and diffuse optical tomography of breast after indocyanine green enhancement," Proc. Natl. Acad. Sci. USA **97**, 2767–2772 (2000).
11. A. Becker, C. Hessenius, K. Licha, B. Ebert, U. Sukowski, W. Semmler, B. Wiedenmann, and C. Grotzinger, "Receptor-targeted optical imaging of tumors with near-infrared fluorescent ligands," Nat. Biotechnol. **19**, 327–331 (2001).
12. J. E. Bugaj, S. Achilefu, R. B. Dorshow, and R. Rajagopalan, "Novel fluorescent contrast agents for optical imaging of *in vivo* tumors based on a receptor-targeted dye-peptide conjugate platform," J. Biomed. Opt. **6**, 122–133 (2001).
13. M. S. Patterson and B. W. Pogue, "Mathematical model for time-resolved and frequency-domain fluorescence spectroscopy in biological tissues," Appl. Opt. **33**, 1963–1974 (1994).
14. E. M. Sevick-Muraca, G. Lopez, J. S. Reynolds, T. L. Troy, and C. L. Hutchinson, "Fluorescence and absorption contrast mechanisms for biomedical optical imaging using frequency-domain techniques," Photochem. Photobiol. **66**, 55–64 (1997).
15. R. Roy and E. M. Sevick-Muraca, "Three-dimensional unconstrained and constrained image-reconstruction techniques applied to fluorescence, frequency-domain photon migration," Appl. Opt. **40**, 2206–2215 (2001).
16. M. A. O'Leary, D. A. Boas, X. D. Li, B. Chance, and A. G. Yodh, "Fluorescence lifetime imaging in turbid media," Opt. Lett. **21**, 158–160 (1996).
17. D. Paithankar, A. Chen, B. Pogue, M. Patterson, and E. Sevick-Muraca, "Imaging of fluorescent yield and lifetime from multiply scattered light reemitted from random media," Appl. Opt. **36**, 2260–2272 (1997).
18. H. Jiang, "Frequency-domain fluorescent diffusion tomography: a finite-element-based algorithm and simulations," Appl. Opt. **37**, 5337–5343 (1998).
19. J. Chang, H. L. Graber, and R. L. Barbour, "Luminescence optical tomography of dense scattering media," J. Opt. Soc. Am. A **14**, 288–299 (1997).
20. V. Ntziachristos and R. Weissleder, "Experimental three-dimensional fluorescence reconstruction of diffuse media by use of a normalized Born approximation," Opt. Lett. **26**, 893–895 (2001).
21. A. B. Milstein, S. Oh, K. J. Webb, C. A. Bouman, Q. Zhang, D. A. Boas, and R. P. Millane, "Fluorescence optical diffusion tomography," Appl. Opt. **42**, 3081–3094 (2003).
22. E. Shives, Y. Xu, and H. Jiang, "Fluorescence lifetime tomography of turbid media based on an oxygen-sensitive dye," Opt. Express **10**, 1557–1562 (2002).
23. J. C. Ye, K. J. Webb, C. A. Bouman, and R. P. Millane, "Optical diffusion tomography using iterative coordinate descent optimization in a Bayesian framework," J. Opt. Soc. Am. A **16**, 2400–2412 (1999).
24. J. C. Ye, C. A. Bouman, K. J. Webb, and R. P. Millane, "Nonlinear multigrid algorithms for Bayesian optical diffusion tomography," IEEE Trans. Image Process. **10**, 909–922 (2001).
25. A. B. Milstein, S. Oh, J. S. Reynolds, K. J. Webb, C. A. Bouman, and R. P. Millane, "Three-dimensional Bayesian opti-

cal diffusion tomography with experimental data," Opt. Lett. **27**, 95–97 (2002).
26. S. Oh, A. B. Milstein, R. P. Millane, C. A. Bouman, and K. J. Webb, "Source–detector calibration in three-dimensional Bayesian optical diffusion tomography," J. Opt. Soc. Am. A **19**, 1983–1993 (2002).
27. F. Gao, H. Zhao, and Y. Yamada, "Improvement of image quality in diffuse optical tomography by use of full time-resolved data," Appl. Opt. **41**, 778–791 (2002).
28. V. Ntziachristos, J. Culver, M. Holboke, A. G. Yodh, and B. Chance, "Optimal selection of frequencies for diffuse optical tomography," in *Biomedical Topical Meetings*, Vol. 38 of OSA Trends in Optics and Photonics Series (Optical Society of America, Washington, D.C., 2000), pp. 475–477.
29. C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948).
30. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
31. R. F. Wagner, D. G. Brown, and M. S. Pastel, "Application of information theory to the assessment of computed tomography," Med. Phys. **6**, 83–94 (1979).
32. M. Fuderer, "The information content of MR images," IEEE Trans. Med. Imaging **7**, 368–380 (1988).
33. L. Shao, A. O. Hero, W. L. Rogers, and N. H. Clinthorne, "The mutual information criterion for SPECT aperture evaluation and design," IEEE Trans. Med. Imaging **8**, 322–336 (1989).
34. A. O. Hero and L. Shao, "Information analysis of single photon emission computed tomography with count losses," IEEE Trans. Med. Imaging **9**, 117–127 (1990).
35. J. P. Culver, V. Ntziachristos, M. J. Holbrooke, and A. G. Yodh, "Optimization of optode arrangements for diffuse optical tomography: a singular-value analysis," Opt. Lett. **26**, 701–703 (2001).
36. H. Xu, H. Dehghani, B. W. Pogue, R. Springett, K. D. Paulsen, and J. F. Dunn, "Near infrared imaging in the small animal brain: optimization of fiber positions," J. Biomed. Opt. **8**, 102–110 (2003).
37. J. Stott and D. A. Boas, "A practical comparison between time-domain and frequency-domain diffusive optical imaging systems," in *Biomedical Topical Meetings*, Vol. 71 of OSA Trends in Optics and Photonics Series (Optical Society of America, Washington, D.C., 2002), pp. 626–628.
38. T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression* (Prentice Hall, Englewood Cliffs, N.J., 1971).
39. S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).
40. J. J. Duderstadt and L. J. Hamilton, *Nuclear Reactor Analysis* (Wiley, New York, 1976).
41. A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. 1.
42. J. S. Reynolds, C. A. Thompson, K. J. Webb, F. P. LaPlant, and D. Ben-Amotz, "Frequency domain modeling of reradiation in highly scattering media," Appl. Opt. **36**, 2252–2259 (1997).
43. C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," IEEE Trans. Image Process. **2**, 296–310 (1993).
44. K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," IEEE Trans. Signal Process. **41**, 534–548 (1993).
45. E. K. P. Chong and S. H. Zak, *An Introduction to Optimization* (Wiley, New York, 1996).
46. R. C. Haskell, L. O. Svaasand, T.-T. Tsay, T.-C. Feng, M. S. McAdams, and B. J. Tromberg, "Boundary conditions for the diffusion equation in radiative transfer," J. Opt. Soc. Am. A **11**, 2727–2741 (1994).
47. S. S. Saquib, C. A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields with applications to Bayesian tomography," IEEE Trans. Image Process. **7**, 1029–1044 (1998).
48. M. L. J. Landsman, G. Kwant, G. A. Mook, and W. G. Zijlstra, "Light-absorbing properties, stability, and spectral stabilization of indocyanine green," J. Appl. Physiol. **40**, 575–583 (1976).
49. R. C. Benson and H. A. Kues, "Fluorescence properties of indocyanine green as related to angiography," Phys. Med. Biol. **23**, 159–163 (1978).
50. M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging* (Institute of Physics, Philadelphia, Pa., 1998).
51. A. D. Klose and A. H. Hielscher, "Optical tomography using the time-independent equation of radiative transfer—Part 2: inverse model," J. Quant. Spectrosc. Radiat. Transf. **72**, 715–732 (2002).
52. D. Boas, T. Gaudette, and S. Arridge, "Simultaneous imaging and optode calibration with diffuse optical tomography," Opt. Express **8**, 263–270 (2001).
53. J. J. Stott, J. P. Culver, S. R. Arridge, and D. A. Boas, "Optode positional calibration in diffuse optical tomography," Appl. Opt. **42**, 3154–3162 (2003).
54. N. Iftimia and H. Jiang, "Quantitative optical image reconstructions of turbid media by use of direct-current measurements," Appl. Opt. **39**, 5256–5261 (2000).