IMAGING OF NEAR-INFRARED FLUORESCENCE, ABSORPTION, AND SCATTER IN TURBID MEDIA

A Thesis

Submitted to the Faculty

of

Purdue University

by

Adam B. Milstein

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2004

To My Parents

ACKNOWLEDGMENTS

First and foremost, I thank my parents Norman and Elaine, my sister Lisa, and my brother Robert for their unconditional love and support throughout my life.

I am fortunate enough to have not one, but two exceptional advisers, Prof. Charles A. Bouman and Prof. Kevin J. Webb. I thank them for their unwavering guidance, patience, career advice, and kindness.

I thank my other committee members, Prof. Philip S. Low and Prof. Peter C. Doerschuk, for dedicating their time and energy. In particular, I thank Prof. Low for the fruitful collaborative research efforts which have given me essential exposure to real-life medical applications. I am also grateful to Prof. Rick P. Millane for his indispensable advice and commitment to my success. I must also thank Prof. David A. Boas for his kindness and encouragement in my work with his group at Massachusetts General Hospital.

I thank Seungseok Oh, whose tireless efforts and generosity truly made much of what appears in this report possible. In addition, I thank Dr. Michael D. Kennedy, Dr. Quan Zhang, Dr. Jonathan J. Stott, and Dr. Jeffery S. Reynolds for their crucial assistance, contributions, and scientific expertise. I also thank Prof. Evan D. Morris, Dr. Mark A. Webster, and Mustafa Kamasak for many helpful discussions.

I thank Dr. Richard D. Juday for his invaluable advice, guidance, and friendship, which helped build the foundation for my research career.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Figure Page

Figure Page

ABSTRACT

Milstein, Adam B. Ph.D., Purdue University, August, 2004. Imaging of near-infrared fluorescence, absorption, and scatter in turbid media. Major Professor: Charles A. Bouman and Kevin J. Webb.

A nonlinear, Bayesian optimization scheme is presented for reconstructing fluorescent yield and lifetime, the absorption coefficient, and the diffusion coefficient in turbid media, such as biological tissue. The method utilizes measurements at both the excitation and emission wavelengths for reconstructing all unknown parameters. The effectiveness of the reconstruction algorithm is demonstrated by simulation and by application to experimental data from a tissue phantom containing the fluorescent agent indocyanine green.

1. INTRODUCTION

Optical diffusion tomography (ODT) is emerging as a powerful tissue imaging modality [1, 2]. In ODT, images are comprised of the spatially dependent absorption and scattering properties of the tissue. Boundary measurements from several sources and detectors are used to recover the unknown parameters from a scattering model described by a partial differential equation. Contrast between the properties of diseased and healthy tissue might then be used in clinical diagnosis. In principle, sinusoidally modulated, continuous-wave (CW), or pulsed excitation light is launched into the biological tissue, where it undergoes multiple scattering and absorption before exiting. The measured intensity and phase (or delay) information may be used to reconstruct three-dimensional (3-D) maps of the absorption and scattering properties by optimizing a fit to diffusion model computations. As a result of the nonlinear dependence of the diffusion equation photon flux on the unknown parameters and the inherently 3-D nature of photon scattering, this inverse problem is computationally intensive and must be solved iteratively.

A relatively modest intrinsic contrast between the optical parameters of diseased and healthy breast tissue has been reported in some studies [3, 4]. The use of exogenous fluorescent agents has the potential to improve the contrast and thus to facilitate early diagnosis. In recent years, the use of fluorescent indicators as exogenous contrast agents for *in vivo* imaging of tumors with near-infrared (NIR) or visible light has shown great promise, attracting considerable interest [5–14]. In experimental studies with animal subjects [5–7, 9, 10, 13, 14], fluorescence has been successfully used to visualize cancerous tissue *in vivo* near the skin surface. In addition, Ntziachristos *et al.* [12] have used optical diffusion tomography after indocyanine green (ICG) injection to image the absorption of a malignant breast tumor in a human sub-

ject. The injected fluorophore may preferentially accumulate in diseased tissue due to increased blood flow from tumor neovascularization [9]. Alternatively, the agent may have different decay properties in diseased tissue, which could be useful in localizing tumors independently of fluorophore concentration [7]. In addition, contrast between tumors and surrounding tissue may be substantially improved by the use of diagnostic agents that selectively target receptors specific to cancer cells [8,10,13,14].

In frequency-domain fluorescence optical diffusion tomography, sinusoidally modulated light at the fluorophore's excitation wavelength is launched into the tissue. The excited fluorophore, when it decays to the ground state, emits light at a longer (emission) wavelength, and this emission is measured by an array of detection devices. These emission data are then used to perform a volumetric reconstruction of the yield (a measure of the fluorescence efficiency) and the lifetime (the fluorescent decay parameter). However, the multiple scattering in tissue complicates the reconstruction [15, 16]. The emission intensity of the fluorophore is proportional to the optical intensity at the excitation wavelength at that position, which depends, in turn, on the optical parameters of the scattering domain at the excitation wavelength. A rigorous reconstruction of fluorescence property maps should also therefore include reconstructions of absorption and scattering parameters at the excitation and emission wavelengths. In addition, reconstruction of the unknown absorption and scattering coefficients by use of ODT can function as an adjunct image to the fluorescence image in screening for tumors.

Fluorescence imaging simulations with (3-D) [17] and two-dimensional (2-D) [18–20] geometries have reconstructed fluorescence yield and lifetime parameters. These simulations have generally assumed that the absorption and scattering parameters are known in advance, except for Roy and Sevick-Muraca [17], who also reconstructed the excitation wavelength absorption. In an early experimental result, Chang *et al.* [21] used a transport theory model to reconstruct fluorescent yield in a heterogeneous tissue phantom containing Rhodamine 6G. Their study used CW data recorded in a 2-D plane geometry. Recently, Ntziachristos and Weissleder [22] used a normalized

Born approximation to reconstruct 3-D, fluorescent heterogeneities containing the NIR cyanine dye Cy5.5 embedded in a tissue phantom. Under the assumption of known background optical properties and absorbers limited to a perturbative regime, their technique can circumvent the need for recording background measurements before contrast agent administration.

The development of nonlinear inversion methods for optical diffusion tomography is necessary due to the fundamentally limited accuracy of methods which linearize the forward model [23]. Previously, Ye *et al.* have presented a nonlinear Bayesian approach [24,25] and shown that it produces high quality images compared to previous methods such as the distorted Born iterative method [26]. The method formulates the inversion as the optimization of an objective function which incorporates a model of the detection system and *a priori* knowledge about the image properties. We have found that a neighborhood regularization scheme used in a Bayesian framework reduces artifacts characteristic of previous approaches which impose a penalty on the norm of the image updates [24]. The inversion can be made more computationally efficient by multigrid techniques [25].

We have extended Ye's previous 2-D Bayesian formulation and iterative coordinate descent (ICD) optimization method for absorption imaging to 3-D, and we address the problem of estimating source-detector coupling, the background diffusion coefficient, and detector noise variance, thereby circumventing difficult and inconvenient calibration measurements on homogeneous phantoms [27–38]. The estimation of source-detector coupling loss and background parameters using a preprocessing technique with an assumed homogeneous domain has been described, [39] and the source-detector coupling coefficients have been estimated as part of a linear generalized inverse. [40] We incorporate estimation of the ancillary parameters into the Bayesian framework and update the estimates throughout the inversion procedure. Laboratory data are used to assess the algorithm and the merits of using multiple modulation frequencies.

We have also extended the approach to include fluorescence yield and lifetime in the inverse problem [41, 42]. We present a new inversion algorithm and a measurement scheme for reconstructing all the unknown fluorescence, absorption, and diffusion parameters. Numerical simulations validate the scheme and demonstrate its computational efficacy. We use the method to image a spherical heterogeneity in a tissue phantom by use of transmission data collected by a CW imaging device. The heterogeneity contains ICG, a fluorescent diagnostic agent approved by the FDA for use in the NIR range, where biomedical imaging with light is most practical. In subsequent work we have extended the FODT approach to include direct reconstruction of pharmacokinetic parameters from measurements of time-varying fluorescence [43].

In addition, we have performed an experiment to localize a fluorescing mouse tumor submerged under varying thicknesses of lipid suspension [44, 45]. The depth of the tumor was recovered after a one dimensional measurement scan with few measurements. The experiment suggests the possibility that clinically useful tumor localization can be obtained from a relatively simple measurement, even with very limited data.

2. THREE-DIMENSIONAL BAYESIAN OPTICAL DIFFUSION IMAGING USING EXPERIMENTAL DATA

2.1 Measurement Models

The frequency domain diffusion equation [46] models the propagation of modulated light in a highly scattering medium. The complex modulation envelope of the photon flux $\phi_k(r)$ at position r due to a point source at position s_k satisfies

$$\nabla \cdot [D(r)\nabla\phi_k(r)] + (-\mu_a(r) - j\omega/c)\phi_k(r) = -\delta(r - s_k), \tag{2.1}$$

where c is the speed of light in the medium, ω is the modulation frequency, $D(r)$ is the diffusion coefficient, and $\mu_a(r)$ is the absorption coefficient. Extrapolated Dirchlet boundary conditions, accounting for refractive index mismatch, may be imposed to model absorbing boundaries. [47]

We consider the case of a spatially variable absorption coefficient and a constant (but unknown) diffusion coefficient. The set of absorption coefficients and the diffusion coefficient are denoted by the vector $\mathbf{x}$, where $x = [\mu_a(r_1), \cdots, \mu_a(r_N), D]^T$, and the domain of the scattering region is discretized into N points at positions r_i. The forward model is expressed as a complex vector $f(x) = [f_{\omega_1}(x), f_{\omega_2}(x), \cdots, f_{\omega_Q}(x)]^T$ where $f_{\omega_i}(x)$ is the computed data vector for $\omega = \omega_i$ corresponding to all source and detector pairs. The true measurements corresponding to f(x) are collected together into the column vector y of length $P = KMQ$, where there are K source positions, M detector positions, and Q modulation frequencies. The objective of the optical imaging problem is to determine the image x from the measurements y.

2.2 Inverse Problem

Formulating the solution in a Bayesian framework [24], we compute the maximum *a posteriori* (MAP) estimate of x (the image) and simultaneously maximize with respect to α (a parameter which scales detector noise variance) and γ (the source-detector coupling coefficient). More precisely, the MAP estimate is given by

$$\hat{x} \;=\; \arg\max_{x\geq 0}\max_{\alpha}\max_{\gamma}\{\;\log p(y|x,\alpha,\gamma)+\log p(x)\;\}, \tag{2.2}$$

where $p(y|x,\alpha,\gamma)$ is the data likelihood and $p(x)$ is the prior density for the image. The data likelihood is formed using a Gaussian model. [24] For the prior density, we use the generalized Gaussian Markov random field (GGMRF) model [24]

$$p(x)=\frac{1}{\sigma^N z(p)}\exp\left[-\frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i-x_j|^p\right] \qquad 1\leq p\leq 2, \tag{2.3}$$

where σ and p are hyperparameters (with $p=2$ corresponding to the Gaussian case), $\mathcal{N}$ consists of all pairs of neighboring nodes, and b_{i-j} represents coefficients in a 26 node neighborhood system with values inversely proportional to node separation.

We assume here that γ is real and is the same for all source-detector pairs, which is appropriate for our experimental arrangement. With α and γ unknown, and using the data and prior density functions, (2.2) can be written

$$\hat{x}=\arg\min_{x\geq 0}\min_{\gamma}\min_{\alpha}\left\{\frac{1}{\alpha}||y-\gamma f(x)||^2_\Lambda + P\log\alpha+\frac{1}{p\sigma^p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i-x_j|^p\right\}, \tag{2.4}$$

where Λ is the inverse of a diagonal covariance matrix and $||w||^2_\Lambda = w^H\Lambda w$. Viewing the argument in (2.4) as a cost function, we sequentially update α, γ, and x in an iterative optimization scheme. Minimization with respect to α, assuming γ and x are constant, gives $\hat{\alpha}=\frac{1}{P}||y-\hat{\gamma}f(\hat{x})||^2_\Lambda$. Minimizing with respect to γ, for constant α and x, gives

$$\hat{\gamma}=\frac{Re\left\{f^H(\hat{x})\Lambda y\right\}}{f^H(\hat{x})\Lambda f(\hat{x})}, \tag{2.5}$$

where H denotes the conjugate transpose. The update with respect to D consists of one iteration of a 1-D Newton's method, with step

$$\Delta\hat{D} \;=\; \frac{Re\left\{E^H f_D'^H(\hat{x})\Lambda z\right\}}{\hat{\gamma}E^H f_D'^H(\hat{x})\Lambda f_D'(\hat{x})E}, \tag{2.6}$$

where $E = [1, 1, ..., 1]^T$, $z = y - \hat{\gamma}f(\hat{x})$, and $f_D'(\hat{x})$ is the $P \times N$ Fréchet derivative [26] of $f(\cdot)$ at $\hat{x}$ with respect to D. The update of the μ_a components of x employs the iterative coordinate descent (ICD) algorithm. [24]

2.3 Experiment

Measurements were made on an optically clear culture flask containing a black plastic cylinder of diameter 0.7 cm embedded in a turbid suspension, as shown in Figure 2.1 [27]. The region of the flask containing the suspension had dimensions 8.1 cm × 2.9 cm × 8.1 cm. The suspension was a phosphate-buffered saline solution of Intralipid diluted to a concentration of 0.4%. The data were collected using an inexpensive apparatus (depicted schematically in Figure 2.1) which contains an infrared LED operating at 890 nm and a silicon p-i-n photodiode. [46] The source was placed at a fixed central position on one side of the flask. On the other side, the detector was mounted on a translation stage and moved to 25 locations at intervals of 0.2 cm (± 2.5 cm). Magnitude and phase data were collected in the range of 10 MHz to 81 MHz using an RF network analyzer.

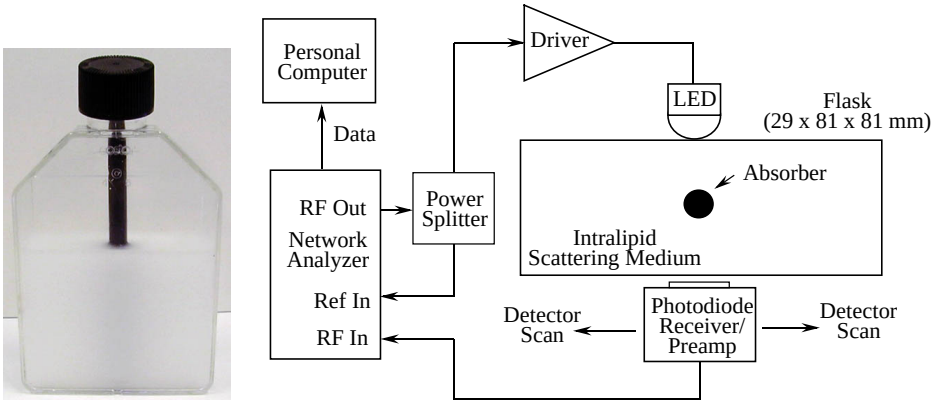
In order to investigate the possible benefits of using multiple modulation frequencies, we selected data acquired at 10 MHz, 46 MHz, and 81 MHz. At each frequency, the 25 measurements were increased to 50 by using the symmetry of the problem to assume that the same data would result from the detectors and source switching sides. Inversions were performed using individual frequencies and also using the three frequencies simultaneously. In all inversions, the domain (including the extrapolated boundary region computed for a refractive index of 1.33) was discretized into 65 × 33 × 65 nodes, giving grid spacings of 1.4 × 1.1 × 1.4 mm. The values of μ_a, D, and γ were initialized to 10^{-5} cm^{-1}, 0.05 cm, and 1.0, respectively. For

the prior density, we used $\sigma = 2.0$ cm^{-1} and $p = 2.0$. The optimization procedure described above was run for 100 iterations, where one iteration consists of an update of all unknowns.

Reconstructed images of the absorption coefficient are shown for the measurement ($z = 0$) plane in Figure 2.2. Figures 2.2(a)-(c) show single-frequency reconstructions obtained using modulation frequencies of 10 MHz, 46 MHz, and 81 MHz, respectively. Figure 2.2(d) shows the reconstruction obtained using all three modulation frequencies. All reconstructions show a circular, centrally located absorber of approximately the correct dimensions. Since the diffusion approximation does not apply in regions of high absorption, the actual values of the absorption within the cylinder cannot be quantitatively interpreted. Figure 2.3 shows isosurface plots of the absorption contoured at 1/4 of the maximum value, corresponding to the reconstructions in Figure 2.2. Despite the fact that data were collected only in a single plane and no quasi-2-D assumptions were incorporated into the inversion geometry or the prior model, the isosurfaces all resemble the cylindrical form of the absorber. The reconstruction using all three modulation frequencies was slightly more accurate, suggesting that the use of multiple frequencies may offer slight advantages. All of the reconstructions contain some artifacts in the vicinity of the detectors, possibly due to the slight asymmetry in the measured data, the failure of the diffusion approximation in the absorbing cylinder, or the influence of the Green's function singularity when evaluating the Fréchet derivative near the detectors. The estimate of D was close to 0.08 cm, which is consistent with previous measurements [46], except for the 10 MHz case, where it was somewhat lower. Using a single value of D (i.e., both outside and inside the absorbing cylinder) is not strictly correct and could introduce some error into the estimate.

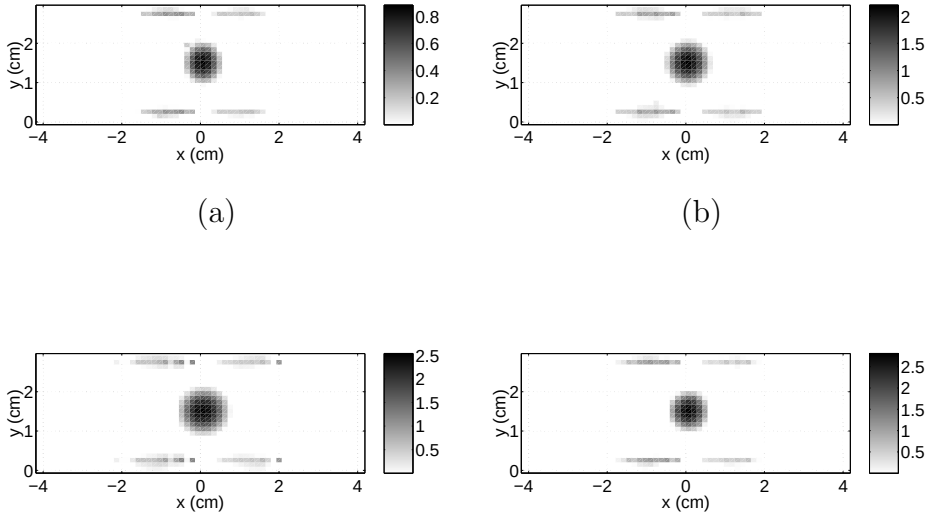
2.4 Conclusion

In summary, we have presented a 3-D Bayesian inversion technique for optical diffusion absorption imaging and applied it to laboratory data. Estimation of source-detector coupling, background diffusion coefficient, and detector noise allow a fully automated reconstruction procedure. The results are geometrically accurate and show that the use of multiple modulation frequencies may be useful. They also demonstrate that use of a full 3-D model with Bayesian regularization allows recovery of 3-D images from limited two-dimensional measurements. Accurate and efficient 3-D inversion methods such as this will be essential for practical optical diffusion imaging.



(a) (b)

Fig. 2.1. (a) Culture flask with the absorbing cylinder embedded in the scattering Intralipid solution. (b) Apparatus used to collect the data.



(a) (b)

(c) (d)

Fig. 2.2. Reconstructed images of the absorption coefficient using data at modulation frequencies of (a) 10 MHz (b) 46 MHz (c) 81 MHz (d) 10, 46, and 81 MHz.

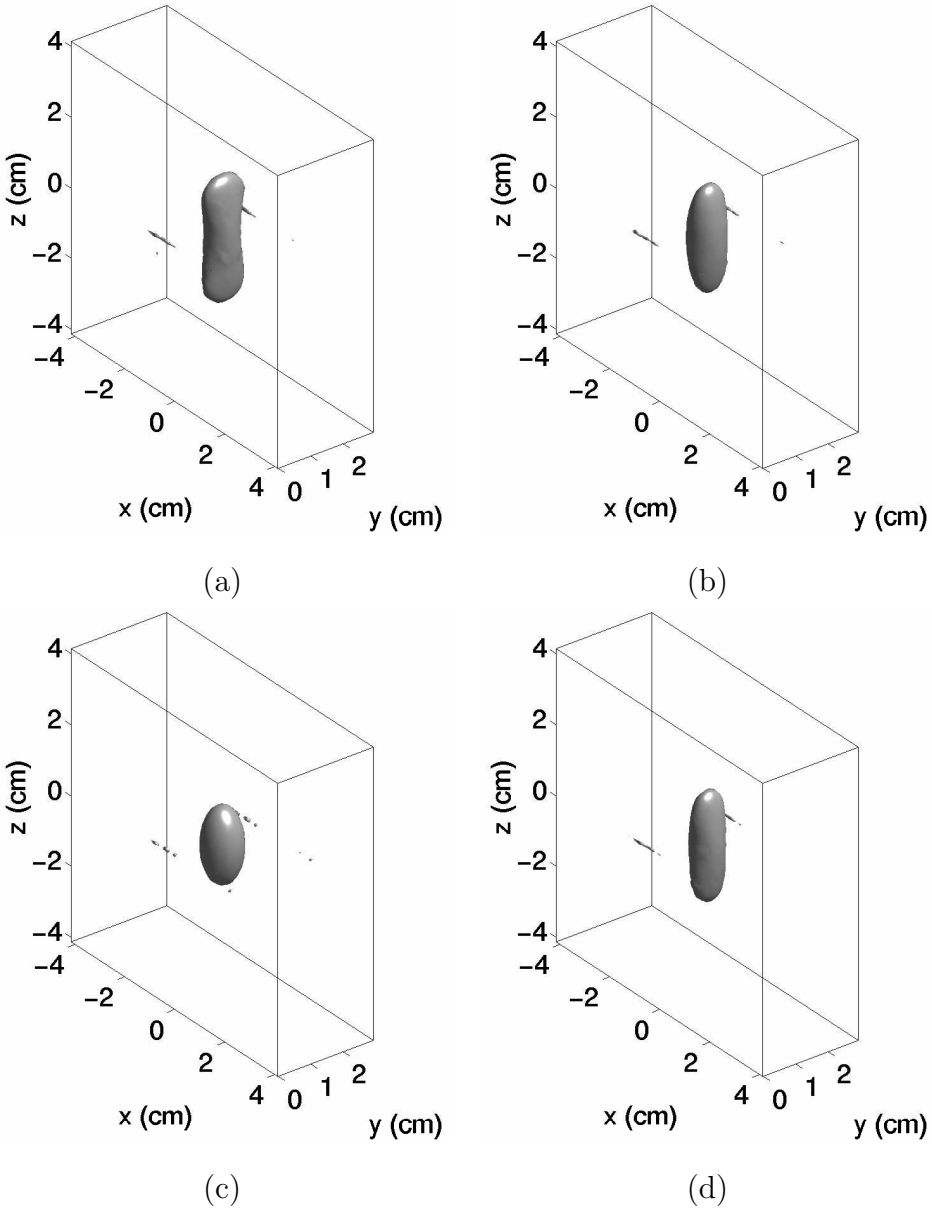


Fig. 2.3. Isosurfaces of the absorption coefficient images contoured at 1/4 of the maximum value using data at: (a) 10 MHz (b) 46 MHz (c) 81 MHz (d) 10, 46, and 81 MHz.

3. FLUORESCENCE OPTICAL DIFFUSION TOMOGRAPHY

3.1 Models

The transport of modulated light (at modulation angular frequency ω, i.e., $e^{j\omega t}$ variation) in a fluorescent, highly scattering medium with an external source at the excitation wavelength is modeled by using the coupled diffusion equations [15,16,48]:

$$\nabla \cdot [D_x(r)\nabla\phi_x(r,\omega)] - [\mu_{a_x}(r) + j\omega/c]\,\phi_x(r,\omega) = -\delta(r - r_{s_k}) \tag{3.1}$$

$$\nabla \cdot [D_m(r)\nabla\phi_m(r,\omega)] - [\mu_{a_m}(r) + j\omega/c]\,\phi_m(r,\omega) = -\phi_x(r,\omega)\eta\mu_{a_f}(r) \times \frac{1 - j\omega\tau(r)}{1 + [\omega\tau(r)]^2}, \tag{3.2}$$

where the subscripts x and m, respectively, denote excitation and emission wavelengths λ_x and λ_m, $\phi(r,\omega)$ is the complex modulation envelope of the photon flux, $\delta(\ \mathrm{r})$ is the Dirac function, and r_{s_k} is the location of the excitation point source. We also assume single exponential decay in this model. The optical parameters are the diffusion coefficients $D(r)$ and the absorption coefficients $\mu_a(r)$. The fluorescence parameters are the lifetime $\tau(r)$ and the fluorescent yield $\eta\mu_{a_f}(r)$. The fluorescent yield 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). Note the right hand side of (3.2), where the light absorbed by fluorophores and subsequently emitted at the emission wavelength, is incorporated into an effective source term. In the case of an external point source at the emission wavelength, the flux is governed by

$$\nabla \cdot [D_m(r)\nabla\phi_m(r,\omega)] - [\mu_{a_m}(r) + j\omega/c]\,\phi_m(r,\omega) = -\delta(r - r_{s_k}). \tag{3.3}$$

In the most general case, the unknown parameters in (3.1) and (3.2) are μ_{a_x}, μ_{a_m}, D_x, D_m, τ, and $\eta\mu_{a_f}$. Reconstructions of the D_x and μ_{a_x} images may be obtained using data from sources and detectors at the excitation wavelength λ_x. Similarly, D_m and μ_{a_m} may be obtained using data from sources and detectors at the emission wavelength λ_m. Finally, having found these parameters, using sources at λ_x and detectors filtered at λ_m will yield the fluorescence parameters. Figure 3.1 depicts this measurement approach schematically.

After discretizing the domain into N voxels of equal size, the unknown parameters can be regarded as three image vectors, each corresponding to a measurement set. Let r_i denote the position of the i^{th} voxel centroid, i.e., the location of a node in a Cartesian finite difference representation of (3.1)-(3.3). We define the image vectors as

$$\begin{aligned} x_x &= \begin{bmatrix} x_{xa} \\ x_{xb} \end{bmatrix} \\ &= [\mu_{a_x}(r_1) \cdots \mu_{a_x}(r_N), D_x(r_1) \cdots D_x(r_N)]^T \qquad (3.4) \\ x_m &= \begin{bmatrix} x_{ma} \\ x_{mb} \end{bmatrix} \\ &= [\mu_{a_m}(r_1) \cdots \mu_{a_m}(r_N), D_m(r_1) \cdots D_m(r_N)]^T \qquad (3.5) \\ x_f &= \begin{bmatrix} x_{fa} \\ x_{fb} \end{bmatrix} \\ &= [\gamma(r_1) \cdots \gamma(r_N), \tau(r_1) \cdots \tau(r_N)]^T, \qquad (3.6) \end{aligned}$$

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. In addition, we reparameterize the fluorescence unknowns $\{\eta\mu_{a_f}, \tau\}$ to $\{\gamma, \tau\}$ using

$$\gamma(r, \omega) = \eta\mu_{a_f}(r) \frac{1}{1 + [\omega\tau(r)]^2}, \qquad (3.7)$$

which, when substituted into (3.2), gives

$$\nabla \cdot [D_m(r)\nabla\phi_m(r,\omega)] - [\mu_{a_m}(r) + j\omega/c]\,\phi_m(r,\omega) = -\phi_x(r,\omega)\gamma(r,\omega) \times [1 - j\omega\tau(r)]\,. \quad (3.8)$$

As explained in Appendix A, this new parameterization is useful because, in a sequential optimization scheme, it takes advantage of the inherent linearity of the fluorescence inverse problem while allowing regularization to be applied to τ directly. The sets of flux measurements corresponding to the above image vectors may be defined, respectively, as y_x, y_m, and y_f.

3.2 Inversion

The estimation of each of the unknown images $\{x_x, x_m, x_f\}$ from the corresponding observations $\{y_x, y_m, y_f\}$ is an ill-posed, typically underdetermined, inverse problem. As in previous work [24, 25, 27, 28], we address this by formulating the inverse problem in a Bayesian framework. This framework allows the incorporation of *a priori* information, and it encapsulates all available information about the problem model into an objective function to be optimized. Let x denote one of the images of (3.4)-(3.6), and let y denote its corresponding observations. We use Bayes' rule to compute the maximum *a posteriori* (MAP) estimate, given by

$$\hat{x}_{MAP} = \arg\max_{x\geq 0}\{\ p(y|x)p(x)\ \}, \quad (3.9)$$

where $p(y|x)$ is the data likelihood and $p(x)$ is the prior density for the image. The data likelihood can be formed from a Gaussian model by considering, for example, the physical properties of a photocurrent shot noise-limited measurement system [24]. This gives

$$p(y|x) = \frac{1}{(\pi\alpha)^P|\Lambda|^{-1}} \exp\left[-\frac{||y - f(x)||^2_\Lambda}{\alpha}\right], \quad (3.10)$$

where P is the number of measurements, f is the appropriate forward operator, α is a scalar parameter that scales the noise variance, and, for an arbitrary vector w, $||w||_\Lambda^2 = w^H \Lambda w$ (where H denotes Hermitian transpose), and $\frac{\alpha}{2}\Lambda^{-1}$ is the covariance matrix. In a small signal shot noise model, the measurements are independent and normally distributed with a mean equal to the exact (noiseless) measurement and a variance proportional to the exact measurement at a modulation frequency of zero (DC). Following Ye *et al.* [24], we approximate the DC flux for the i^{th} datum as $|y_i|$. The resulting covariance matrix is given by

$$\alpha\Lambda^{-1} \quad = \quad \alpha \text{diag}[|y_1|, |y_2|, ...|y_P|]. \tag{3.11}$$

For the prior density $p(x)$, we use the generalized Gaussian Markov random field (GGMRF) model, which enforces smoothness in the solution while preserving sharp edge transitions [24, 49]. For each node (representing a voxel) inside the image, we form a three-dimensional neighborhood from the 26 adjacent nodes. Let $x^T = [x_a^T \;\; x_b^T]$, as in (3.4)-(3.6). Assuming independence of x_a and x_b, the density function is given by

$$\begin{aligned} p(x) &= p(x_a) \cdot p(x_b) & (3.12) \\ &= \left[\frac{1}{\sigma_a^N z(p_a)} \exp\left(-\frac{1}{p_a \sigma_a^{p_a}} \sum_{\{i,j\} \in \mathcal{N}_a} b_{i-j} \left| x_i - x_j \right|^{p_a} \right) \right] \cdot \\ &\quad \left[\frac{1}{\sigma_b^N z(p_b)} \exp\left(-\frac{1}{p_b \sigma_b^{p_b}} \sum_{\{i,j\} \in \mathcal{N}_b} b_{i-j} \left| x_i - x_j \right|^{p_b} \right) \right], & (3.13) \end{aligned}$$

where the subscripts a and b have the same meaning as in (3.4)-(3.6), x_i denotes the i^{th} node of x, the set $\mathcal{N}$ consists of all pairs of neighboring nodes, and b_{i-j} is the weighting coefficient corresponding to the i^{th} and j^{th} nodes. The coefficients b_{i-j} are assigned to be inversely proportional to the node separation in a cube-shaped node layout, with the requirement that that $\sum_j b_{i-j} = 1$. The constants p and σ control the shape and scale of the distribution, and the factor $z(p)$ is a normalization term.

As in previous work [25], we incorporate α into the inverse problem as an unknown for each image. We have found that this tends to improve the robustness and speed

of convergence. As a result, we perform a joint MAP estimation of both x and α for each image:

$$\hat{x}_x = \arg \max_{x_x \geq 0, \alpha_x} \{ p(x_x|y_x, \alpha_x) \} \tag{3.14}$$

$$\hat{x}_m = \arg \max_{x_m \geq 0, \alpha_m} \{ p(x_m|y_m, \alpha_m) \} \tag{3.15}$$

$$\hat{x}_f = \arg \max_{x_f \geq 0, \alpha_f} \{ p(x_f|y_f, \alpha_f, \hat{x}_x, \hat{x}_m) \}. \tag{3.16}$$

The estimations of x_x and x_m are performed independently of each other, using (3.1) and (3.3) as the respective forward models. Subsequently, these estimates are incorporated into the coupled diffusion equations (3.1) and (3.2) to estimate x_f.

Let x and α correspond to one of the images in (3.14)-(3.16). Ye *et al.* [25] showed that the above reconstructions are equivalent to maximizing the log posterior probability $l(x)$, which can be derived using (3.9), (3.10), and (3.13):

$$\begin{aligned} l(x) = & -P \ln ||y - f(x)||_\Lambda^2 - \frac{1}{p_a \sigma_a^{p_a}} \sum_{\{i,j\} \in \mathcal{N}_a} b_{i-j} |x_i - x_j|^{p_a} - \\ & \frac{1}{p_b \sigma_b^{p_b}} \sum_{\{i,j\} \in \mathcal{N}_b} b_{i-j} |x_i - x_j|^{p_b} \end{aligned} \tag{3.17}$$

Optimizing $l(x)$ may be implemented by alternating closed form updates of $\hat{\alpha}$ with updates of $\hat{x}$ [25]:

$$\hat{\alpha} = \frac{1}{P} ||y - f(\hat{x})||_\Lambda^2 \tag{3.18}$$

$$\hat{x} \simeq \arg \max_{x \geq 0} \{ \ln p(y|x, \hat{\alpha}) + \ln p(x|\hat{\alpha}) \}, \tag{3.19}$$

where $\simeq$ implies an update iteration, rather than a full optimization. The $\hat{x}$ updates represent more computationally expensive steps toward optimizing (3.9) than the $\hat{\alpha}$ updates. For each image, we form an objective function from (3.10) and (3.13):

$$\begin{aligned} c(x_x, \hat{\alpha}_x) = & \frac{1}{\hat{\alpha}_x} ||y_x - f_x(x_x)||_{\Lambda_x}^2 + \frac{1}{p_{xa} \sigma_{xa}^{p_{xa}}} \sum_{\{i,j\} \in \mathcal{N}_{xa}} b_{i-j} \left| x_{xa_i} - x_{xa_j} \right|^{p_{xa}} \\ & + \frac{1}{p_{xb} \sigma_{xb}^{p_{xb}}} \sum_{\{i,j\} \in \mathcal{N}_{xb}} b_{i-j} \left| x_{xb_i} - x_{xb_j} \right|^{p_{xb}} \end{aligned} \tag{3.20}$$

$$c(x_m, \hat{\alpha}_m) = \frac{1}{\hat{\alpha}_m} ||y_m - f_m(x_m)||^2_{\Lambda_m} + \frac{1}{p_{ma}\sigma_{ma}^{p_{ma}}} \sum_{\{i,j\}\in\mathcal{N}_{ma}} b_{i-j} \left|x_{ma_i} - x_{ma_j}\right|^{p_{ma}} + \frac{1}{p_{mb}\sigma_{mb}^{p_{mb}}} \sum_{\{i,j\}\in\mathcal{N}_{mb}} b_{i-j} \left|x_{mb_i} - x_{mb_j}\right|^{p_{mb}} \quad (3.21)$$

$$c(x_f, \hat{x}_x, \hat{x}_m, \hat{\alpha}_f) = \frac{1}{\hat{\alpha}_f} ||y_f - f_f(x_f, \hat{x}_x, \hat{x}_m)||^2_{\Lambda_f} + \frac{1}{p_{fa}\sigma_{fa}^{p_{fa}}} \sum_{\{i,j\}\in\mathcal{N}_{fa}} b_{i-j} \left|x_{fa_i} - x_{fa_j}\right|^{p_{fa}} + \frac{1}{p_{fb}\sigma_{fb}^{p_{fb}}} \sum_{\{i,j\}\in\mathcal{N}_{fb}} b_{i-j} \left|x_{fb_i} - x_{fb_j}\right|^{p_{fb}}. \quad (3.22)$$

The variables have the same meaning as in (3.10) and (3.13), and their subscripts have the same meaning as in (3.4)-(3.6). Note that forward operator f_f is a function of x_f and the estimates $\hat{x}_x$ and $\hat{x}_m$. In principle, one could jointly optimize (3.20)-(3.22) over x_x, x_m, and x_f, but for computational simplicity, we first optimize (3.20) and (3.21) and subsequently incorporate the estimates into (3.22). With the objective functions (3.20)-(3.22) established, an optimization algorithm to minimize these costs is needed, which is described in the next section.

3.3 Iterative coordinate descent optimization

The optimizations of (3.20)-(3.22) are performed using the iterative coordinate descent (ICD) algorithm [24,27,50], a sequential single-site update scheme similar to the Gauss-Seidel method used in other problems. One ICD scan consists of forming a local quadratic approximation to the cost function, followed by an update of each image element individually to minimize the approximate objective function. On each subsequent scan, the Fréchet derivative of the nonlinear forward operator is recomputed, and a new quadratic approximation is made.

Once again, let x denote one of the three images to be optimized. During the scan, the individual voxels of x are sequentially updated in random order. At the beginning of the scan, $f(x)$ is first expressed using a Taylor expansion as

$$||y - f(x)||_{\Lambda}^2 \simeq ||y - f(\hat{x}) - F'(\hat{x})\Delta x||_{\Lambda}^2 \ , \tag{3.23}$$

where $\Delta x = x - \hat{x}$, and $F'(\hat{x})$ represents the Fréchet derivative of $f(x)$ with respect to x at $x = \hat{x}$. Using (3.23), we formulate the approximate cost function

$$\begin{aligned} c(x, \hat{\alpha}) \ &\simeq \ \frac{1}{\hat{\alpha}} ||z - F'(\hat{x})x||_{\Lambda}^2 + \frac{1}{p_a {\sigma_a}^{p_a}} \sum_{\{i,j\} \in \mathcal{N}_a} b_{i-j} \left| x_i - x_j \right|^{p_a} \\ &\quad + \frac{1}{p_b {\sigma_b}^{p_b}} \sum_{\{i,j\} \in \mathcal{N}_b} b_{i-j} \left| x_i - x_j \right|^{p_b} \ , \end{aligned} \tag{3.24}$$

where

$$z \ = \ y - f(\hat{x}) + F'(\hat{x})\hat{x} \ . \tag{3.25}$$

With the other image elements fixed, the ICD update for $\hat{x}_i$ is given by

$$\begin{aligned} \hat{x}_i \ &= \ \arg\min_{x_i \geq 0} \left\{ \frac{1}{\hat{\alpha}} \left|\left| y - f(\hat{x}) - \left[F'(\hat{x})\right]_{*(i)} (x_i - \hat{x}_i) \right|\right|_{\Lambda}^2 + \right. \\ &\quad \left. \frac{1}{p\sigma^p} \sum_{j \in \mathcal{N}_i} b_{i-j} \left| x_i - \hat{x}_j \right|^p \right\} , \end{aligned} \tag{3.26}$$

where $\left[f'(\hat{x})\right]_{*(i)}$ is the i^{th} column of the Fréchet matrix and $\mathcal{N}_i$ is the set of nodes neighboring node i, and p and σ are chosen appropriately from $\{p_a, p_b\}$ and $\{\sigma_a, \sigma_b\}$. This one-dimensional minimization is solved by use of a simple half-interval search [24]. The Fréchet matrices used for each image are given in Appendix A. Appendix B summarizes the ICD optimization algorithm in pseudocode form.

Previously, we have found that multiresolution techniques can reduce the computational burden and improve robustness of convergence for the optical diffusion tomography problem [25]. Hence, for large computational domains, it may be beneficial to perform several ICD scans at a reduced resolution followed by interpolation as an initialization step for the full-resolution problem.

3.4 Simulations

Figure 3.2 shows cross section images of a 17.3 × 17.3 × 6 cm tissue phantom having background values $\mu_{a_{x,m}} = 0.01$ cm^{-1}, $D_{x,m} = 0.047$ cm, $\tau = 0$ ns and $\eta\mu_{a_f} = 0$ cm^{-1}. A slightly off-center spherical heterogeneity with diameter of roughly 3 cm is present, with $\mu_{a_x} = 0.05$ cm^{-1}, $\mu_{a_m} = 0.01$ cm^{-1}, D_x and $D_m = 0.30$ cm, $\tau = 0.55$ ns, and $\eta\mu_{a_f} = 0.02$ cm^{-1}. Figure 3.2(g) shows the location and size of the fluorophore as the $\eta\mu_{a_f} = 0.01$ cm^{-1} isosurface. As shown in Figure 3.3, the bottom face of the domain contains 16 sources (modulated at 70 MHz) arranged in a 4 × 4 grid pattern. On the top face, 16 detectors are placed in an identical grid. Using multigrid finite differences [51] to solve the diffusion equations, we generated synthetic measurements. Additive noise was introduced using the approximate shot noise model of (3.10) and (3.11), giving an average signal-to-noise ratio (SNR) of 34 dB and a maximum SNR of 41 dB. In the forward solution, an extrapolated zero-flux boundary condition [52] was used to model the free space absorbing boundaries.

For each of the x_x, x_m, and x_f inversions, 20 ICD iterations at a resolution of 17 × 17 × 9 nodes, followed by 20 ICD iterations at a resolution of 33 × 33 × 17 nodes, were performed. For the nonlinear x_x and x_m problems, multigrid finite differences were used to solve the forward model prior to each ICD image update. During the inversions, the log posterior probability was evaluated as the convergence criterion. For each image, convergence (with subsequent iterations changing the images very little) was obtained in approximately 10 minutes of computation on a AMD Athlon 1333 MHz workstation. While automatic estimation of the GGMRF hyperparameters p and σ is in principle possible using a maximum likelihood estimation technique [53], we follow Ye *et al.* [24] and use parameter values which empirically give good results. For each reconstruction, the estimates were initialized with homogeneous images equal to the correct background values, as the ICD method's convergence is slow for low spatial frequency image components [50]. We have shown previously that multigrid inversion methods in conjunction with ICD

updates alleviate this difficulty [25, 36, 54], but, again, we do not address them in this investigation.

Reconstructions $\hat{\mu}_{a_{x,m}}$, $\hat{D}_{x,m}$, $\hat{\tau}$, and $\hat{\eta\mu}_{a_f}$ are shown in Figure 3.4. We used $p = 2.0$ in all of the reconstructions, and σ values of 0.015 cm^{-1}, 0.02 cm, 0.5 ns, and 0.002 cm^{-1} in computing $\hat{\mu}_a$, $\hat{D}$, $\hat{\tau}$, and $\hat{\gamma}$, respectively. The reconstructions are qualitatively and quantitatively accurate.

We conjecture that changes in the lifetime parameter τ may be useful in distinguishing between diseased and healthy tissue environments. Hence, it is useful to determine if such changes are within the accuracy of the reconstruction algorithm. The simulation was repeated for four fluorophores, each with a different value of τ: 0.1375 ns, 0.275 ns, 0.55 ns, and 1.10 ns. The procedure outlined above was performed for each image. To determine a single lifetime value for each reconstructed image, we used a weighted average:

$$\hat{\tau}_{avg} = \frac{\sum_{i=0}^{N-1} \hat{\gamma}(r_i)\hat{\tau}(r_i)}{\sum_{i=0}^{N-1} \hat{\gamma}(r_i)}. \tag{3.27}$$

The weighted average is reasonable, as the reconstruction $\hat{\tau}(r)$ could be significant in spurious regions where $\eta\hat{\mu}_{a_f} \approx 0$. It is also similar to the weighting that occurs in the source term of (3.8), which represents the effect of τ on the data. Figure 3.5 shows a plot of $\hat{\tau}_{avg}$ as a function of the true value. The result suggests that the method can track even small changes in diagnostic lifetime imaging applications.

We also investigated the propagation of error from the $\hat{x}_x$ and $\hat{x}_m$ images into the $\hat{x}_f$ result. We created a series of incorrect images for $\hat{x}_x$ and $\hat{x}_m$, and we used the incorrect guesses in reconstructing $\hat{x}_f$. The incorrect $\hat{x}_x$ and $\hat{x}_m$ images were homogeneous. We tested the effect of varying μ_{a_x}, D_x, μ_{a_m}, and D_m, and we computed an error metric to quantify the change. For the $\hat{\eta\mu}_{a_f}$ results, the error metric was the normalized root mean squared error (NRMSE), defined as

$$\text{NRMSE} = \left[\frac{\sum_{i=0}^{N-1} |\hat{\eta\mu}_{a_f}(r_i) - \eta\mu_{a_f}(r_i)|^2}{\sum_{i=0}^{N-1} |\eta\mu_{a_f}(r_i)|^2} \right]^{1/2}. \tag{3.28}$$

For the $\hat{\tau}$ results, the NRMSE is less appropriate, as $\hat{\tau}$ may be nonzero in areas where $\eta\hat{\mu}_{a_f} \approx 0$. Hence, the error metric for $\hat{\tau}$ was the fractional error, $|\hat{\tau}_{avg} - \tau_{true}|/\tau_{true}$.

The plots in Figure 3.6(a)-(d) show the NRMSE for $\hat{\eta\mu}_{a_f}$, as a function of background μ_{a_x}, D_x, μ_{a_m}, and D_m, respectively. The plots in Figure 3.7(a)-(d) show the fractional error for $\hat{\tau}_{avg}$, also as a function of background μ_{a_x}, D_x, μ_{a_m}, and D_m, respectively. Whenever one parameter was varied, the others were all set to the correct image's background value. In all plots, the "+" symbols depict the error metric values resulting from the erroneous images, and an "×" symbol shows the error metric value resulting from computing reconstructions $\hat{x}_x$ and $\hat{x}_m$ in advance. For $\hat{\eta\mu}_{a_f}$, the NRMSE is above 0.5 in all cases, even when the full reconstruction was done. This was likely due to blurring of the sharp edges as a result of using the GGMRF prior model [49] with $p = 2$. Ignoring the heterogeneities, but using the correct background values, noticeably increased the NRMSE. However, the NRMSE changed relatively little over a range of incorrect background values for all of the parameters, μ_{a_x}, D_x, μ_{a_m}, and D_m. The reconstructed $\hat{\eta\mu}_{a_f}$ images also looked qualitatively similar. This suggests that the reconstructed yield was fairly robust to errors in $\hat{x}_x$ and $\hat{x}_m$. However, $\hat{\tau}$ was highly sensitive to the use of the incorrect images.

3.5 Experimental results

To further evaluate the proposed reconstruction method, we performed fluorescence measurements [41]. The data were recorded using a CW imaging device and a versatile phantom box (shown in Figure 3.8), both described in detail elsewhere [55]. In the absence of phase information, we did not consider τ, and we assumed D to be constant and known based on physical considerations. The instrument has laser diode sources available at both 690 nm and 830 nm, and it has avalanche photodiode (APD) detectors. While 690 nm excitation is not ideal for ICG, a published excitation spectrum [56] indicates that the detected 830 nm emission intensity, using 690 nm excitation, is about 30% of the maximum value (obtained using 780 nm excitation), making it acceptable for this demonstration. The box had internal dimensions of 16 × 16 × 3.8 cm, where the last dimension is the vertical thickness. As shown

in Figure 3.9, 9 source fibers were connected to the bottom plate (at $z = -1.9$ cm), and 14 detector fibers were connected to the top plate (at $z = 1.9$ cm). A hollow, surface frosted glass sphere of outer diameter 2 cm and thickness of about 2 mm was mounted with its center near $z = 0.7$ cm, as depicted in Figure 3.10. This sphere was mounted on a small plastic stand on the bottom of the box. It was also connected to a closed circulation channel via thin, translucent rubber tubes leading outside the box, allowing fluorophore solution to be titrated into the sphere from an external reservoir over the course of the experiment. The titration allowed comparable data to be taken both without and with the fluorophore present for analysis and calibration purposes.

The box was filled with a suspension of 0.4% Intralipid. Assuming 690 nm excitation and 830 nm emission, this results in background values of approximately $D_x = 0.071$ cm, $D_m = 0.082$ cm, $\mu_{a_x} = 0.0052$ cm^{-1}, and $\mu_{a_m} = 0.03$ cm^{-1} [57, 58]. The sphere was initially filled with the same suspension, creating an essentially homogeneous domain (apart from the glass sphere, rubber tubes, and plastic stand). Before administration of the ICG, two sets of measurements, $y_x^{(base)}$ and $y_m^{(base)}$, were recorded using sources at 690 nm and at 830 nm, respectively. We refer to the measurements recorded before ICG administration as baseline measurements. Subsequently, ICG was introduced into the sphere at a concentration of 1.0 μmol/L. For the purposes of reconstructing μ_{a_x}, measurements, which we define as $y_x^{(uncal)}$, were recorded using 690 nm sources with no optical bandpass filters installed over the detectors. We neglect the fluorescence signal in these measurements, as published quantum efficiency values [16, 59] imply that its effect on the data is two or three orders of magnitude below the effects due to absorption. Following a similar procedure, 830 nm data, which we call $y_m^{(uncal)}$, were recorded (with no filters) for reconstructing μ_{a_m}. Upon completion of these measurements, 830 nm bandpass filters with 12 nm FWHM (Newport 10LF10-830) were installed in front of the detectors to perform the fluorescence measurements. Due to a limited number of filters, only 9 of the 14

detectors were used for recording the filtered fluorescence measurements (as shown in Figure 3.9).

The installation of the filters required disconnection of the detector fibers from the detection devices. In principle, the disconnection and subsequent reconnection of the fibers could invalidate the previous baseline calibrations by potentially changing the detectors' coupling efficiencies. Such effects might adversely affect the y_x and y_m measurements. Hence, a new baseline calibration procedure was performed. This need for multiple calibrations is a limitation in the design of the experiment which could be alleviated by using a different detection scheme. For example, the instrument used by Ntziachristos and Weissleder [22], in which a CCD camera imaged a detection fiber array, required only the installation of a single fluorescence filter without perturbing the detection fibers. Incorporating the unknown calibration parameters into the inverse problem [28, 40] may also alleviate this difficulty.

To perform the new calibration, the ICG-Intralipid mixture was pumped out of the sphere and replaced with new Intralipid without ICG. Baseline measurements $y_f^{(base)}$ with 690 nm sources and 830 nm detection were made. Subsequently, a new ICG-Intralipid mixture identical in concentration to the previous one was titrated into the sphere. With the ICG now present, fluorescence measurements $y_f^{(uncal)}$ were recorded using 690 nm source excitation and 830 nm detection.

Before applying the reconstruction algorithm for x_x and x_m, calibrations using the baseline data were performed. Synthetic data $y_x^{(comp)}$ (for 690 nm sources) and $y_m^{(comp)}$ (for 830 nm sources) were computed for a homogeneous phantom with $D_x = 0.071$ cm, $D_m = 0.082$ cm $\mu_{a_x} = 0.0052$ cm^{-1}, and $\mu_{a_m} = 0.03$ cm^{-1} on a $33 \times 33 \times 17$ grid. Calibrations were performed by normalizing to these computed data:

$$y_{x_i} = y_{x_i}^{(uncal)} \frac{y_{x_i}^{(comp)}}{y_{x_i}^{(base)}} \tag{3.29}$$

$$y_{m_i} = y_{m_i}^{(uncal)} \frac{y_{m_i}^{(comp)}}{y_{m_i}^{(base)}}, \tag{3.30}$$

where the subscript i represents the i^{th} component of the data vector. This baseline calibration procedure estimates the unknown scaling and coupling efficiencies in

the measurements. For the much dimmer fluorescence measurements, the baseline data $y_f^{(base)}$ contained significant background signal. Calibrations were performed to account for the unknown coupling efficiencies and to remove these background components from the fluorescence data:

$$y_{f_i} = \left(y_{f_i}^{(uncal)} - y_{f_i}^{(base)}\right) \frac{y_{x_i}^{(comp)}}{y_{x_i}^{(base)}}, \tag{3.31}$$

where we have used the 690 nm calibration factors. The resulting fluorescence data contain an unknown scale factor, due to the unknown filter attenuation of the 690 nm excitation light relative to the 830 nm fluorescence light.

The reconstructions $\hat{\mu}_{a_x}$ and $\hat{\mu}_{a_m}$ are shown in Figures 3.11 and 3.12, respectively. For each inversion, a volume representing the whole box was discretized into $33 \times 33 \times 17$ voxels. The $\hat{\mu}_{a_x}$ computation used $\sigma = 0.015$ cm^{-1} and $p = 2$, and the $\hat{\mu}_{a_m}$ computation used $\sigma = 0.03$ cm^{-1} and $p = 2$. For both images, the ICD algorithm was run for 20 iterations on a 927 MHz Pentium III workstation, taking approximately 10 minutes. The resulting $\hat{\mu}_{a_x}$ and $\hat{\mu}_{a_m}$ images show a heterogeneity with accurate shape, though with artifacts present in the region close to the top plate. In both images, the sphere's vertical positions are similar, but below the true location by approximately 4 or 5 mm. The similarity of the two reconstructions, despite the fact that they are based on independent data sets, suggests that this error is due to a systematic effect in the reconstruction method. They may be a result of calibration errors, as the assumption of a diffuse, homogeneous medium in the baseline calibrations neglected the presence of the low-scattering glass sphere, the plastic stand used to hold the sphere, and the thin rubber tubes used for pumping in ICG solution. Small errors in the assumed D_x and D_m values might also contribute to artifacts in the reconstructions. In addition, placing the sphere close to the detectors may have resulted in modeling errors under the diffusion approximation. In $\hat{\mu}_{a_x}$, the reconstructed ICG absorption is slightly smaller than the predicted value of 0.039 cm^{-1} which one would expect from the results of Sevick-Muraca *et al.* [16], after correcting for the use of 690 nm, rather than 780 nm, excitation with the afore-

mentioned 30% factor [56]. The $\hat{\mu}_{a_m}$ image has higher contrast than the $\hat{\mu}_{a_x}$ image, in contrast to a published absorption spectrum for ICG 6.5 μmol/L, which shows higher absorption at 690 nm than at 830 nm. It is possible that ICG's instability in aqueous solution causes some variability in its optical spectrum, as Landsman *et al.* [60] observed a shift in the absorption peak toward longer wavelengths with decreasing concentration. In addition, the effect of an Intralipid suspension on ICG's absorption spectrum has not been investigated in detail, to our knowledge.

Figure 3.13 shows the fluorescent yield reconstruction $\eta\hat{\mu}_{a_f}$. As a result of the unknown scale factor in the fluorescence data, the image is in arbitrary units (AU). Making use of $\hat{\mu}_{a_x}$ and $\hat{\mu}_{a_m}$, the ICD algorithm, using $p = 2.0$ and $\sigma = 5.0$ AU , was run for 20 iterations (about 3 minutes). The iterations were computationally inexpensive due to the linearity of the fluorescence inverse problem. In contrast to the absorption reconstructions, the reconstructed fluorophore's center is slightly higher than that of the true fluorophore. Though no quantitative information is available, the size and shape are approximately correct.

3.6 Conclusion

We have presented a computationally efficient Bayesian inversion strategy for reconstructing fluorescence, absorption, and scattering properties, and demonstrated the method in a simulation study and in a tissue phantom experiment to image fluorescence from ICG in a spherical heterogeneity. The results show potential for use of optical diffusion tomography with fluorescence as a tool for localizing fluorescent contrast agents in clinical diagnostic applications.

It is worth noting that our experimental calibration procedure simplifies the problem substantially by choosing a background absorption value in advance and observing changes with respect to baseline measurements. We have found that this procedure reconstructs similar-looking absorbers, over a wide range of background values. Ongoing work with nonlinear multigrid reconstruction algorithms [25], auto-

matic optode calibration [28], and more accurate transport models [61] may improve reconstruction accuracy with less favorable calibrations.

Fluorescence imaging may offer some clinical advantages over absorption imaging. One advantage is the possibility of using lifetime as a diagnostic tool. For the yield imaging problem, the error propagation study suggests that qualitative results may be obtainable without doing the full x_x and x_m inversions. Although the reconstructed lifetime was highly sensitive to incorrect background properties, the reconstructed yield was not. Hence, as Ntziachristos and Weissleder [22] have observed, simplified first-order models are more easily applied to the fluorescent yield imaging problem than to the full absorption imaging problem. This enables a simpler experimental approach which requires no baseline data. For qualitative localization of tumors, this could prove to be a decisive advantage of fluorescence imaging over absorption imaging.

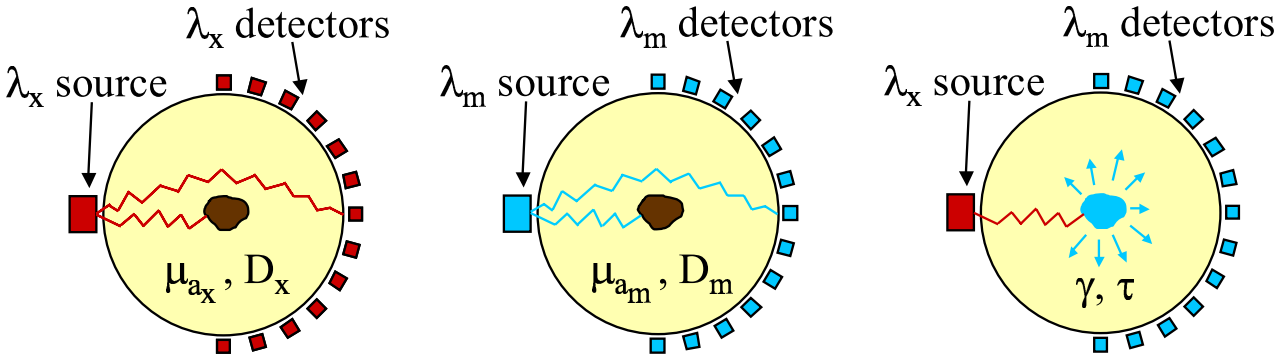


Fig. 3.1. Proposed measurement scheme.

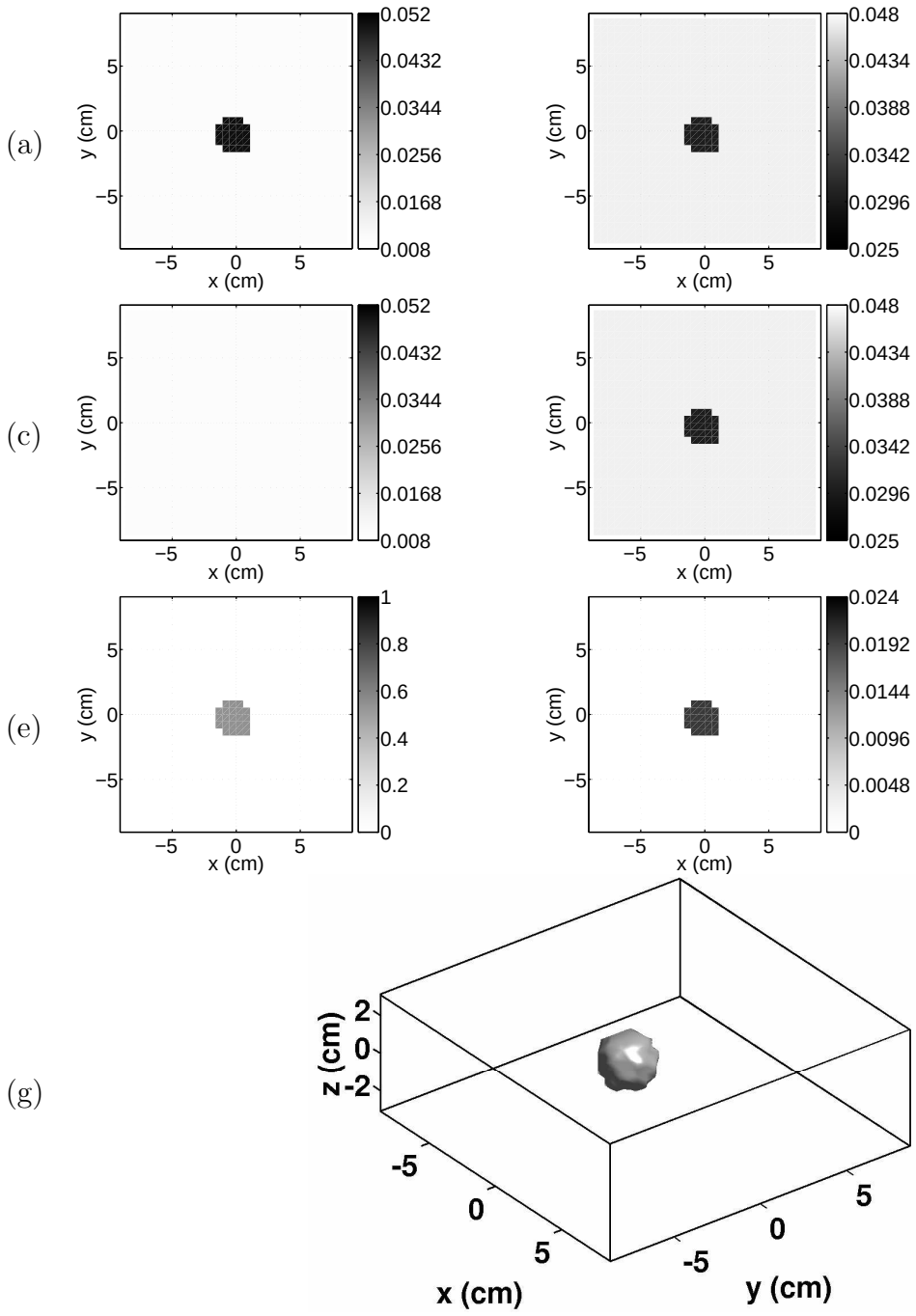


Fig. 3.2. True phantom, with cross sections of the widest part of the heterogeneity: (a) μ_{a_x} in cm^{-1} (b) D_x in cm (c) μ_{a_m} in cm^{-1} (d) D_m in cm (e) τ in ns (f) $\eta\mu_{a_f}$ in cm^{-1} (g) $\eta\mu_{a_f} = 0.01$ cm^{-1} isosurface

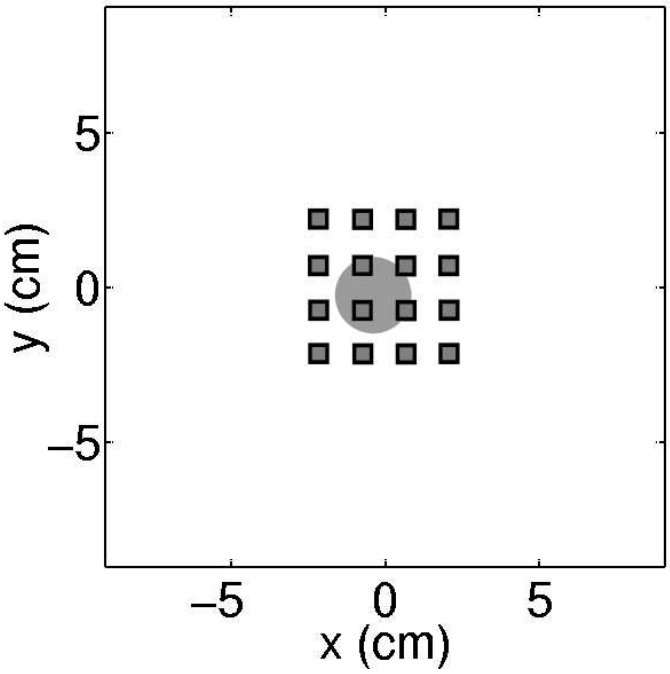


Fig. 3.3. Grid used for both sources and detectors in the simulation, with the relative location of the sphere depicted.

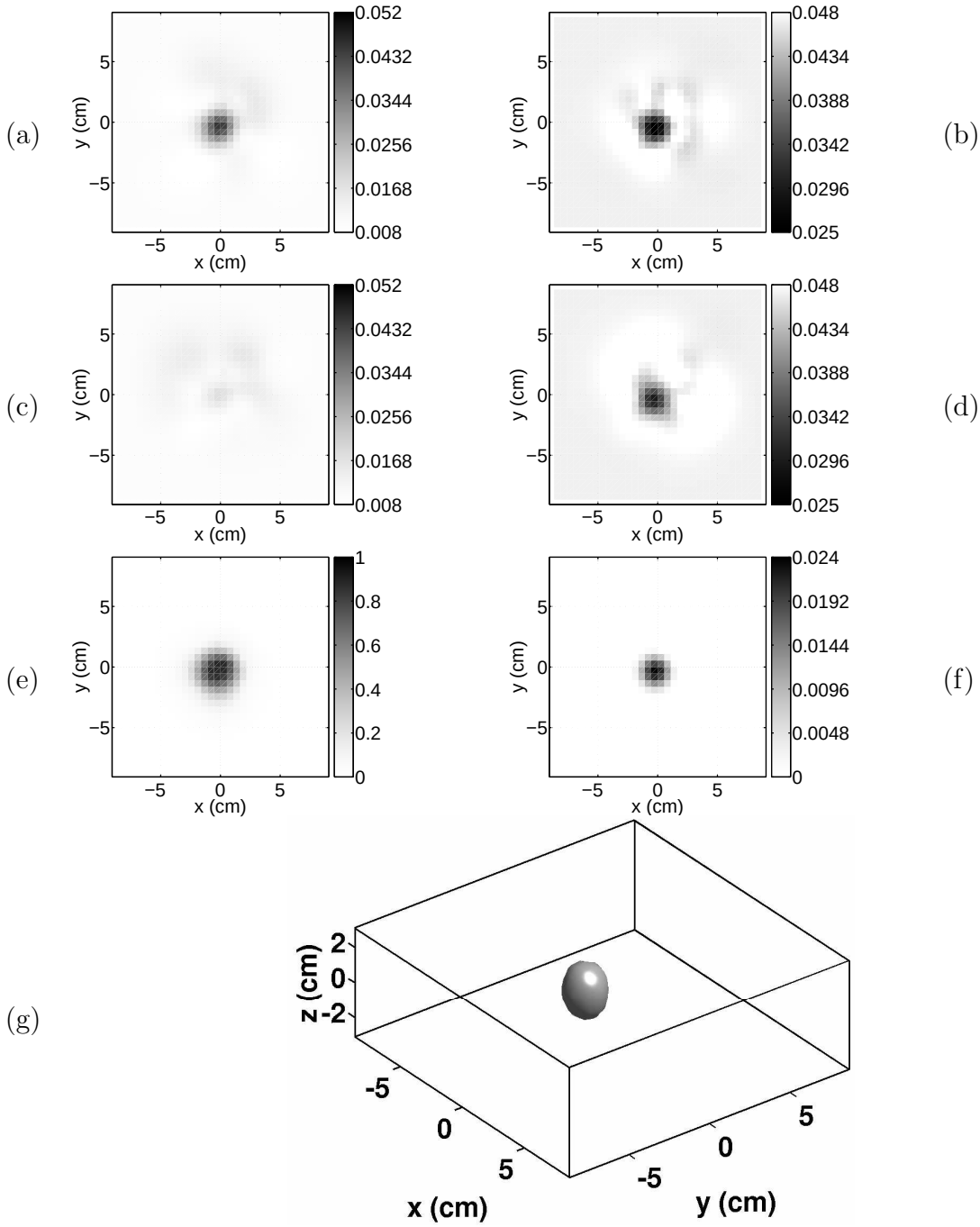


Fig. 3.4. Reconstructed phantom: (a) $\hat{\mu}_{a_x}$ in cm^{-1} (b) $\hat{D}_x$ cm (c) $\hat{\mu}_{a_m}$ in cm^{-1} (d) $\hat{D}_m$ in cm (e) $\hat{\tau}$ in ns (f) $\hat{\eta}\hat{\mu}_{a_f}$ in cm^{-1} (g) $\hat{\eta}\hat{\mu}_{a_f} = 0.01$ cm^{-1} isosurface

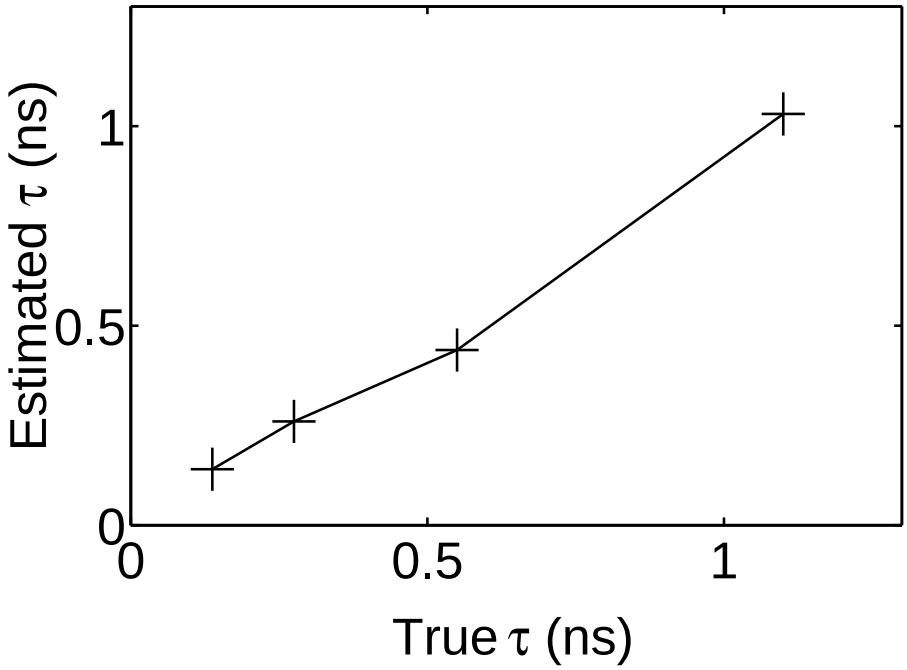


Fig. 3.5. Plot of estimate $\hat{\tau}_{avg}$ versus the true value of τ. The trend is almost linear, as desired.

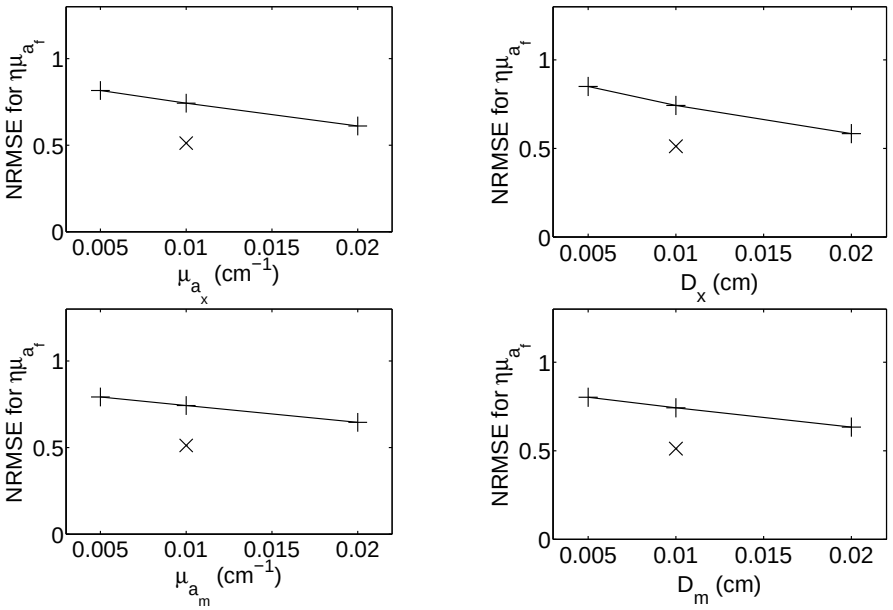


Fig. 3.6. NRMSE for $\hat{\eta\mu}_{a_f}$ due to changes in assumed constant background values for: (a) μ_{a_x} in cm^{-1} (b) D_x cm (c) μ_{a_m} in cm^{-1} (d) D_m in cm

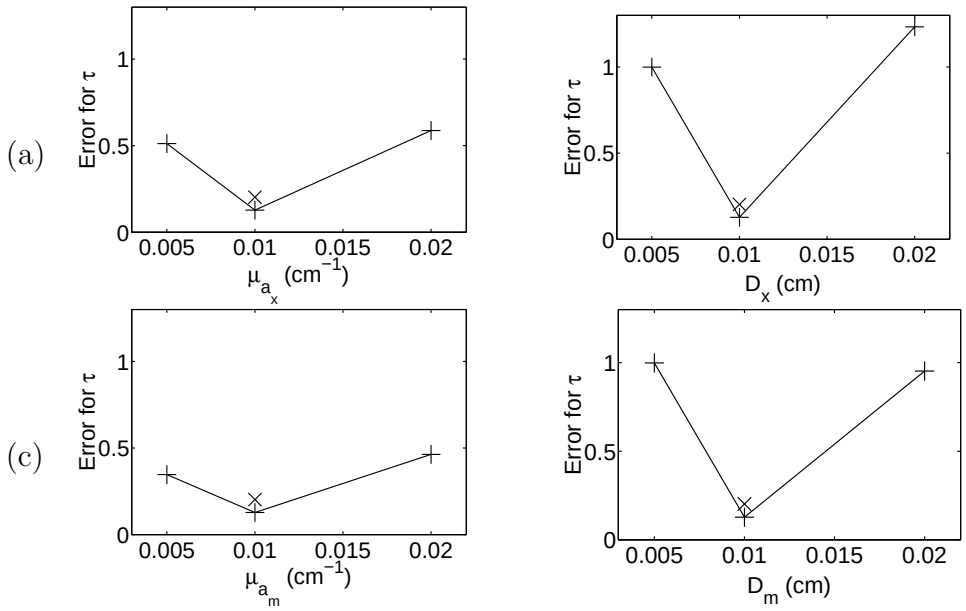


Fig. 3.7. Fractional error for $\hat{\tau}_{avg}$ due to changes in assumed constant background values for: (a) μ_{a_x} in cm^{-1} (b) D_x cm (c) μ_{a_m} in cm^{-1} (d) D_m in cm

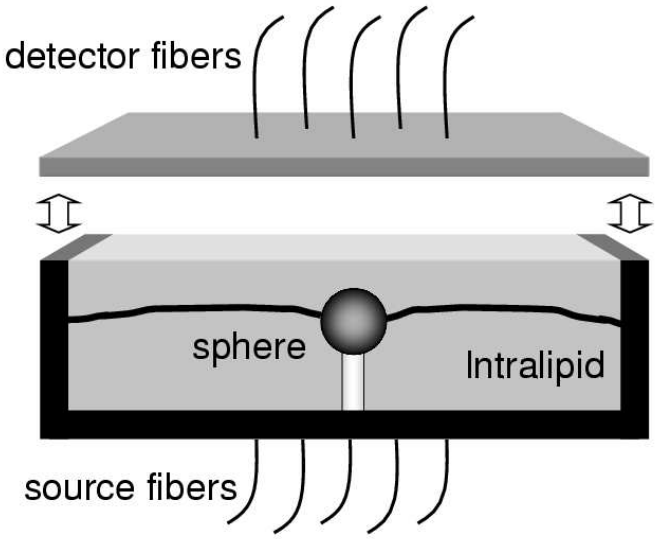
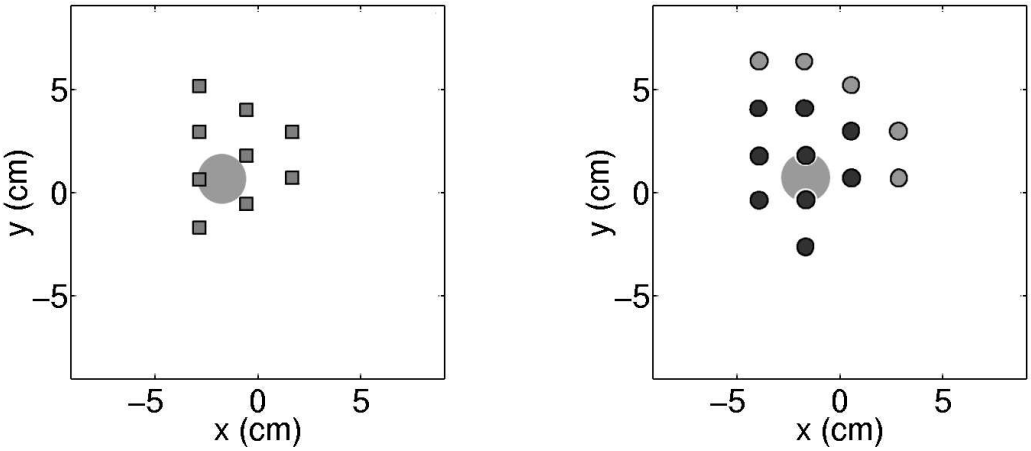
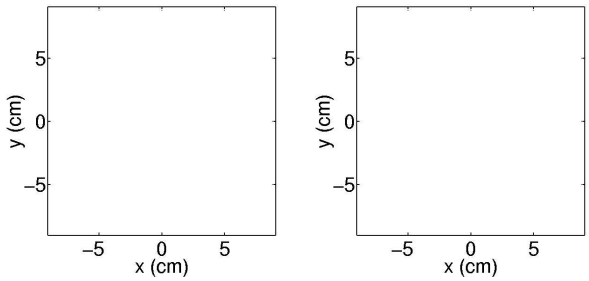


Fig. 3.8. Phantom box schematic, showing the fibers, the spherical heterogeneity, and the removable lid.

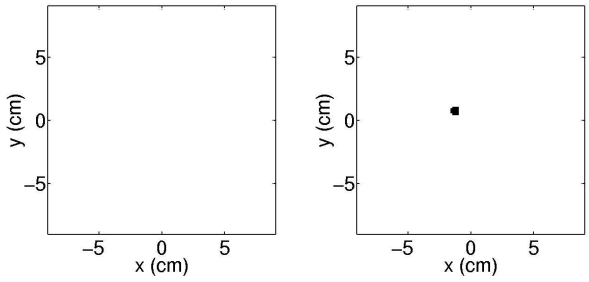


(a) Bottom plate (sources) (b) Top plate (detectors)

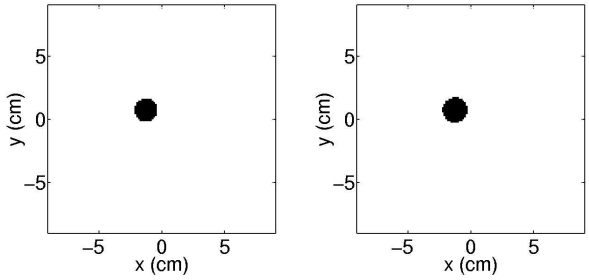
Fig. 3.9. Source and detector layout for experiment. The blackened detector symbols represent detector positions used in the fluorescence measurements. The relative location of the sphere is also depicted.



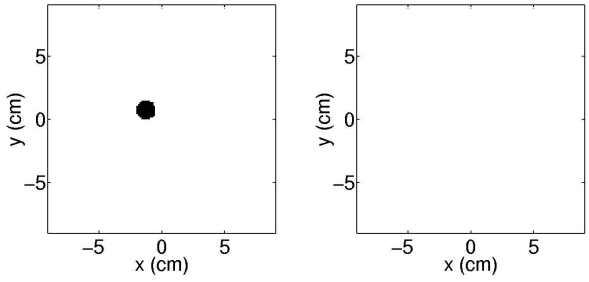
(a) z=-1.82 cm (b) z=-1.30 cm



(c) z=-0.78 cm (d) z=-0.26 cm

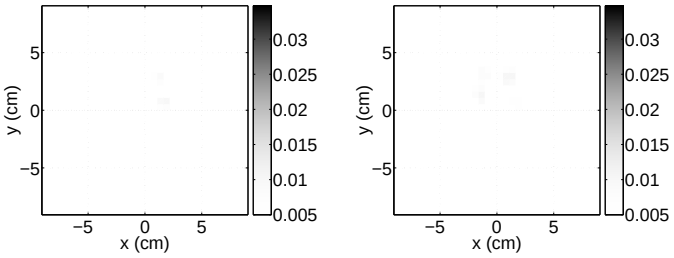


(e) z=0.26 cm (f) z=0.78 cm

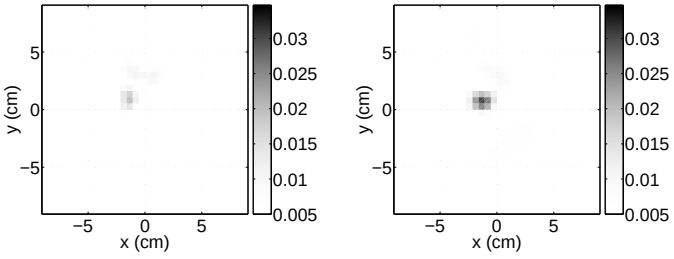


(g) z=1.30 cm (h) z=1.82 cm

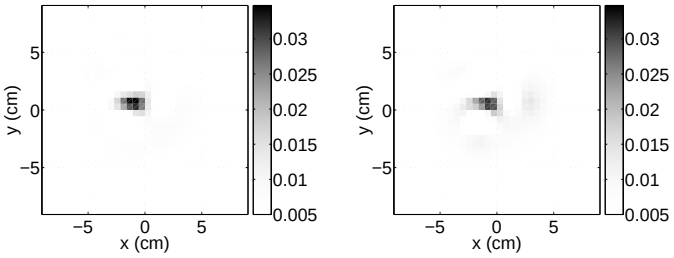
Fig. 3.10. True fluorophore location



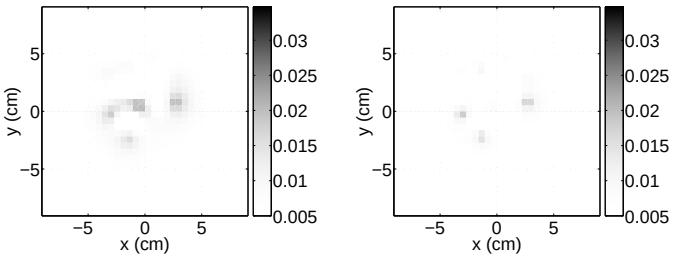
(a) z=-1.82 cm (b) z=-1.30 cm



(c) z=-0.78 cm (d) z=-0.26 cm

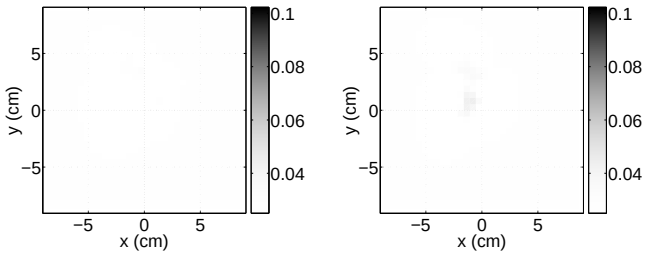


(e) z=0.26 cm (f) z=0.78 cm

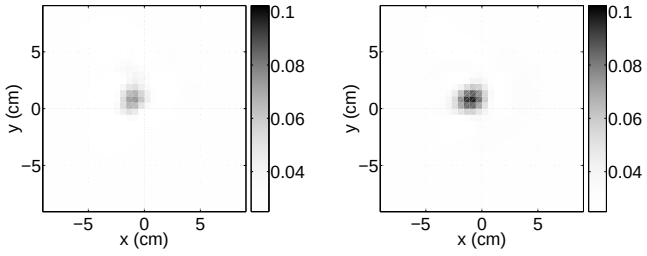


(g) z=1.30 cm (h) z=1.82 cm

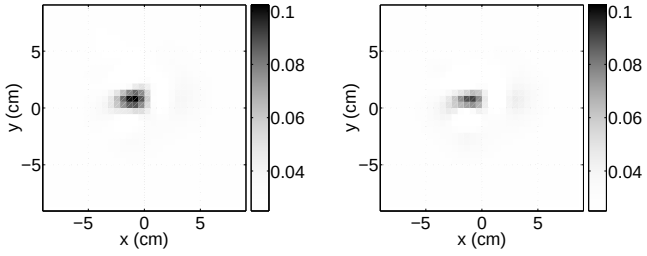
Fig. 3.11. Reconstructions of μ_{a_x} in cm^{-1}



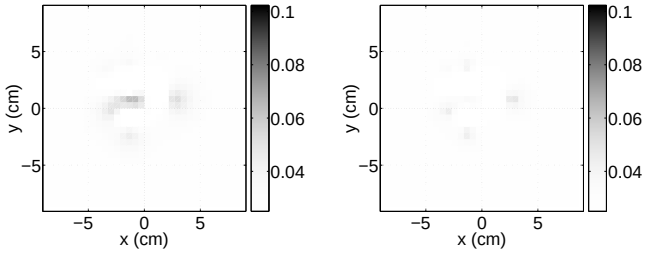
(a) z=-1.82 cm (b) z=-1.30 cm



(c) z=-0.78 cm (d) z=-0.26 cm

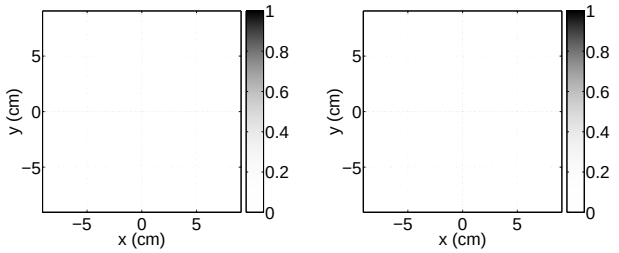


(e) z=0.26 cm (f) z=0.78 cm

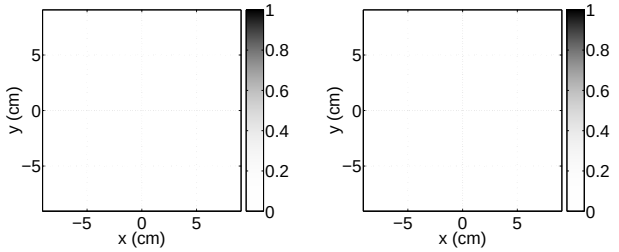


(g) z=1.30 cm (h) z=1.82 cm

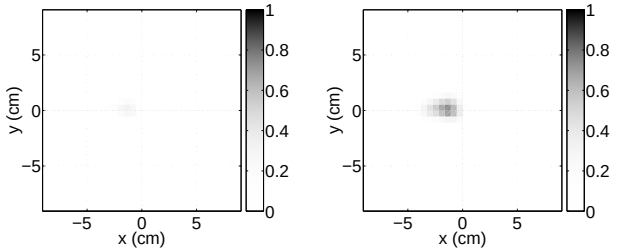
Fig. 3.12. Reconstructions of μ_{a_m} in cm^{-1}



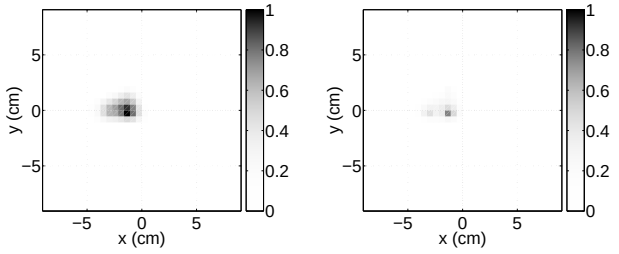
(a) z=-1.82 cm (b) z=-1.30 cm



(c) z=-0.78 cm (d) z=-0.26 cm



(e) z=0.26 cm (f) z=0.78 cm



(g) z=1.30 cm (h) z=1.82 cm

Fig. 3.13. Reconstructions of $\eta\mu_{a_f}$ in arbitrary units

4. FLUORESCENCE OPTICAL DIFFUSION TOMOGRAPHY USING MULTIPLE-FREQUENCY DATA

4.1 Background

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 [62, 63]. Intuitively, one might expect that additional modulation frequencies would provide additional information which 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 multi-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 using multiple modulation frequencies [42]. 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 the 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 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 [64, 65] for evaluating an experimental configuration. The performance met-

ric 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 (CT) [66] and magnetic resonance imaging (MRI) [67]. 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 (SPECT) [68, 69]. While information theory has not been applied to the ODT problem, several groups have used singular value analysis to evaluate source/detector arrangements [70, 71] or data types [72]. 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 [73], it provides a lower bound for the mean squared 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.

4.2 Fluorescence Optical Diffusion Tomography Models

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 i^{th} voxel centroid, and let η

denote $\eta\mu_{a_f}$, for brevity. We define three image vectors, with each corresponding to a particular measurement:

$$\begin{aligned} x_x &= \left[\ x_{xA}^T, x_{xB}^T\ \right]^T \\ &= [\mu_{a_x}(r_1)\cdots\mu_{a_x}(r_N), D_x(r_1)\cdots D_x(r_N)]^T \qquad (4.1) \\ x_m &= \left[\ x_{mA}^T, x_{mB}^T\ \right]^T \\ &= [\mu_{a_m}(r_1)\cdots\mu_{a_m}(r_N), D_m(r_1)\cdots D_m(r_N)]^T \qquad (4.2) \\ x_f &= \left[\ x_{fA}^T, x_{fB}^T\ \right]^T \\ &= [\eta(r_1)\cdots\eta(r_N), \tau(r_1)\cdots\tau(r_N)]^T, \qquad (4.3) \end{aligned}$$

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.

4.3 Inverse Problem

4.3.1 Definitions

Previously [27, 41], we have shown how to reconstruct x_x and x_m in a Bayesian framework that can incorporate multiple frequencies. Hence, suppose 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. (3.2) as part of the source term. Let

$$h(x_f, r, \omega) = \eta(r)\frac{1 - j\omega\tau(r)}{1 + [\omega\tau(r)]^2}. \qquad (4.4)$$

Also, let $g(r_{s_k}, r_{d_{m'}}; \omega, x)$ be the diffusion equation Green's function for the problem domain computed 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

$$\begin{aligned}\phi_f(r_{s_k}, r_{d_{m'}}; \omega, x_f) = \\ \int h(x_f, r, \omega) g_x(r_{s_k}, r; \omega, x_x) g_m(r, r_{d_{m'}}; \omega, x_m) d^3r,\end{aligned} \tag{4.5}$$

where we omit the dependence on x_x and x_m for simplicity. From Eq. (3.2),

$$\begin{aligned}h(x_f, r, \omega) g_x(r_{s_k}, r; \omega, x_x) \quad = \\ \phi_x(r, \omega) \frac{\eta(r)[1 - j\omega\tau(r)]}{1 + [\omega\tau(r)]^2}.\end{aligned} \tag{4.6}$$

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

$$f_\omega(x_f) = \begin{bmatrix} \phi_f(r_{s_1}, r_{d_1}; \omega, x_f) \\ \phi_f(r_{s_1}, r_{d_2}; \omega, x_f) \\ \vdots \\ \phi_f(r_{s_1}, r_{d_M}; \omega, x_f) \\ \phi_f(r_{s_2}, r_{d_1}; \omega, x_f) \\ \vdots \\ \phi_f(r_{s_K}, r_{d_M}; \omega, x_f) \end{bmatrix}. \tag{4.7}$$

Let Q be the number of modulation frequencies used, and

$$f(x_f) = \begin{bmatrix} f_{\omega_1}(x_f)^T, f_{\omega_2}(x_f)^T \cdots f_{\omega_Q}(x_f)^T \end{bmatrix}^T. \tag{4.8}$$

Similarly, we define the measurement vector y as:

$$y = \begin{bmatrix} y_{\omega_1}^T, y_{\omega_2}^T \cdots y_{\omega_Q}^T \end{bmatrix}^T \tag{4.9}$$

corresponding to the same order used in Eq. (4.8). 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, due to the reciprocity theorem [74].

For the discretized problem, we define matrix operators to approximate the integration in Eq. (4.5):

$$G^x(\omega) = \begin{bmatrix} g_x(r_{s_1}, r_1; \omega, x_x) & \cdots & g_x(r_{s_1}, r_N; \omega, x_x) \\ \vdots & \ddots & \vdots \\ g_x(r_{s_K}, r_1; \omega, x_x) & \cdots & g_x(r_{s_K}, r_N; \omega, x_x) \end{bmatrix} \tag{4.10}$$

$$G^m(\omega) = \begin{bmatrix} g_m(r_{d_1}, r_1; \omega, x_m) & \cdots & g_m(r_{d_1}, r_N; \omega, x_m) \\ \vdots & \ddots & \vdots \\ g_m(r_{d_M}, r_1; \omega, x_m) & \cdots & g_m(r_{d_M}, r_N; \omega, x_m) \end{bmatrix} \tag{4.11}$$

$$J_\omega = V \cdot \begin{bmatrix} G^x_{1,1}(\omega)G^m_{1,1}(\omega) & \cdots & G^x_{1,N}(\omega)G^m_{1,N}(\omega) \\ \vdots & \ddots & \vdots \\ G^x_{1,1}(\omega)G^m_{M,1}(\omega) & \cdots & G^x_{1,N}(\omega)G^m_{M,N}(\omega) \\ G^x_{2,1}(\omega)G^m_{1,1}(\omega) & \cdots & G^x_{2,N}(\omega)G^m_{1,N}(\omega) \\ \vdots & \ddots & \vdots \\ G^x_{K,1}(\omega)G^m_{M,1}(\omega) & \cdots & G^x_{K,N}(\omega)G^m_{M,N}(\omega) \end{bmatrix} \tag{4.12}$$

where V is the volume of a voxel. Let

$$h_\omega(x_f) = \begin{bmatrix} h(x_f, r_1, \omega) & \cdots & h(x_f, r_N, \omega) \end{bmatrix}^T \tag{4.13}$$

Neglecting discretization error,

$$f_\omega(x_f) = J_\omega h_\omega(x_f). \tag{4.14}$$

Suppressing the $\hat{x}_x$ and $\hat{x}_m$ arguments for brevity, we can rewrite Eq. (3.22) as:

$$\begin{aligned} c(x_f, \hat{\alpha}_f) = \frac{1}{\hat{\alpha}_f} \sum_{q=1}^{Q} \left|\left| y_{f\omega_q} - J_{\omega_q} h_{\omega_q}(x_f) \right|\right|^2_{\Lambda_{f\omega_q}} \\ + \frac{1}{\rho_{fA}\sigma_{fA}^{\rho_{fA}}} \sum_{\{i,j\}\in\mathcal{N}_{fA}} b_{i-j} \left| x_{fA_i} - x_{fA_j} \right|^{\rho_{fA}} \\ + \frac{1}{\rho_{fB}\sigma_{fB}^{\rho_{fB}}} \sum_{\{i,j\}\in\mathcal{N}_{fB}} b_{i-j} \left| x_{fB_i} - x_{fB_j} \right|^{\rho_{fB}}. \end{aligned} \tag{4.15}$$

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 [41]. However, since the resulting parameters must contain frequency terms in their denominators, they cannot be used for the multiple frequency case. Hence, we perform a nonlinear optimization of Eq. (4.15) directly over $\{\eta, \tau\}$.

4.3.2 Iterative Coordinate Descent

As in our previous work [24,25,27,28,41], we use the iterative coordinate descent (ICD) algorithm [50], a Gauss-Seidel approach, to optimize (4.15). The voxels are scanned in random order, and the cost function is optimized with respect to each individual voxel. Previously [41], we have shown how to optimize (3.20) and (3.21). Hence, we focus on (3.22), 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 i^{th} element of x. With all other image elements fixed, the ICD update for the estimate $\hat{x}_i$ is given by

$$\begin{aligned} \hat{x}_i \quad \leftarrow \quad & \arg\min_{x_i \geq 0} \left\{ \frac{1}{\hat{\alpha}} \sum_{q=1}^{Q} \left|\left| y_{\omega_q} - \left[J_{\omega_q}\right]_{*(i)} h(x, r_i, \omega_q) \right|\right|^2_{\Lambda_{\omega_q}} \right. \\ & \left. + \frac{1}{\rho \sigma^\rho} \sum_{j \in \mathcal{N}_i} b_{i-j} \left| x_i - \hat{x}_j \right|^\rho \right\} , \end{aligned} \tag{4.16}$$

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 $\{\sigma_A, \sigma_B\}$. $\left[J_{\omega_q}\right]_{*(i)}$ denotes the i^{th} column of J_{ω_q}. Suppose we have an initial guess $\tilde{x}$, and let $z_{\omega_q} = y_{\omega_q} - f_{\omega_q}(\tilde{x})$. Then, Eq. (4.16) is equivalent to

$$\begin{aligned}
\hat{x}_i \;\leftarrow\; & \arg\min_{x_i \geq 0} \Big\{ \\
& \frac{1}{\hat{\alpha}} \sum_{q=1}^{Q} ||z_{\omega_q} - [J_{\omega_q}]_{*(i)}[h(x, r_i, \omega_q) \\
& - h(\tilde{x}, r_i, \omega_q)]||^2_{\Lambda_{\omega_q}} \\
& + \frac{1}{\rho\sigma^\rho} \sum_{j \in \mathcal{N}_i} b_{i-j} \left| x_i - \hat{x}_j \right|^\rho \Bigg\} \;, \\
= \;& \arg\min_{x_i \geq 0} \Big\{ \\
& \frac{1}{\hat{\alpha}} \sum_{q=1}^{Q} \Big(\theta_{1,\omega_q}[h(x, r_i, \omega_q) - h(\tilde{x}, r_i, \omega_q)] \\
& + \frac{\theta_{2,\omega_q}}{2}[h(x, r_i, \omega_q) - h(\tilde{x}, r_i, \omega_q)]^2 \Big) \\
& + \frac{1}{\rho\sigma^\rho} \sum_{j \in \mathcal{N}_i} b_{i-j} \left| x_i - \tilde{x}_j \right|^\rho \Bigg\} ,
\end{aligned} \tag{4.17}$$

where

$$\theta_{1,\omega_q} = -2\mathrm{Re}\left\{ [J_{\omega_q}]^H_{*i} \Lambda_{\omega_q} z_{\omega_q} \right\} \tag{4.18}$$

$$\theta_{2,\omega_r} = 2[J_{\omega_q}]^H_{*i} \Lambda_{\omega_q} [J_{\omega_q}]_{*i} \tag{4.19}$$

In Eq. (4.17), 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 (1-D) line search over x_i. This property enables 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 [75].

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

4.4 Simulation

We performed a simulation incorporating multiple modulation frequencies by use of the proposed algorithm. Figure 4.1 (a) shows a cubic tissue phantom, with 5 sources on one side and 5 detectors on the other. Two small η heterogeneities were placed near each other, slightly off center. The other parameters were constant, with $\mu_{a_{x,m}} = 0.047$ cm $^{-1}$, $D = 0.27$ cm, $\tau = 0$ 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 [52] with internal sources interpolated among the nearest grid nodes. The sources were placed 0.08 cm inside the physical boundary, corresponding approximately to 1 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.* [24] and assuming equal input source power for each modulation frequency. The average signal to noise ratio (SNR) was 21.2 dB for the 78.4 MHz data and 14.7 dB for the 314 MHz data. Using 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, for all three reconstructions. For each reconstruction, using $\sigma = 0.021$ cm^{-1} gave substantially the best qualitative results. The algorithm was run to convergence, with subsequent iterations changing the cost function negligibly. To avoid singularities near the sources and detectors, we did not update within in a 2-voxel border of the computational boundary. Figure 4.1 (c)-(e) shows 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.

4.5 Experiment

4.5.1 Design and Procedure

To test the method, we prepared a time-domain fluorescence and absorption imaging experiment, shown schematically in Figure 4.2(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 [41], 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 Figure 4.2. 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 data. The 0.8 W average output of the laser was split using the reflection off of 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.

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 used in a calibration scheme which we have used previously [41]. 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, samples were recorded at 250 ps delay intervals. Due to 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 durations. 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, the laser was tuned 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 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.

To remap the raw data images into $2-D$ uniform grid coordinates, 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 was estimated in the least-squares sense. 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 slab model [47] with a direct line search. 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 baseline measurements. These factors were used to calibrate the data with the procedure of our previous work [41].

4.5.2 Reconstructions

To reconstruct μ_{a_x} and μ_{a_m} from the measured data, we used the ICD algorithm presented previously [27, 41] to optimize Eqs. (3.20) and (3.21). 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. 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 [53], we follow our previous work [24,41] and use parameter values which 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 $\sigma = 0.005$ cm^{-1} and $\rho = 2$. The estimates $\hat{\mu}_{a_x}$ and $\hat{\mu}_{a_m}$ were incorporated into the coupled diffusion equations, and η and τ were reconstructed using the proposed algorithm.

Figures 4.3 and 4.4 show the reconstructions of μ_{a_x} and μ_{a_m}, respectively, obtained using 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} which one would obtain by dividing a previously reported [16] 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 [59, 60].) The μ_{a_m} peak value is consistent with our previous results [41]. Aside from a significant artifact near one of the sources, the images also appear to be qualitatively reasonable. The reconsructed sphere's center is approximately 4 mm below the center (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 using 78, 314, and 627 data. Figure 4.5 shows the reconstruction of η, obtained using $\sigma = 2.5 \times 10^{-5}$ 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 data. The image is qualitatively accurate, with few artifacts and position similar to the reconstructed absorbers. Dividing the peak fluorescent yield by the peak absorption of the heterogeneity at λ_x, we obtain a quantum efficiency of about 0.018. This result is slightly higher than the value of 0.016 obtained by Sevick-Muraca *et al.* [16], who measured a micromolar aqueous solution with a spectrofluorometer. Figure 4.6 shows the reconstructed τ, obtained using $\sigma = 0.1$ ns and $\rho = 2$. Quantitatively, it is similar to a literature-reported lifetime of 0.56 ns [19]. The image appears spread out, due 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 equations).

Figures 4.7 and 4.8 show reconstructions of η and τ obtained using only 78 MHz data. The results are very similar to the reconstructions obtained using 78, 314, and 627 MHz data together. Hence, for this experiment, the additional modulation frequencies (perhaps surprisingly) do not appear to contribute information which is useful for reconstructing the sphere.

4.5.3 Effects of Regularization

Figures 4.9 and 4.10 show reconstructions of η and τ, respectively, obtained over a range of σ values. The η reconstructions in Figure 4.9 qualitatively show a progression from over- to underregularization with increasing σ (where the τ prior parameters were fixed to $\rho = 2$ and $\sigma = 0.1$ ns, as used above). Figure 4.9(a), which we deem to be overregularized, appears overly broadened, with blurred edges. In contrast, Figure 4.9(e) is underregularized, with sharp peaks appearing in the image. The τ images in Figure 4.10 were also computed over a range of σ values, with the η prior parameters fixed to $\rho = 2$ and $\sigma = 2.5 \times 10^{-4}$ 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 qualitatively reasonable, and not excessively broadened. Figures 4.10(d) and 4.10(e), which are deemed to be underregularized, have much of their structure outside the voxels containing the fluorophore in Figure 4.5 rather than within it.

One method of selecting σ for the fluorescence lifetime problem might be to choose the largest one which gives physically acceptable images. Less regularization implies a more accurate fit to the data irrespective of image quality. As in our previous work [41], we define the weighted average lifetime:

$$\hat{\tau}_{avg} = \frac{\sum_{i=1}^{N} \hat{\eta}(r_i)\hat{\tau}(r_i)}{\sum_{i=1}^{N} \hat{\eta}(r_i)}. \tag{4.20}$$

This weighted average accounts for the possibility that the reconstruction could be significant in spurious regions where $\eta\hat{\mu}_{a_f} \approx 0$. It is also related to the weighting that occurs in the source term of (3.2), and thus represents the effect of τ on the data y_f. Figure 4.10(f) shows $\hat{\tau}_{avg}$ as a function of σ. As σ increases, $\hat{\tau}_{avg}$ asymptotically approaches a constant value of about 0.49 ns, just under the literature-reported value of 0.56 ns [19]. The stability of $\hat{\tau}_{avg}$ with increasing σ suggests that selecting the largest σ which produces a physically acceptable image is reasonable. However, despite the stability of $\hat{\tau}_{avg}$, the image becomes more distorted (and less physically meaningful) with increasing σ due to the inherent tradeoff between regularization and accurate fitting of the data. We selected $\sigma = 0.1$ ns, which gave $\hat{\tau}_{avg} = 0.205$ ns, as the best compromise. Although this gives a less than ideal result for $\hat{\tau}_{avg}$, it is physically reasonable, with the lifetime distributed evenly throughout the fluorophore.

4.6 Mutual Information Performance Measure

4.6.1 Motivation and Mathematics

The results of Sections 4.4 and 4.5 suggest that the additional modulation frequencies provide useful information for some problems, but not for others. Here,

we propose to use the mutual information, defined in information theory [64, 65, 73], as 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 [65]

$$I(X;Y) \;=\; H(Y) \;-\; H(Y|X), \tag{4.21}$$

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

$$H(Y) \;=\; E[-\log p_Y(Y)], \tag{4.22}$$

and

$$H(Y|X) \;=\; E[-\log p_{Y|X}(Y|X)]. \tag{4.23}$$

where $p_Y(y)$ and $p_{Y|X}(y|x)$ are the densities of X and $Y|X$ respectively. 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".

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

$$p_X(x) \;=\; \frac{1}{\sqrt{(2\pi\sigma)^N}}|C|^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}x^tCx\right\} \tag{4.24}$$

where the matrix C has elements

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

As in previous sections, we let

$$E[Y|X] = JX \tag{4.25}$$

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. (3.11). Using these definitions, we show in Appendix D that the mutual information is given by

$$I(X;Y) \;=\; \log\left|I + \frac{\sigma^2}{\alpha}\Lambda J C^{-1} J^H\right| \tag{4.26}$$

where I is the identity matrix.

While (4.26) 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 is also reduced. The minimum MSE estimate of X given Y is given by the conditional expectation $\hat{X} = E[X|Y]$. Furthermore, for each image X there exists a distortion rate function $D_X(\cdot)$ which lower bounds the achievable distortion at a specified rate [73]. In communication applications, the rate is the amount of information content per unit time which 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. For any particular choice of Y, we then know that

$$E[||X - \hat{X}||^2] \geq D_X(I(X;Y)) \; . \tag{4.27}$$

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

4.6.2 Application to Simulation and Experiment Models

Figure 4.11(a) shows the mutual information as a function of α for the same geometry and statistical models assumed in the simulation study of Section 4.4. 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 (multi-frequency) data provide significant information for the set of problems modeled by the same image and measurement statistics as in the simulation.

In Figure 4.11(b), the 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 [25]. In contrast to Figure 4.11(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, is not beneficial.

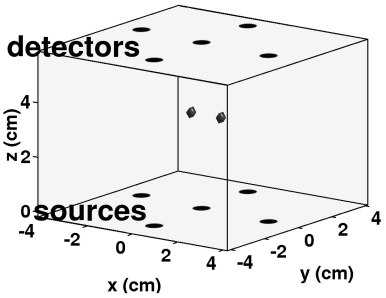
4.7 Conclusions

The proposed algorithm 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. Even when the numbers of sources and detectors are reduced, single and multiple frequency reconstructions produce substantially similar reconstructions of fluorescent yield and lifetime. We suppose that the additional, higher-resolution

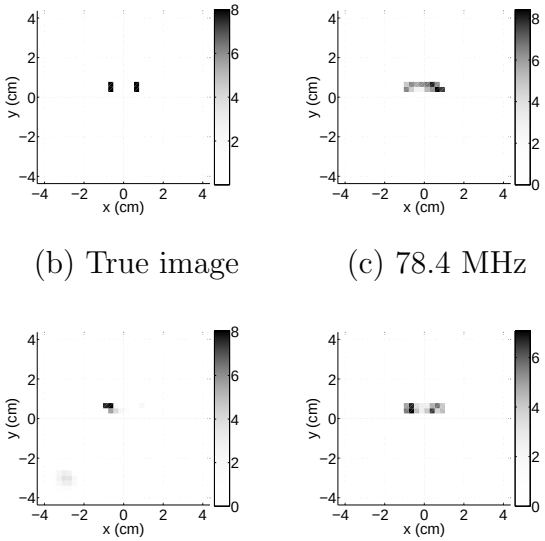
image components observable by the use of the additional modulation frequencies may be largely absent from the true sphere image. Simulations with single, spherical images seem to confirm that the sphere is fully resolved using just one frequency, in agreement with the experiment.

The 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 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. In principle, for Gaussian problems, it is possible to use other, related performance metrics such as the reconstruction MSE. However, the reconstruction MSE is computed as the trace of the inverse of a very large matrix [76], making it less computationally tractable than the mutual information.

As in a previous study [41], 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 [22], suggest that the baseline calibration step can be circumvented, under certain simplifying assumptions about the absorption problem. Using higher-order transport equations [77] to improve modeling accuracy or estimating source/detector and boundary condition coefficients numerically [28, 39, 40, 78] may also alleviate the need for baseline calibrations.



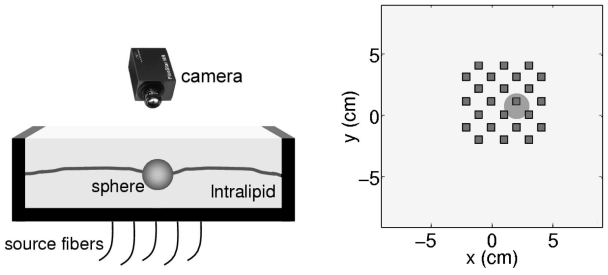
(a) Source/detector layout



(b) True image (c) 78.4 MHz

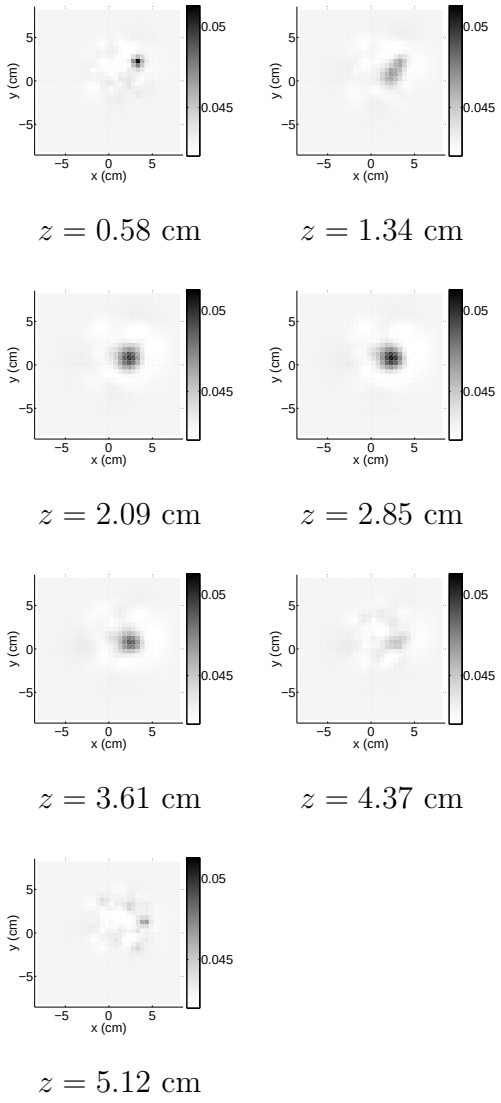
(d) 314 MHz (e) Both

Fig. 4.1. Reconstruction 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



(a) (b)

Fig. 4.2. (a) Schematic of the experimental setup, showing the box and 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.



$z = 0.58$ cm $z = 1.34$ cm

$z = 2.09$ cm $z = 2.85$ cm

$z = 3.61$ cm $z = 4.37$ cm

$z = 5.12$ cm

Fig. 4.3. Reconstruction of μ_{a_x} (cm^{-1}), obtained using 78, 314, and 627 MHz data

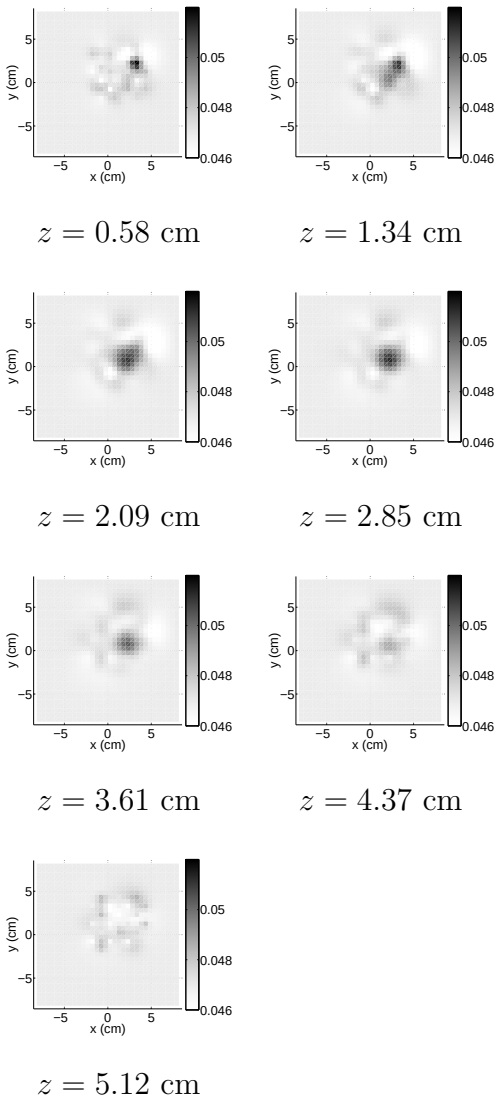
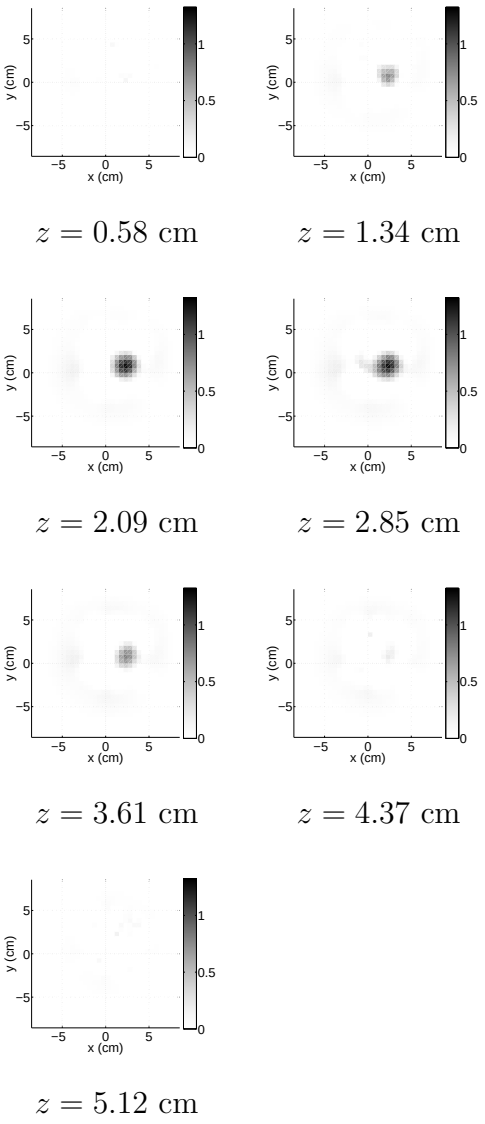


Fig. 4.4. Reconstruction of μ_{a_m} (cm^{-1}), obtained using 78, 314, and 627 MHz data



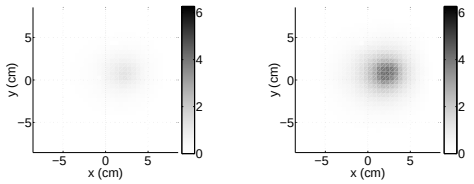
$z = 0.58$ cm $z = 1.34$ cm

$z = 2.09$ cm $z = 2.85$ cm

$z = 3.61$ cm $z = 4.37$ cm

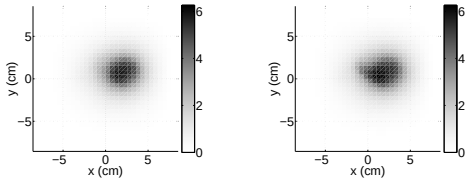
$z = 5.12$ cm

Fig. 4.5. Reconstruction of η (in 10^{-4} cm^{-1}), obtained using 78, 314, and 627 MHz data



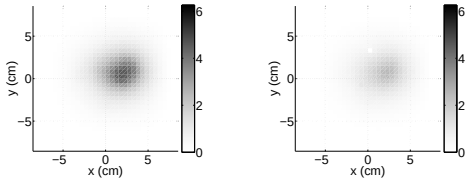
$z = 0.58$ cm

$z = 1.34$ cm



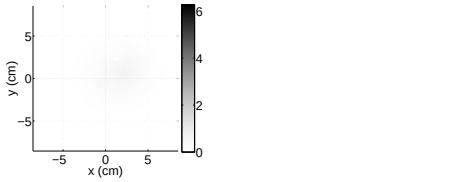
$z = 2.09$ cm

$z = 2.85$ cm



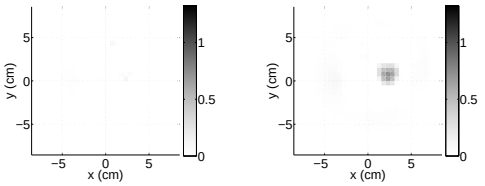
$z = 3.61$ cm

$z = 4.37$ cm



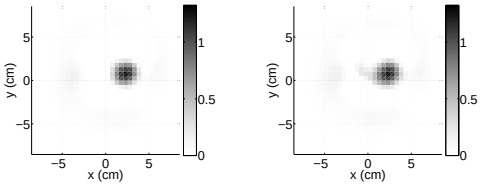
$z = 5.12$ cm

Fig. 4.6. Reconstruction of τ (in 10^{-10} s), obtained using 78, 314, and 627 MHz data



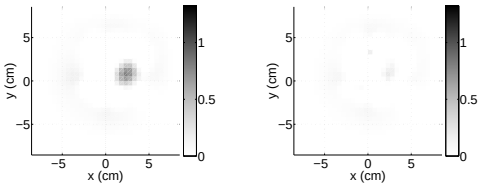
$z = 0.58$ cm

$z = 1.34$ cm



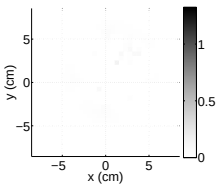
$z = 2.09$ cm

$z = 2.85$ cm



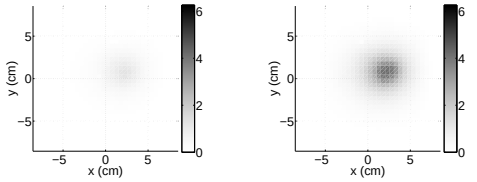
$z = 3.61$ cm

$z = 4.37$ cm



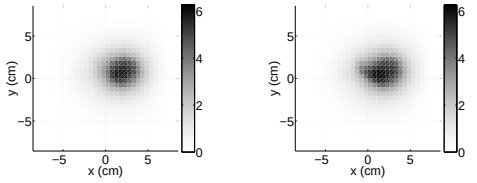
$z = 5.12$ cm

Fig. 4.7. Reconstruction of η (in 10^{-4} cm^{-1}), obtained using only 78 MHz data. The result is similar to the result obtained using multiple modulation frequencies.



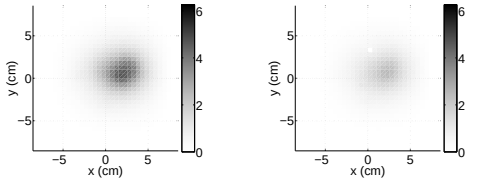
$z = 0.58$ cm

$z = 1.34$ cm



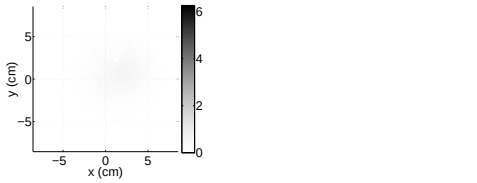
$z = 2.09$ cm

$z = 2.85$ cm



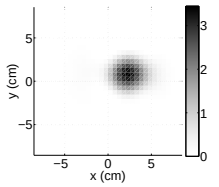
$z = 3.61$ cm

$z = 4.37$ cm

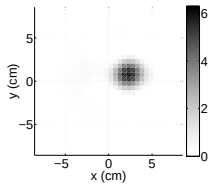


$z = 5.12$ cm

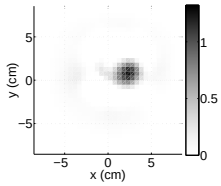
Fig. 4.8. Reconstruction of τ (in 10^{-10} s), obtained using only 78 MHz data. The result is similar to the result obtained using multiple modulation frequencies.



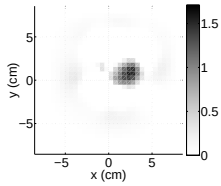
(a) $\sigma = 2.5 \times 10^{-6}$ cm^{-1}



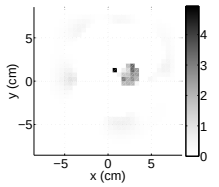
(b) $\sigma = 5.0 \times 10^{-6}$ cm^{-1}



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

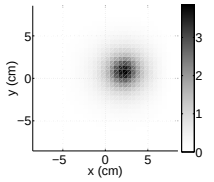


(d) $\sigma = 5.0 \times 10^{-5}$ cm^{-1}

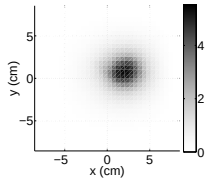


(e) $\sigma = 2.5 \times 10^{-4}$ cm^{-1}

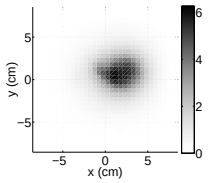
Fig. 4.9. Reconstructions of η (in 10^{-4} cm^{-1}) using various values of σ, showing a progression from overregularization to underregularization. The $z = 2.85$ cm cross sections are shown. The τ model used $\sigma = 1 \times 10^{-10}$ s in all cases.



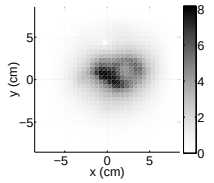
(a) $\sigma = 1 \times 10^{-11}$ s



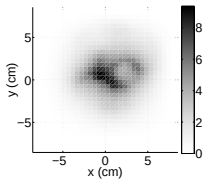
(b) $\sigma = 5 \times 10^{-11}$ s



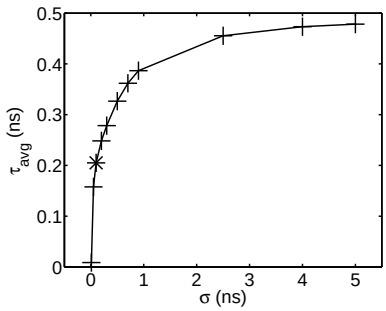
(c) $\sigma = 1 \times 10^{-10}$ s



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

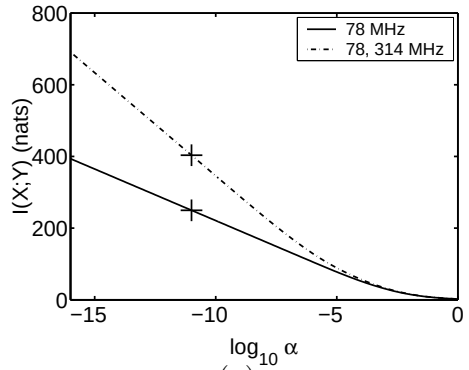


(e) $\sigma = 5 \times 10^{-10}$ s

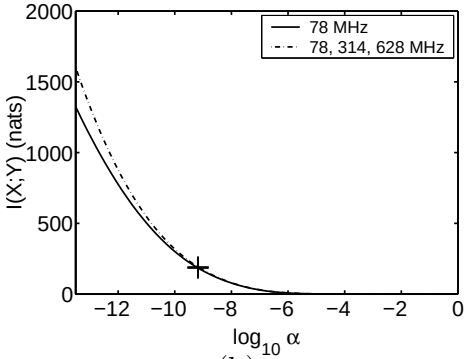


(f)

Fig. 4.10. (a)-(e) The $z = 2.85$ cm cross sections of reconstructed τ (in 10^{-10} s) for various σ, showing a progression from overregularization to underregularization. (f)$\hat{\tau}_{avg}$ as a function of σ for the τ reconstruction. The $\times$ symbol represents the value of σ which was used in generating the data of Figure 4.6. The η model used $\sigma = 2.5 \times 10^{-5}$ cm^{-1} in all cases.



(a)



(b)

Fig. 4.11. 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 results 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.

5. ESTIMATION OF KINETIC MODEL PARAMETERS IN FLUORESCENCE OPTICAL DIFFUSION TOMOGRAPHY

5.1 Background

As in positron emission tomography [79], the reconstruction of optical contrast agent kinetics can provide useful physiological information. Several groups of researchers have measured the dynamic behavior of injected optical contrast agents in animal or human subjects [12, 80–84]. Gurfinkel *et al.* [82] have used an intensified CCD camera to measure the pharmacokinetics of fluorescent agents in a canine with mammary tumors and fit the image sequence to a biexponential decay function which arises from a compartmental model. The study employed indocyanine green (ICG), which is believed to act as a nonselective blood pool agent, and carotene-conjugated 2-devinyl-2-(1-hexyloxyethyl) pyropheophorbide (HPPH-car), a photosensitizer which is believed to accumulate selectively in diseased tissue. A model parameter related to the dye's uptake rate showed significant contrast between diseased and surrounding tissue for HPPH-car, but not for ICG. Cuccia *et al.* [84] have measured the dynamics of two light-absorbing dyes, ICG and methylene blue (MB), in an adenocarcinoma rat tumor model by use of an optical probe with magnetic resonance imaging coregistration. Due to its small molecular weight of 373.9 Da, the MB temporal dynamics were dominated by blood flow effects. From the MB measurements, the authors observed variations in perfusion within the rat tumor. In contrast, ICG binds to albumin in the blood, with a resulting effective molecular weight of 66 kDa. Hence, ICG's temporal dynamics are dominated by the movement of albumin across the capillary membrane between the plasma and the extravascular,

extracellular space (EES). The authors used ICG dynamics to compute a physiologic parameter related to capillary permeability.

Previously, the time-varying absorption coefficient has been reconstructed in a cylindrical phantom [85] and in the human brain [86] by solving the inverse problem separately for each image in a time sequence. However, in some cases, the unknown image may not reasonably be considered constant over the instrument's measurement time. In addition, independent reconstruction of each image in the sequence ignores correlations in the image over time. Kolehmainen *et al.* have presented a state-estimation approach to the time-varying optical diffusion tomography problem which models the unknown image as a stochastic process governed by a stochastic difference equation [87]. This method solves the inverse problem by using extended Kalman filter and Kalman smoother techniques. The authors demonstrate their method with synthetic data from a two-dimensional phantom and, in a subsequent investigation, on real hemodynamic data from the human motor cortex [88]. This approach has shown promise for dynamic imaging problems where the time variation cannot be accurately parameterized by a known, deterministic model alone. However, in practical three-dimensional imaging problems, reconstructing a time sequence of images and updating large estimator covariance matrices may pose some difficulty, due to storage and computation requirements. Other dynamic imaging approaches have been investigated, including space/time regularization operators [89–91], principal components analysis [92], and temporal B-splines [93–95]. In many tracer experiments, a compartmental model [79,96] can accurately describe tracer kinetics by use of a system of first order differential equations. Previously, maximum likelihood approaches for direct reconstruction of kinetic model parameter images from PET data have been presented [97,98]. Recently, Kamasak *et al.* [99] have presented a Bayesian approach for dynamic PET which directly reconstructs images of the compartmental model's parameter images using all the data, while imposing spatial regularization. This approach results in substantially improved accuracy compared with previous

dynamic imaging methods which do not directly reconstruct the kinetic parameter images.

Here, we present a Bayesian, three-dimensional reconstruction approach for time-varying fluorescence optical diffusion tomography problems with nonlinear parameterizations of some known functional form. We demonstrate the method in a simulation study for the important case of a double exponential model, where the unknown parameters are the two amplitude coefficients and the two rate constants. This case can arise from a compartmental model in some applications, and it is similar to the behavior observed by Gurfinkel *et al.* [82] and Cuccia *et al.* [84]. The reconstruction approach is closely related to the methods of Kamasak *et al.* [99], in that it uses a statistical framework to directly reconstruct kinetic model parameters and a similar optimization scheme. We use all of the measured data to reconstruct the model parameter images directly, rather than reconstructing and storing a time sequence of fluorescence images. Our approach explicitly accounts for the fact that different sources are illuminated at different times. We also draw upon our recent work, in which we presented a nonlinear Bayesian inversion approach for the ODT/FODT problems and applied it to experimental data [27, 28, 41, 42]. We use parametric iterative coordinate descent (PICD) optimization [42, 50, 99], which is efficient and convenient for enforcing non-negativity constraints, and we use the generalized Gaussian Markov random field (GGMRF) prior model [49] for spatial regularization in the parameter images.

5.2 Models

Suppose the fluorescent yield $\eta(r, t)$ varies with time, on a scale comparable to the total acquisition time of the tomography instrument. The time variation might be a result of drug kinetics, which may be of physiological interest. Here, we consider the

case where $\eta(r,t)$ can be expressed as a (possibly nonlinear) function of U parameters which do not vary in time:

$$\eta(r,t) = \tilde{\eta}\left(\gamma_1(r), \cdots, \gamma_U(r), t\right), \tag{5.1}$$

where $\tilde{\eta}$ is a known function.

One important case which follows this framework is the compartmental model [96]. In a compartmental model, the body consists of a number of compartments, conceptual regions where the drug's concentration is assumed to be uniform. A system of differential equations describes the exchange of the drug among the different compartments. Previously [82, 84], the pharmacokinetics of ICG in animal subjects has been described by use of a three-compartment model, depicted in Figure 5.1. The compartments were the plasma, the tissue (the extracellular, extravascular space), and the kidneys and liver, which tend to clear the fluorophore out of the blood pool. Let $c_P(t)$ and $c_T(t)$ be the concentration of fluorophore in the plasma and tissue compartments, respectively. Cuccia *et al.* [84] initially assumed a biexponential decay model for $c_P(t)$, but ultimately observed only single exponential behavior in their experiment due to the relatively long elimination time of ICG compared with their measurement duration. Hence, we will assume single exponential decay for $c_P(t)$:

$$c_P = A\exp(-\kappa_3 t) \tag{5.2}$$

where A is taken to be the initial fluorophore concentration in the plasma and κ_3 is the rate constant for fluorophore elimination. We assume that the plasma input function does not vary throughout the imaging domain, *i.e.*, that $c_T(t)$ is not large enough to significantly effect the overall rate of elimination. A similar assumption was made by Cuccia *et al.* [84] We also let κ_1 and κ_2 be the rate constants for ICG entering and leaving the tissue. Then the concentrations are obtained by solving a differential equation for $c_T(t)$:

$$\frac{dc_T}{dt} = \kappa_1 c_P - \kappa_2 c_T. \tag{5.3}$$

To obtain a volumetric image, we solve (5.3) with initial condition $c_T(0) = 0$ to obtain $c_T(t)$ in each voxel:

$$c_T = \left(\frac{\kappa_1 A}{\kappa_2 - \kappa_3}\right)\left[\exp(-\kappa_3 t) - \exp(-\kappa_2 t)\right]. \tag{5.4}$$

The experimentally observed fluorescent yield is proportional to the concentration of fluorophore in the imaging domain. Within each voxel, the fluorophore concentration is some weighted sum of the tissue and blood compartments. Hence, we may write $\eta(r,t)$ as

$$\eta(r,t) = w_P(r)c_P(t) + w_T(r)c_T(r,t). \tag{5.5}$$

Substituting (5.2) and (5.4) into (5.5) yields the biexponential solution:

$$\eta(r,t) = \gamma_1(r)\exp[-\gamma_4(r)t] - \gamma_2(r)\exp[-\gamma_3(r)t] \tag{5.6}$$

where

$$\gamma_1 = A\left(w_P + \frac{w_T\kappa_1}{\kappa_2 - \kappa_3}\right) \tag{5.7}$$

$$\gamma_2 = A\left(\frac{w_T\kappa_1}{\kappa_2 - \kappa_3}\right) \tag{5.8}$$

$$\gamma_3 = \kappa_2 \tag{5.9}$$

$$\gamma_4 = \kappa_3. \tag{5.10}$$

From $\gamma_1, \ldots, \gamma_4$, it is possible to obtain the parameters κ_2, κ_3, (Aw_P), and $(Aw_T\kappa_1)$. Hence, we directly reconstruct images of the biexponential model parameters γ_1,γ_2, γ_3, and γ_4. To enforce the spatial independence of $c_P(t)$, one can constrain γ_4 to be the same everywhere. Alternatively, one can reconstruct $\gamma_4(r)$ to check the self-consistency of the model.

Strictly speaking, the time-dependence of $\eta(r,t)$ should correspond with a time-dependent perturbation in μ_{a_x}. In previous work, we have observed that perturbational changes in μ_{a_x} do not have a strong effect on reconstructed fluorescence [41]. Hence, for simplicity, we will not consider the reconstruction of time-varying μ_{a_x} here.

5.2.1 Tomography Problem

Previously, for the stationary case, we have shown how to reconstruct $\mu_{a_x}(r)$, $\mu_{a_m}(r)$, $D_x(r)$, $D_m(r)$, $\tau(r)$, and $\eta(r)$ [41, 42]. Here, we assume that $\mu_{a_x}(r)$, $\mu_{a_m}(r)$, $D_x(r)$, and D_m(r) are known in advance and do not vary with time, and we consider the problem of reconstructing $\tau(r)$ and $\eta(r,t)$ in the time-varying case.

Suppose that measurements are recorded at C measurement times, which we call $t_1, \cdots, t_C$. At each measurement time, measurements are recorded with one or more sources at wavelength λ_x and detectors filtered at λ_m. Figure 5.2 schematically depicts the measurement, with a source and an array of detectors arranged around the domain at each time. Note that the source and detector geometry may be different at different time indices. In particular, practical instruments often sequentially illuminate sources one at a time, while all detectors are used simultaneously at all times, and all are fixed in space.

Consider a domain discretized into N volume elements, or voxels. Let r_i denote the position of the i^{th} voxel centroid. Assuming that $\eta(r,t)$ can be expressed using (5.1), we define the image vector x:

$$x = \left[\begin{array}{cccc} x_{(0)}^T & x_{(1)}^T & \cdots & x_{(U)}^T \end{array} \right]^T, \tag{5.11}$$

where

$$x_{(u)}^T = \left[\begin{array}{ccc} x_{(u),1} & \cdots & x_{(u),N} \end{array} \right]^T, \tag{5.12}$$

$$x_{(0)}^T = \left[\begin{array}{ccc} \tau(r_1) & \cdots & \tau(r_N) \end{array} \right]^T, \tag{5.13}$$

and, for $1 \leq u \leq U$,

$$x_{(u)}^T = \left[\begin{array}{ccc} \gamma_u(r_1) & \cdots & \gamma_u(r_N) \end{array} \right]^T, \tag{5.14}$$

with the superscript T denoting the transpose operation. Note that x is of size $(U+1) \times N$, consisting of $U+1$ concatenated parameter vectors of size N. In the parameterization of (5.6), $U = 4$.

5.3 Inverse Problem

5.3.1 Bayesian Framework

Let y denote the measurement vector whose ordering will be precisely specified in Section 5.3.2. Similarly, let $f(x)$ denote the forward model. As in previous Chapters, we address the ill-posed problem of estimating x from y in a Bayesian framework. We impose positivity constraints for x, and also require that $\eta(r_i, t) \geq 0$ for all r_i and t. For $p_{Y|X}(y|x)$, we once again use an independent Gaussian distribution derived from a shot noise model [24]. The prior model $p_X(x)$ is the Gaussian Markov random field (GGMRF) model [24, 49]. We use upper case to represent the corresponding random variables, and we assume that $X_{(0)}, \cdots X_{(U)}$ are independent:

$$p_X(x) = \prod_{u=0}^{U} p_{X_{(u)}}(x_{(u)}) \tag{5.15}$$

$$= \prod_{u=0}^{U} \frac{1}{\sigma_{(u)}^N \zeta(\rho_{(u)})} \exp\left(-\frac{1}{\rho_{(u)} \sigma_{(u)}^{\rho_{(u)}}} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_{(u),i} - x_{(u),j} \right|^{\rho_{(u)}} \right) \tag{5.16}$$

where the u subscripts correspond to (5.12), $\mathcal{N}$ consists of all pairs of neighboring (adjacent) nodes in a 26-neighbor system, and b_{i-j} is the weighting coefficient corresponding to the i^{th} and j^{th} nodes. As in the previous Chapters, the coefficients b_{i-j} are assigned to be inversely proportional to the node separation in a cube-shaped node layout, where $\sum_j b_{i-j} = 1$.

We incorporate α into the inverse problem as an unknown instrument parameter, as we have found that this tends to improve the robustness and speed of convergence [25]:

$$\hat{x} = \arg \max_{x \geq 0, \alpha \geq 0} \{ \, p_{X|Y}(x|y, \alpha) \, \}. \tag{5.17}$$

We form the log posterior probability $l(x)$ [25]:

$$\begin{aligned} l(x) &= -P \log ||y - f(x)||_{\Lambda}^2 \\ &\quad - \sum_{u=0}^{U} \left(\frac{1}{\rho_{(u)} \sigma_{(u)}^{\rho_{(u)}}} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_{(u),i} - x_{(u),j} \right|^{\rho_{(u)}} \right) \end{aligned} \tag{5.18}$$

and implement its maximization by alternating closed form updates of $\hat{\alpha}$ with updates of $\hat{x}$:

$$\hat{\alpha} \leftarrow \frac{1}{P} ||y - f(\hat{x})||_{\Lambda}^2 \tag{5.19}$$

$$\hat{x} \leftarrow \arg \underset{x \geq 0}{\text{update}} \left\{ \log p_{Y|X}(y|x, \hat{\alpha}) + \log p_X(x) \right\}, \tag{5.20}$$

where "$\leftarrow$" denotes assignment and "arg update" denotes an iteration of some optimizer. The update in (5.20) is equivalent to reducing a cost function

$$\begin{aligned} c(x, \hat{\alpha}) &= \frac{1}{\hat{\alpha}} ||y - f(x)||_{\Lambda}^2 \\ &\quad + \sum_{u=0}^{U} \left(\frac{1}{\rho_{(u)} \sigma_{(u)}^{\rho_{(u)}}} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_{(u),i} - x_{(u),j} \right|^{\rho_{(u)}} \right). \end{aligned} \tag{5.21}$$

5.3.2 Definitions

Define $s_{t_c,k}$ as the location of the k^{th} source at time t_c, and $d_{t_c,m'}$ as the location of the m'^{th} detector at time t_c, and let $g_x(s_{t_c,k}, d_{t_c,m'}; \omega)$ and $g_m(s_{t_c,k}, d_{t_c,m'}; \omega)$ be the diffusion equation Green's functions for wavelength λ_x, and λ_m, respectively. Also, let $\phi_f(s_{t_c,k}, d_{t_c,m'}; \omega, t_c, x)$ be the fluorescence observed at observation position $d_{t_c,m'}$ for an excitation source at $s_{t_c,k}$, where

$$\begin{aligned} &\phi_f(s_{t_c,k}, d_{t_c,m'}; \omega, t_c, x) = \\ &\quad \int \eta(r, t_c) \tfrac{1 - j\omega\tau(r)}{1 + [\omega\tau(r)]^2} g_x(s_{t_c,k}, r; \omega) g_m(r, d_{t_c,m'}; \omega) d^3 r. \end{aligned} \tag{5.22}$$

Suppose that at time t_c we have K_c sources and M_c detectors at a modulation frequency of ω. (Typically, $K_c = 1$ for most systems which illuminate sources sequen-

tially.) Let $f_{\omega,t_c}(x)$ be the forward model for the data taken at t_c with $e^{j\omega t}$-modulated light. Then,

$$f_{\omega,t_c}(x_f) = \begin{bmatrix} \phi_f(s_{t_c,1}, d_{t_c,1}; \omega, t_c, x) \\ \phi_f(s_{t_c,1}, d_{t_c,2}; \omega, t_c, x) \\ \vdots \\ \phi_f(s_{t_c,1}, d_{t_c,M_c}; \omega, t_c, x) \\ \phi_f(s_{t_c,2}, d_{t_c,1}; \omega, t_c, x) \\ \vdots \\ \phi_f(s_{t_c,K_c}, d_{t_c,M_c}; \omega, t_c, x) \end{bmatrix}. \tag{5.23}$$

Let Q be the number of modulation frequencies used, and C be the number of measurement times times. Then:

$$f_{t_c}(x) = \left[f_{\omega_1,t_c}(x)^T, f_{\omega_2,t_c}(x)^T \cdots f_{\omega_Q,t_c}(x)^T \right]^T \tag{5.24}$$

$$f(x) = \left[f_{t_1}(x)^T, f_{t_2}(x)^T \cdots f_{t_C}(x)^T \right]^T. \tag{5.25}$$

Similarly, we define the measurement vector y as:

$$y_{t_c} = \left[y^T_{\omega_1,t_c}, y^T_{\omega_2,t_c} \cdots y^T_{\omega_Q,t_c} \right]^T \tag{5.26}$$

$$y = \left[y^T_{t_1}, y^T_{t_2} \cdots y^T_{t_C} \right]^T \tag{5.27}$$

corresponding to the same order used in (5.25). Note that $g(s_k, d_{m'}, \omega) = g(d_{m'}, s_k, \omega)$ at λ_x and at λ_m, due to reciprocity [74].

Using matrices to approximate the integration of (5.22), we write

$$G^x(\omega, t_c) = \begin{bmatrix} g_x(s_{t_c,1}, r_1; \omega) & \cdots & g_x(s_{t_c,1}, r_N; \omega) \\ \vdots & \ddots & \vdots \\ g_x(s_{t_c,K}, r_1; \omega) & \cdots & g_x(s_{t_c,K}, r_N; \omega) \end{bmatrix} \tag{5.28}$$

$$G^m(\omega, t_c) = \begin{bmatrix} g_m(d_{t_c,1}, r_1; \omega) & \cdots & g_m(d_{t_c,1}, r_N; \omega) \\ \vdots & \ddots & \vdots \\ g_m(d_{t_c,M}, r_1; \omega) & \cdots & g_m(d_{t_c,M}, r_N; \omega) \end{bmatrix}. \tag{5.29}$$

We also define J_{ω,t_c} as

$$J_{\omega,t_c} = V \cdot \begin{bmatrix} G^x_{1,1}(\omega,t_c)G^m_{1,1}(\omega,t_c) & \cdots & G^x_{1,N}(\omega,t_c)G^m_{1,N}(\omega,t_c) \\ \vdots & \ddots & \vdots \\ G^x_{1,1}(\omega,t_c)G^m_{M,1}(\omega,t_c) & \cdots & G^x_{1,N}(\omega,t_c)G^m_{M,N}(\omega,t_c) \\ G^x_{2,1}(\omega,t_c)G^m_{1,1}(\omega,t_c) & \cdots & G^x_{2,N}(\omega,t_c)G^m_{1,N}(\omega,t_c) \\ \vdots & \ddots & \vdots \\ G^x_{K,1}(\omega,t_c)G^m_{M,1}(\omega,t_c) & \cdots & G^x_{K,N}(\omega,t_c)G^m_{M,N}(\omega,t_c) \end{bmatrix} \tag{5.30}$$

where V is the volume of a voxel. Let

$$h(x_{(*),i},\omega,t) = \eta(r_i,t)\frac{1-j\omega\tau(r_i)}{1+[\omega\tau(r_i)]^2} \tag{5.31}$$

$$h_{\omega,t_c}(x) = \begin{bmatrix} h(x_{(*),1},\omega,t_c) & \cdots & h(x_{(*),N},\omega,t_c) \end{bmatrix}^T \tag{5.32}$$

Then

$$f_{\omega,t_c}(x) = J_{\omega,t_c}h_{\omega,t_c}(x) \tag{5.33}$$

if we ignore discretization error. Therefore, (5.21) is equivalent to:

$$\begin{aligned} c(x,\hat{\alpha}) = \frac{1}{\hat{\alpha}}\sum_{c=1}^{C}\sum_{q=1}^{Q}\left|\left|y_{\omega_q,t_c} - J_{\omega_q,t_c}h_{\omega_q,t_c}(x)\right|\right|^2_{\Lambda_{\omega_q,t_c}} \\ +\sum_{u=0}^{U}\frac{1}{\rho_{(u)}\sigma^{\rho_{(u)}}_{(u)}}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\left|x_{(u),i}-x_{(u),j}\right|^{\rho_{(u)}}. \end{aligned} \tag{5.34}$$

The cost function in (5.34) is used in our image reconstruction.

5.3.3 Parametric Iterative Coordinate Descent

To optimize (5.34), we use an algorithm which we call parametric iterative coordinate descent (PICD). It is based on earlier work [24], and it is modified to allow for computationally efficient updates of the kinetic model parameters. The voxels are individually updated in random order by optimizing the cost function with respect

to the parameters at each voxel position. The updates enforce the constraints $x \geq 0$, and also $\gamma_1 \geq \gamma_2$ and $\gamma_3 \geq \gamma_4$, which are necessary and sufficient to ensure that each parameter is nonnegative, and that $\eta(t) \geq 0$ for all time.

In one update scan for $\hat{x}$, all of the unknowns $x_{(u)}$, $u = 0, \cdots 4$ are updated at all N voxel positions. Let the scalar $x_{(u),i}$ denote the i^{th} element of $x_{(u)}$. With all other image elements fixed, the PICD update for the estimate $\hat{x}_{(u),i}$ is given by

$$\begin{aligned} \hat{x}_{(u),i} \quad \leftarrow \quad & \arg \min_{x_{(u),i} \geq 0} \left\{ \frac{1}{\hat{\alpha}} \sum_{c=1}^{C} \sum_{q=1}^{Q} \left|\left| y_{\omega_q,t_c} - \left[J_{\omega_q,t_c}\right]_{*(i)} h(x_{(*),i}, \omega_q, t_c) \right|\right|^2_{\Lambda_{\omega_q,t_c}} \right. \\ & \left. + \frac{1}{\rho_{(u)} \sigma_{(u)}{}^{\rho_{(u)}}} \sum_{j \in \mathcal{N}_i} b_{i-j} \left| x_{(u),i} - \hat{x}_{(u),j} \right|^{\rho_{(u)}} \right\}, \end{aligned} \tag{5.35}$$

where $\mathcal{N}_i$ is the set of nodes neighboring node i and $\rho_{(u)}$ and $\sigma_{(u)}$ are the prior model parameters for $X_{(u)}$. In (5.35), $\left[J_{\omega_q,t_c}\right]_{*(i)}$ denotes the i^{th} column of J_{ω_q,t_c}. Suppose we have an initial guess $\tilde{x}$, and let $z_{\omega_q,t_c} = y_{\omega_q,t_c} - f_{\omega_q,t_c}(\tilde{x})$. Then, (5.35) is equivalent to

$$\begin{aligned} \hat{x}_{(u),i} \quad \leftarrow \quad & \arg \min_{x_{(u),i} \geq 0} \{ \\ & \frac{1}{\hat{\alpha}} \sum_{c=1}^{C} \sum_{q=1}^{Q} ||z_{\omega_q,t_c} - [J_{\omega_q,t_c}]_{*(i)} [h(x_{(*),i}, \omega_q, t_c) \\ & - h(\tilde{x}_{(*),i}, \omega_q, t_c)]||^2_{\Lambda_{\omega_q,t_c}} \\ & + \frac{1}{\rho_{(u)} \sigma_{(u)}{}^{\rho_{(u)}}} \sum_{j \in \mathcal{N}_i} b_{i-j} \left| x_{(u),i} - \hat{x}_{(u),j} \right|^{\rho_{(u)}} \Bigg\} , \\ = \quad & \arg \min_{x_{(u),i} \geq 0} \{ \\ & \frac{1}{\hat{\alpha}} \sum_{c=1}^{C} \sum_{q=1}^{Q} \Big(\theta_{1,\omega_q,t_c} [h(x_{(*),i}, \omega_q, t_c) - h(\tilde{x}_{(*),i}, \omega_q, t_c)] \\ & + \frac{\theta_{2,\omega_q,t_c}}{2} [h(x_{(*),i}, \omega_q, t_c) - h(\tilde{x}_{(*),i}, \omega_q, t_c)]^2 \Big) \\ & + \frac{1}{\rho_{(u)} \sigma_{(u)}{}^{\rho_{(u)}}} \sum_{j \in \mathcal{N}_i} b_{i-j} \left| x_{(u),i} - \tilde{x}_{(u),j} \right|^{\rho} \Bigg\}, \end{aligned} \tag{5.36}$$

where

$$\theta_{1,\omega_q,t_c} = -2\mathrm{Re}\left\{[J_{\omega_q,t_c}]_{*i}^H \Lambda_{\omega_q,t_c} z_{\omega_q,t_c}\right\} \quad (5.37)$$

$$\theta_{2,\omega_r,t_c} = 2[J_{\omega_q,t_c}]_{*i}^H \Lambda_{\omega_q,t_c} [J_{\omega_q,t_c}]_{*i}. \quad (5.38)$$

In (5.36), θ_{1,ω_q,t_c} and θ_{2,ω_q,t_c} are not functions of $x_{(u),i}$, and thus do not need to be recomputed during the nonlinear, one-dimensional (1-D) line search over $x_{(u),i}$. This property enables significant computational savings, as repeated computations of θ_{1,ω_q,t_c} and θ_{2,ω_q,t_c} would require numerous complex multiplications. We perform the minimization over $x_{(u),i}$ by use of a Golden Section search [75].

To enforce the constraints $\gamma_1 \geq \gamma_2$, we initially perform minimizations over $x_{(1),i}$ and over $x_{(2),i}$ and observe whether the inequality constraint is satisfied. If $\hat{x}_{(2),i} > \hat{x}_{(1),i}$, we perform a new line search enforcing $x_{(1),i} = x_{(2),i}$:

$$\begin{aligned}
\hat{x}_{(1),i}, \hat{x}_{(2),i} \leftarrow \arg \min_{x_{(1),i}=x_{(2),i}\geq 0} \{ & \\
& \frac{1}{\hat{\alpha}} \sum_{c=1}^{C} \sum_{q=1}^{Q} \Big(\theta_{1,\omega_q,t_c} [h(x_{(*),i},\omega_q,t_c) - h(\tilde{x}_{(*),i},\omega_q,t_c)] \\
& + \frac{\theta_{2,\omega_q,t_c}}{2} [h(x_{(*),i},\omega_q,t_c) - h(\tilde{x}_{(*),i},\omega_q,t_c)]^2 \Big) \\
& + \frac{1}{\rho_{(1)} {\sigma_{(1)}}^{\rho_{(1)}}} \sum_{j\in\mathcal{N}_i} b_{i-j} \left| x_{(1),i} - \tilde{x}_{(1),j} \right|^{\rho_{(1)}} \\
& + \frac{1}{\rho_{(2)} {\sigma_{(2)}}^{\rho_{(2)}}} \sum_{j\in\mathcal{N}_i} b_{i-j} \left| x_{(2),i} - \tilde{x}_{(2),j} \right|^{\rho_{(2)}} \Bigg\}
\end{aligned} \quad (5.39)$$

A similar procedure is used to enforce the $\gamma_3 \geq \gamma_4$ condition.

We implement the joint estimation of α and x iteratively. One iteration consists of a closed form update of $\hat{\alpha}$ using (5.19), followed by a PICD scan to update $\hat{x}$. Appendix E provides pseudocode for a more detailed specification of the PICD algorithm.

5.4 Simulation

To validate the method, we performed a simulation study. A synthetic time series of data was generated from a cube-shaped phantom containing two heterogeneities. The background properties were $\mu_{a_x} = \mu_{a_m} = 0.047$ cm^{-1} and $D_x = D_m = 0.027$ cm. The heterogeneities had the same τ, but different γ_1, γ_2, and γ_3. The parameter γ_4 was 0, and was not reconstructed. This corresponds to an assumption that elimination time of fluorophore from the plasma is long compared with the measurement time, which is reasonable in cases where we are most interested in the initial drug uptake behavior. The parameter values were selected to result in uptake behavior on the order of seconds, which may be reasonable for a small animal imaging experiment [84].

The phantom was 8 cm $\times$ 8 cm $\times$ 5.7 cm in size, and it was discretized into $33 \times 33 \times 17$ voxels of size 0.26 cm $\times$ 0.26 cm $\times$ 0.38 cm. To generate the synthetic measurements, the diffusion equation was solved numerically using multigrid finite differences [51] and extrapolated zero-flux boundary conditions with interpolated source positions as we have described previously [28]. The simulation used a modulation frequency of 78.4 MHz. Figure 5.3 shows the locations of the sources which were placed on the bottom face of the cube-faced domain. The same positions were used as detector positions on the top face of the domain, simulating a parallel-plate transmission geometry similar to that which has been used for optical mammography previously. The sources were illuminated one at a time at different times, in the order shown in Figure 5.3, and the data consisted of one complete pass through all of the sources, with 441 measurements in all. Simulated shot noise was added, giving an average signal/noise ratio of 28 dB for all the data. The true phantom is shown in Figure 5.4, with cross section images through each of the heterogeneities.

For the reconstructions, the hyperparameter ρ was set to 2, corresponding to the Gaussian Markov random field (GMRF) model. Although automatic estimation of the hyperparameters is possible in principle [53], σ was chosen to empirically give

the best results, as we have done previously [42]. For this problem the best results were given by $\sigma_{(0)} = 2.75 \times 10^{-10}$, $\sigma_{(1)} = 0.5$, $\sigma_{(2)} = 0.5$, and $\sigma_{(3)} = 0.0125$ (where the units of lifetime are seconds and the units of fluorescence are cm^{-1}). The PICD algorithm was run to 50 iterations, which required approximately 2 hours on a 2 GHz AMD Athlon workstation.

The reconstructed parametric images are shown in Fig. 5.5. The results are accurate, although shadowing effects are apparent in the images. In particular, γ_3, which is related to a dye's uptake rate, was reconstructed accurately, enabling a clear distinction between the two objects. Figure 5.6(a)-(d) shows the true images of $\eta(r,t)$ for the two objects, at 4 different times, and Figure 5.6(e) shows plots of $\eta(t)$ for a single point near the center of each object. The reconstructed time variation is also accurate, without shadowing artifacts.

The reconstruction $\hat{\eta}(r,t)$ was obtained by substituting $\hat{\gamma}_1(r)$, $\hat{\gamma}_2(r)$, and $\hat{\gamma}_3(r)$ into (5.6). The results shown in Figure 5.7 indicate that all features are nicely captured. Figure 5.8 shows a convergence plot showing monotone decrease of the cost function versus iteration number. For comparison, we also reconstructed $\eta(r,t)$ by independently reconstructing $\eta(r,t_i)$ at each measurement time t_i, using our previous FODT reconstruction algorithm [42]. For the first simulation, we used the same 441 measurements that was used for the results in Figure 5.9. The reconstructions all used $\rho = 2$, with $\sigma = 0.5$ for η and $\sigma = 2.75 \times 10^{-10}$ for τ (which gave the best empirical results). We performed 21 reconstructions of $\eta(r,t_i)$, using a single source and 21 detectors for each. The results, shown in Figure 5.9, have poor accuracy. For the second simulation, we greatly increased the number of data, using all 21 sources and 21 detectors for each of the reconstructions of $\eta(r,t_i)$ (*i.e.*, 9261 measurements, with 441 measurements used at each time index). The reconstructions in this simulation used $\rho = 2$, with $\sigma = 0.375$ for η and $\sigma = 2.75 \times 10^{-10}$ for τ. The results are shown in Figure 5.10. With this 21-fold increase in data, the reconstructions accuracy approaches that of the parametric imaging method.

5.5 Conclusions

We have presented a method for parametric reconstruction of fluorescent drug kinetics by use of fluorescence optical diffusion tomography. The simulation showed that two heterogeneities with different time-dependent behavior could be reconstructed simultaneously, and clearly distinguished based on uptake-related parameters. In principle, receptor-targeted fluorescent probes may have a significantly faster uptake rate in tumors compared with the surrounding tissue. Hence, the ability to reconstruct the drug uptake kinetics could facilitate tumor imaging with high contrast, compared with methods which do not make full use of the drug dynamics.

The presented approach is flexible, and may be applied to more sophisticated compartmental models. In principle, more complicated kinetic models which incorporate additional compartments or nonlinear saturation effects may be used in the same framework, as they simply increase the complexity of the single-site updates. In addition, the PICD algorithm may be incorporated into a multigrid framework [35] to improve convergence properties for a wide variety of images.

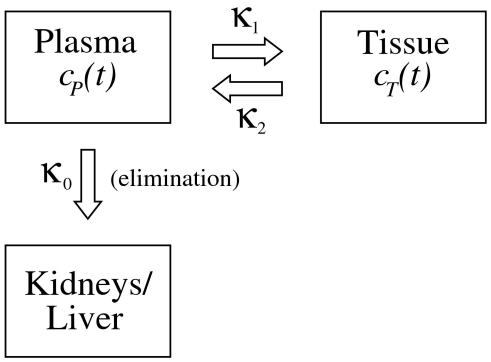


Fig. 5.1. Compartmental model describing the exchange of contrast agent between the tissue and the plasma.

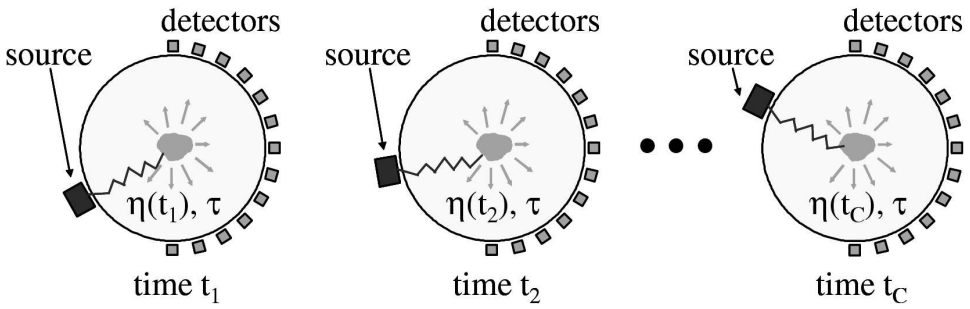


Fig. 5.2. Measurement approach for reconstructing $\eta(t)$ and τ. Note that the measurement geometry may differ at each time.

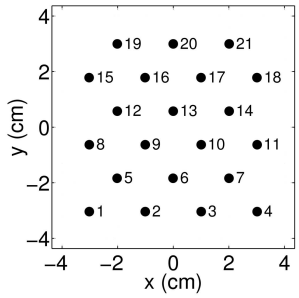
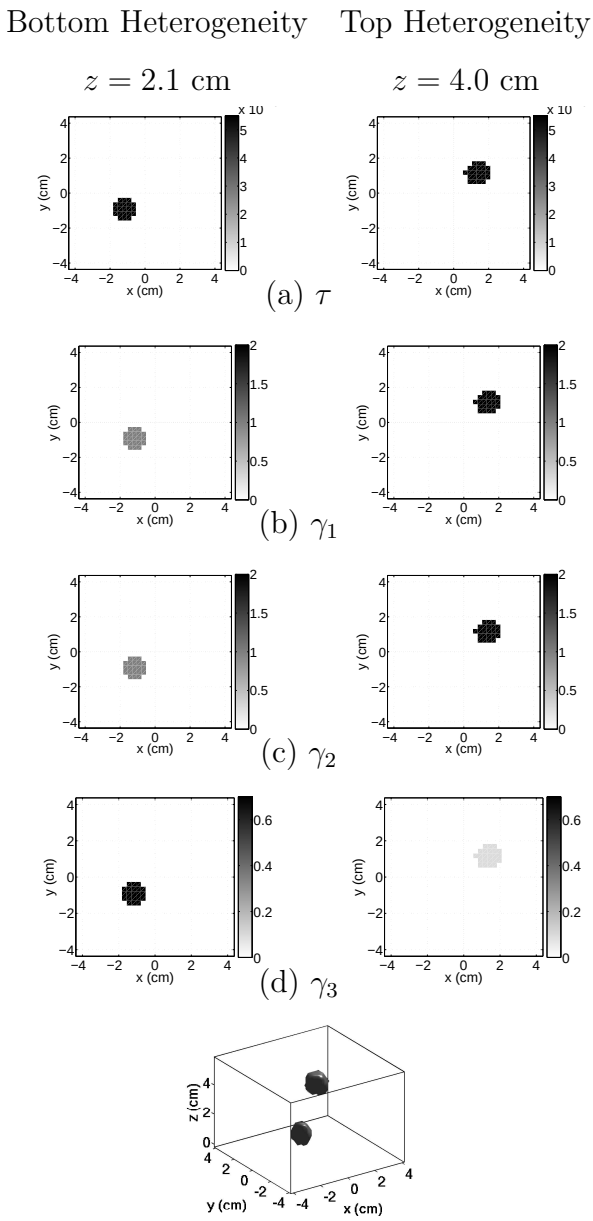


Fig. 5.3. Source and detector locations used in the simulations. The sources were on the bottom face of the cube-shaped phantom, while the detectors were on the top. The sources were illuminated in the order shown, with one source used for each time frame.



(a) τ

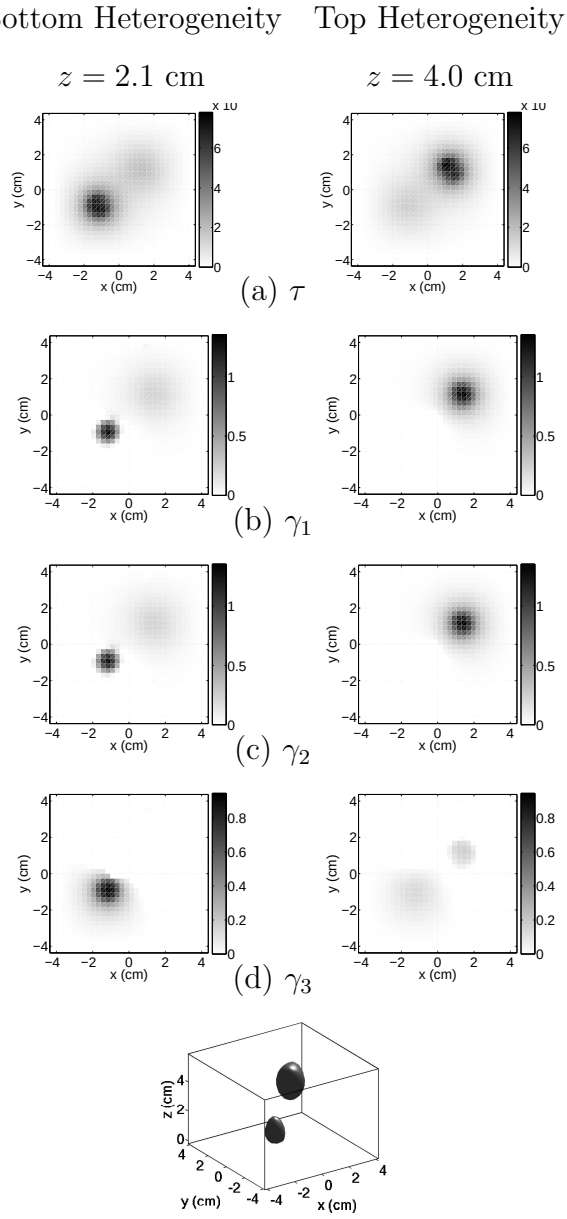
(b) γ_1

(c) γ_2

(d) γ_3

(e) γ_1 Isosurface

Fig. 5.4. True parameter images describing the time-varying fluorescence in simulation study. Cross sections are shown through the top heterogeneity and the bottom heterogeneity. Note the parameter γ_3, which indicates different uptake rates in the two heterogeneities. In (e), an isosurface of the γ_1 reconstruction is shown, contoured at 1/3 the maximum value.



(a) τ

(b) γ_1

(c) γ_2

(d) γ_3

(e) γ_1 Isosurface

Fig. 5.5. Reconstructed parameter images describing the time-varying fluorescence in the simulation study. In (e), an isosurface of the γ_1 reconstruction is shown, contoured at 1/3 the maximum value.

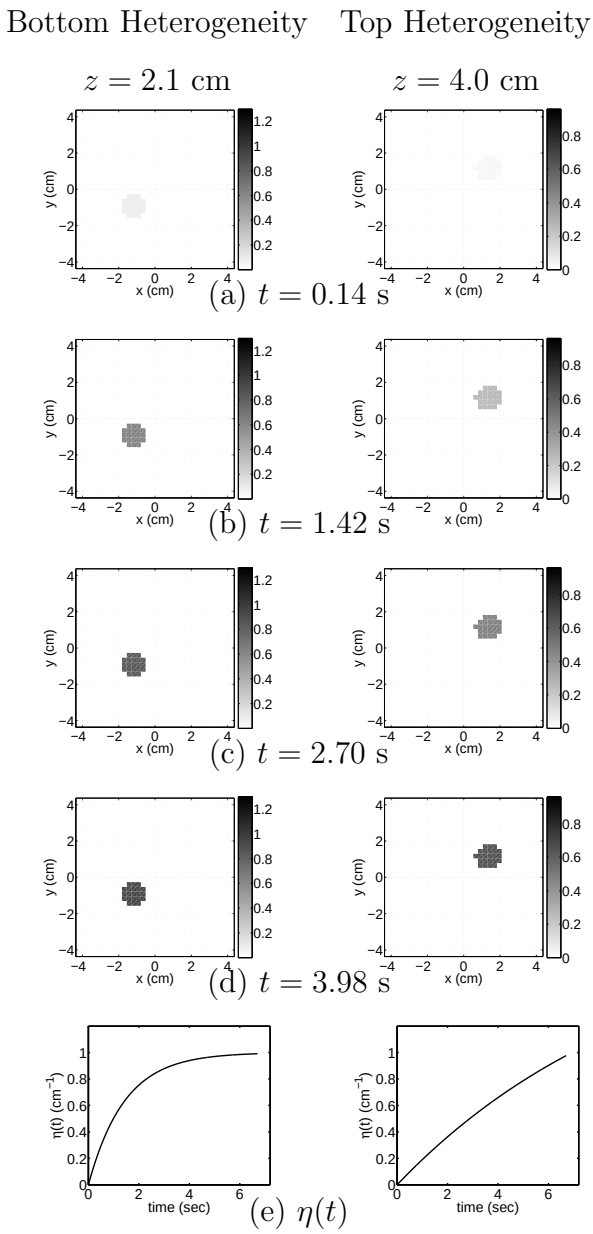


Fig. 5.6. (a)-(d) True fluorescence versus time. (e) $\eta(t)$, for a sample point within each heterogeneity.

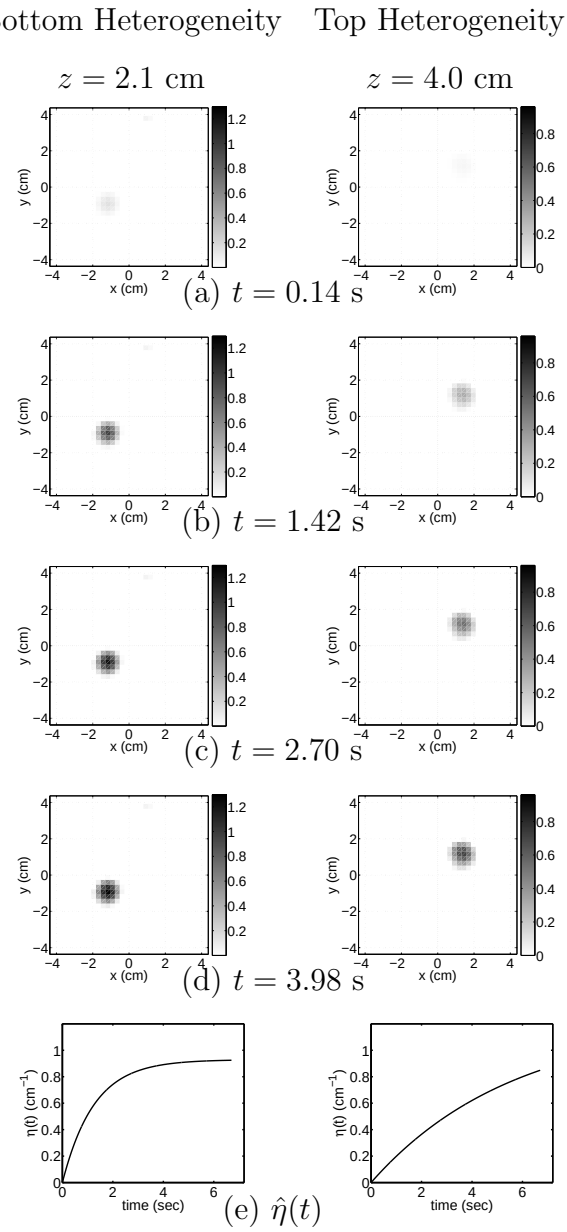


Fig. 5.7. (a)-(d) Fluorescence versus time, reconstructed by parametric ICD method. (e) $\hat{\eta}(t)$, for a sample point within each heterogeneity.

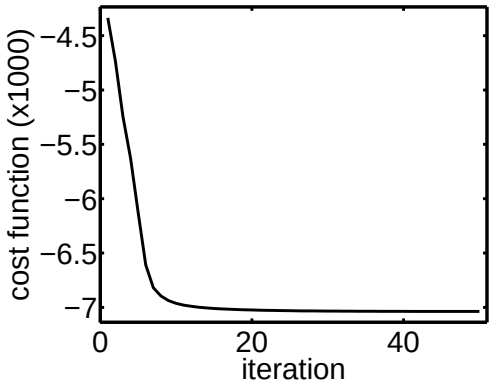


Fig. 5.8. Convergence for PICD algorithm in simulation study.

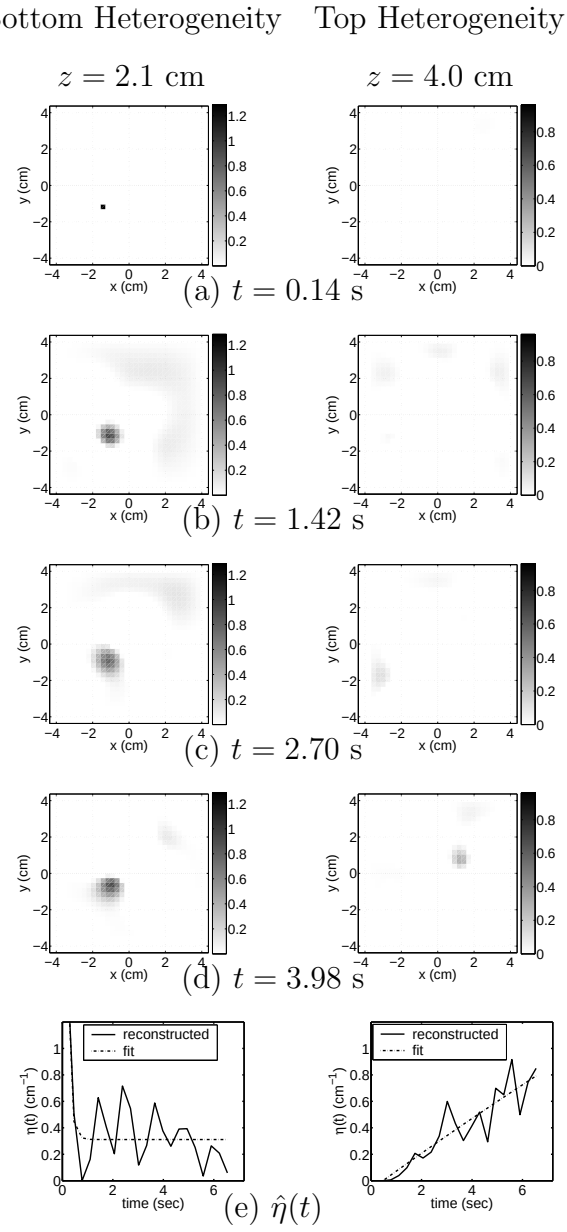


Fig. 5.9. (a)-(d) Fluorescence versus time, reconstructed independently at each time frame, using the same data as the parametric reconstructions. (e) $\hat{\eta}(t)$, for a sample point within each heterogeneity.

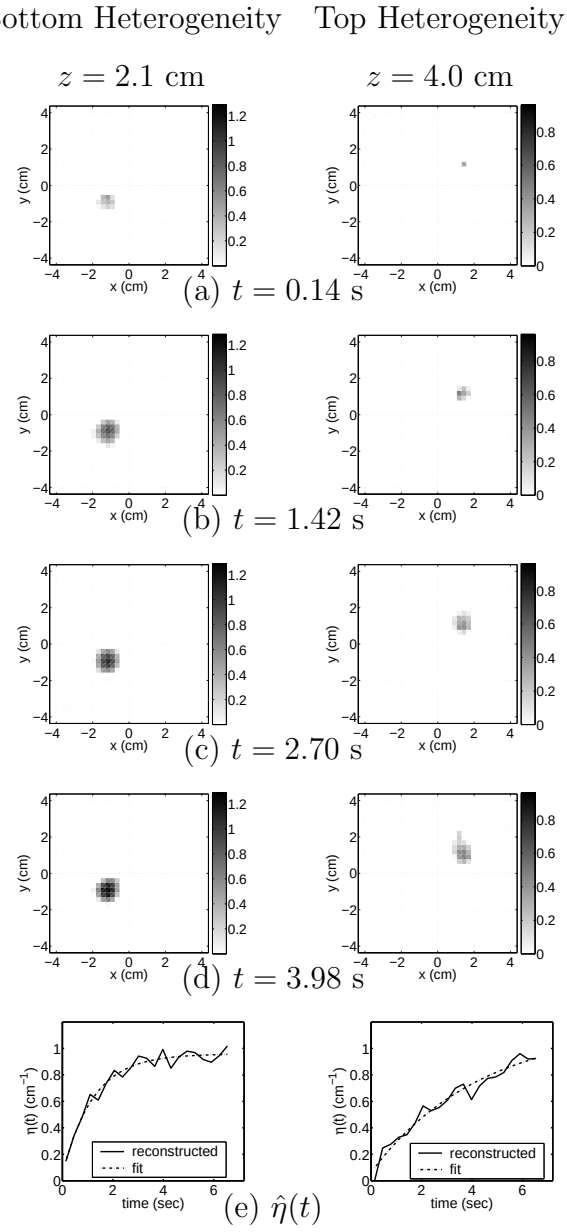


Fig. 5.10. (a)-(d) Fluorescence versus time, reconstructed independently at each time frame, using a 21-fold increase in data over those used in the parametric reconstructions. (e) $\hat{\eta}(t)$, for a sample point within each heterogeneity.

6. DETECTION AND LOCALIZATION OF A FLUORESCING MOUSE TUMOR IN A TURBID MEDIUM

6.1 Background

Currently, the only assured cure for the vast majority of cancers involves the complete resection of all malignant lesions. Achievement of this objective, however, is often limited by the inability of the surgeon to identify and localize all cancerous tissues. In some cases, neoplastic loci are difficult to distinguish from adjacent normal tissue. In other situations, they are not detected by the surgeon due to their small size or obscured location. Even after removal of visible tumor masses during surgical debulking, microscopic tumors not visible to the naked eye can often be the source of recurrent disease [100].

Recent advances in biomedical optical imaging based on fluorescent dyes [101] offer great promise in diagnosing malignant tissue, monitoring tumor therapy progress, and guiding surgical intervention for tumor removal. Optical imaging is safe and relatively inexpensive compared to other modalities, and optical measurement probes have the potential to be small, maneuverable, and unobtrusive. These properties may be particularly attractive in an intraoperative environment where real-time imaging may be required over a period of hours.

Previously, the problem of localizing fluorophores in real animal tissue, tissue-simulating phantoms, or computational simulations has been considered in a variety of studies. Chen *et al.* [102] have used a near-infrared (near-IR) measurement system with dual interfering sources to determine the 2-D location of a fluorescing tumor in a mouse subject. The mouse was injected with a contrast agent that selectively

targeted the tumor due to its increased metabolic activity. Hull *et al.* [103] have accurately determined the depth of a small fluorescent sphere embedded within a turbid medium by use of fluorescence measurements on the surface. The sample was illuminated from the side with an expanded laser beam, and the fluorescence was collected from the top surface with a linear array of detector fibers. Pfister and Scholz [104] have used a multiple signal classification algorithm to localize fluorescent spots under tissuelike scatter in a computational simulation. Fluorescence optical diffusion tomography (FODT), in which full volumetric images of fluorescence are reconstructed from measurements on the boundary, has also been studied [17, 18, 21, 22, 41, 42, 105].

Recently, several advancements have been made in the development of fluorescent contrast agents which specifically target cancer cells. Receptors for the vitamin folic acid are expressed at high levels in a number of tumor types including ovarian [106–108], breast [108], brain [109, 110], kidney [111], lung [108, 112], uterine [108], and others. In non-malignant tissues, expression of folate receptors (FR) is found only in a few tissue types [113–120]. Because of the rare occurrence of folate receptors in normal tissues and their high level of expression in tumors, FRs are a strong candidate for obtaining tumor targeting with little uptake into normal tissues. FR targeting has previously been used for radioimaging, MRI, and optical imaging [121–126]. Optical imaging agents targeted with folate have been developed using near-infrared dyes as well as visible dyes [125, 126], resulting in high tumor-to-background contrast.

Here we present a method for detecting and localizing a fluorescing tumor obscured by a turbid medium and validate the method with an experimental study. For this study, we developed a new targeted folate-indocyanine dye which was designed to operate with excitation at 785 nm, a common wavelength for inexpensive diode lasers. Tissue has relatively low absorption in this near-infrared range [127, 128]. The dye was injected into a mouse and the mouse tumor was placed into a tissue-simulating Intralipid-agarose gel phantom. We used a fluorescence microscope to

perform a 1-D measurement scan, and we estimated the tumor's lateral and depth coordinates. This measurement is related to that of Hull *et al.* [103], but it used a reflection measurement geometry which is more applicable to clinical applications.

We analyze the tumor detection and localization problems in a statistical framework. For the detection problem, we use a binary hypothesis testing approach, and we provide expressions for the tumor's detection probability. We also compute the Cramèr-Rao lower bound for the localization error. These bounds allow one to predict the performance and solvability of a detection and localization problem in advance, before making measurements. Previously, the Cramèr-Rao bound has been computed in other inverse scattering applications to bound localization error [129] and to compute performance limits in parametric, shape-based imaging [130,131]. In addition, hypothesis testing has been explored for linearized inverse scattering problems [132]. We apply these to the problem of detecting and localizing a fluorescent tumor.

6.2 Models

6.2.1 Forward Model

Consider a fluorescence measurement scan, where a single source/detector position is scanned over the top of a tissue surface to probe for a fluorescing submerged tumor. Figure 6.1(a) illustrates this measurement, with photons migrating from the source position to the tumor, and emitted photons migrating back to the same position. The inverse problem is to determine the tumor's position from the fluorescence measurements. We use a forward model based on (2.1) which gives the expected measurements that would result from a tumor at a known position.

For simplicity, we impose a few restrictions, although most of these are straightforward to remove. We consider only the $\omega = 0$ case, where unmodulated light is used, and we assume that $\mu_{a_x} = \mu_{a_m}$ and $D_x = D_m$, so that the diffusion equation Green's functions are the same for λ_x and λ_m. We also model the tumor as a point

fluorophore, and we examine the validity of this approximation in Section 6.2.2. Finally, we model the tissue as a semi-infinite region with homogeneous μ_a and D, so that a closed-form analytical solution to the diffusion equation may be employed. This model allows for rapid solution to the localization inverse problem.

Let y denote the measurement vector, and let $\tilde{f}(r)$ denote the fluorescence data vector expected from the diffusion model due to a point fluorophore at location r . We assume a source wavelength of λ_x and a detector wavelength of λ_m. Let the measurement source positions be identified as r_{s_i} and let the tumor centroid position be denoted by r. In addition, let $g(r_{s_i}, r)$ denote the domain's Green's function obtained as the solution to (2.1), with the source at r_{s_i} and the observation at r. We also denote the recorded measurement at r_{s_i} as y_i. Using the coupled diffusion equations of (3.1) and (3.2) and the approximations noted above, the fluorescence recorded at r_{s_i} from a point fluorophore located at r is given by $\tilde{f}_i(r) = g(r_{s_i}, r)g(r, r_{s_i})w_0$, where the first Green's function represents the excitation light reaching the tumor, the second represents the light emitted by the tumor back to the measurement point, and w_0 is a constant which incorporates the tumor's $\eta\mu_f$ and the efficiency of light coupling into the medium. Using the reciprocity theorem [74], the Green's function's source and observation positions can be interchanged, and $\tilde{f}_i(r) = [g(r_{s_i}, r)]^2 w_0$.

Suppose the air-tissue interface is located at $z = 0$, and that the region $z > 0$ consists of tissue. We use the extrapolated zero-flux boundary condition for the diffusion equation, where the extrapolation distance $l_s = 5.03D$ is chosen to model a scattering domain interface where the scatterers are assumed to be in water with refractive index 1.33 [47]. We apply the method of images to enforce the $\phi(z = -l_s) = 0$ boundary condition, as shown in Figure 6.1(b). The excitation source is modeled as an effective isotropic point source located one transport length inside the medium at $z_{s_i} = 3D$ [47]. We assume a semi-infinite, homogeneous medium, with a

point fluorophore at $r = (x, y, z)$, and the measurement position at $r_{s_i} = (x_{s_i}, y, z_{s_i})$. Define

$$r_a = \sqrt{(x - x_{s_i})^2 + (z - z_{s_i})^2} \tag{6.1}$$

$$r_b = \sqrt{(x - x_{s_i})^2 + (z + z_{s_i} + 2l_s)^2} \tag{6.2}$$

The forward model is given by [47]

$$\tilde{f}_i(r) = \left[\frac{\exp(-kr_a)}{r_a} - \frac{\exp(-kr_b)}{r_b}\right]^2 w \tag{6.3}$$

$$= f_i(r)w, \tag{6.4}$$

where we have bundled multiplicative constants from the closed-form Green's function and w_0 into a new parameter w, $k = \sqrt{\mu_a/D}$ and the definition of f_i is implied. Note that we consider a 2-D geometry, where the laser excitation and the tumor are located at the same y position, so no y terms appear in (6.3).

6.2.2 Validity of Point Tumor Source Model

In Section 6.2.1, we assumed that a fluorescing tumor can be approximated as a single point. To investigate the validity of this assumption for scan measurements similar to the one in our experiment, we performed a numerical simulation study with a fluorophore dispersed throughout a spherical domain of varying size and depth [44]. The 3 cm × 3 cm × 3 cm tissue phantom of Figure 6.2(a) with homogeneous background $\mu_a = 0.03$ cm^{-1} and $D = 0.033$ cm was considered. Two different-sized tumors were considered: a 2 mm fluorescent sphere and a 1 cm sphere. Each had $\eta\mu_{a_f} = 0.02$ cm^{-1}, $\mu_a = 0.030$ cm^{-1} and the same scattering properties as the background. A 1-D scan (using unmodulated light) across the top of the surface over the tumor was simulated, assuming a conceptual instrument containing a source and detector in the same position. The region was discretized into 65 × 65 × 65 voxels, and 21 measurements were simulated using multigrid finite differences to solve the diffusion equation. Figure 6.2(b) shows the simulated measurements of fluorescence

for the small tumor, normalized to the maximum, at four different depths. Figure 6.2(c) shows the simulated measurements of fluorescence for the large tumor at three different depths. Figure 6.2(d) superimposes the two results. Note in Figure 6.2(d) that the measured intensity profile is relatively invariant to the tumor's size, giving very similar results as a function of tumor depth for the two different tumor sizes considered. This result suggests that a simple point tumor model is sufficiently accurate for localizing a real tumor in a similar 1-D scanning experiment.

6.2.3 Detector Noise Model

For the detector noise, we assume the same shot noise model presented previously by Ye *et al.* [24] Let y denote the vector of measurements, and n denote the corresponding detector noise vector. We assume that n is independent, zero-mean, and Gaussian, with covariance given by Υ, where

$$[\Upsilon]_{ii} = \alpha |y_i| \tag{6.5}$$

and α is a scalar parameter of the measurement system. Previously, Ye *et al.* [25] have presented a method for estimating α from the measured data while solving the ODT inverse problem. In Section 6.5, we apply this estimation method to obtain the value of α which we use to compute statistical performance bounds related to our experimental study.

6.3 Detection

Here, we describe a procedure for detecting the presence of a tumor, and we compute the probability of detecting a tumor for a specified false alarm rate. As we demonstrate in Section 6.5.4, the probability of detection may be plotted as a function of tumor position for a particular measurement system, allowing one to characterize the instrument's diagnostic capabilities. The detection problem may be viewed as a binary hypothesis testing problem [133]. Let hypothesis H_0 correspond

to the absence of a tumor, and let the composite hypothesis $H_{1,\boldsymbol{\theta}}$ correspond to the presence of a tumor parameterized by the vector $\theta = [\ x \quad z \quad w\ ]^T$, where x and z give the location and w incorporates all scalar factors in the measurement. Let y denote the measurement vector of length P, and let n be the independent Gaussian noise vector with covariance Υ introduced in Section 6.2.3. The densities for y under both hypotheses are given by

$$p_0(y) = \frac{1}{\sqrt{(2\pi)^P|\Upsilon|}} \exp\left(-\frac{1}{2}||y||^2_{\Upsilon^{-1}}\right) \tag{6.6}$$

$$p_{1,\theta}(y) = \frac{1}{\sqrt{(2\pi)^P|\Upsilon|}} \exp\left(-\frac{1}{2}||y - wf(r)||^2_{\Upsilon^{-1}}\right) \tag{6.7}$$

where $r = [\ x \quad z\ ]^T$, and $||u||^2_V = u^H V u$.

Suppose for the moment that θ is known, and let P_F be the false alarm rate that one wishes to achieve. The Neyman-Pearson lemma [133] indicates that a likelihood ratio test (LRT) produces the highest probability of detection for a specified false alarm rate of P_F. We form the log likelihood ratio $L(y,\theta) = \ln \frac{p_{1,\theta}(y)}{p_0(y)}$, compare it to a threshold $\tilde{k}_{P_F}(\theta)$ (which is determined by P_F), and declare a tumor present if $L(y,\theta) > \tilde{k}_{P_F}(\theta)$.

By writing out the ratio of (6.7) and (6.6) and taking the logarithm, we obtain

$$L(y,\theta) = h^T(\theta)y - c(\theta) \tag{6.8}$$

where $h^T(\theta) = wf^T(r)\Upsilon^{-1}$ and $c(\theta) = \frac{1}{2}w^2 f^T(r)\Upsilon^{-1}f(r)$. Hence, the LRT is equivalent to comparing $q = h^T(\theta)y$ to a threshold, which we call $k_{P_F}(\theta)$. Note that the LRT is similar to matched filtering, as we evaluate a cross-correlation between the data y and the forward model $f(r)$, and compare it to a threshold.

The threshold $k_{P_F}(\theta)$ may be determined by rewriting the LRT in a more revealing form, using q rather than y. The 1-D Gaussian distributions $p_0(q)$ and $p_{1,\theta}(q)$ are obtained by computing the required moments:

$$\begin{aligned} E_0[q] &= h^T(\theta)E_0[y] && (6.9)\\ &= 0 && (6.10)\\ E_{1,\theta}[q] &= E_{1,\theta}[h^T(\theta)y] \\ &= w^2 f^T(\theta)\Upsilon^{-1}f(\theta) && (6.11)\\ \sigma_q^2 &= E_0[h^T(\theta)nn^T h(\theta)] && (6.12)\\ &= h^T(\theta)\Upsilon h(\theta) && (6.13) \end{aligned}$$

By definition, the false alarm rate P_F is given by

$$P_F = \int_{k_{P_F}}^{\infty} p_0(q)dq = 1 - \mathrm{erf}_*\left(\frac{k_{P_F}}{\sigma_q}\right), \tag{6.14}$$

where we define

$$\mathrm{erf}_*(r) = \int_{-\infty}^{r} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt. \tag{6.15}$$

Hence, the threshold k_{P_F} which allows us to achieve a false alarm rate P_F is $k_{P_F} = \sigma_q \mathrm{erf}_*^{-1}(1 - P_F)$.

To assess the achievable performance of the tumor detection approach, we compute the probability of a successful tumor detection. The receiver operating characteristic (ROC) [133], is defined as the probability of detection $P_D(P_F)$ specified as a function of the false alarm rate P_F. Using the threshold k_{P_F} computed above, we obtain

$$\begin{aligned} P_D &= \int_{k_{P_F}}^{\infty} p_{1,\theta}(q)dq && (6.16)\\ &= 1 - \mathrm{erf}_*\left(\frac{k_{P_F} - \bar{q}}{\sigma_q}\right) && (6.17)\\ &= 1 - \mathrm{erf}_*\left(\mathrm{erf}_*^{-1}(1 - P_F) - \frac{\bar{q}}{\sigma_q}\right), && (6.18) \end{aligned}$$

where $\bar{q} = E_{1,\theta}[q]$. From (6.18), the ROC is specified for every value of θ (thus, for any possible tumor location).

In practice, the tumor position and the true value of w are not known. Because the LRT assumes known θ, we must use a modification of the LRT called the generalized likelihood ratio test (GLRT) [133]. In the GLRT, we first suppose that the tumor is present, and we compute the maximum likelihood (ML) estimate $\hat{\theta} = \arg\max_\theta p_{1,\theta}(y)$. We then perform the LRT, using $\hat{\theta}$ in place of θ. Due to the fact that the GLRT must be used, the ROC specified in (6.18) cannot be achieved in practice, and must be considered an upper bound for the true probability of detection [133]. Computing the ML estimate $\hat{\theta}$ is the problem of localization, which is addressed in the following section.

6.4 Localization

If a tumor is present, we wish to localize it by estimating its x and z coordinates. In the process, we must also estimate w as a nuisance parameter. We use ML estimation to compute

$$\hat{\theta} = \arg\max_\theta p_{1,\theta}(y). \tag{6.19}$$

By taking the logarithm of $p_{1,\theta}(y)$, the above optimization is equivalent to minimizing an objective function

$$c(r) = \min_w ||y - wf(r)||^2_{\Upsilon^{-1}}. \tag{6.20}$$

Setting the derivative of $||y - wf(r)||^2_{\Upsilon^{-1}}$ with respect to w to zero results in the following equivalent representation of $c(r)$:

$$\tilde{w}(r) = \frac{f^T(r)\Upsilon^{-1}y}{y^T\Upsilon^{-1}y} \tag{6.21}$$

$$c(r) = ||y - \tilde{w}(r)f(r)||^2_{\Upsilon^{-1}}. \tag{6.22}$$

The tumor position $\hat{r}$ is then estimated by determining the position r which minimizes $c(r)$ in (6.22):

$$\hat{r} = \arg\min_{r} c(r) \tag{6.23}$$

$$\hat{w} = \tilde{w}(\hat{r}) \tag{6.24}$$

With the use of closed-form Green's functions in the computation of f, this minimization can be accomplished very quickly, even by evaluating $c(r)$ over an entire region of interest.

6.4.1 Cramèr-Rao Bound

To assess the precision of our localization procedure for tumors of various depths, we compute statistical bounds on the estimator variance using a model of the measurement system. Specifically, we compute the Cramèr-Rao lower bound [134] for the estimator covariance, to determine a measurement's best-case performance limits. For simplicity, we neglect bias in our computations of the Cramèr-Rao bound, as has been done previously [129]. Let $C = E[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T]$ and let J be the Fisher information matrix [134] defined as $J_{mn} = E\left[-\frac{\partial^2}{\partial\theta_m\partial\theta_n}\ln p_{1,\theta}(y)\right]$. The Cramèr-Rao bound requires that $C \geq J^{-1}$ (i.e., $C - J^{-1}$ is nonnegative definite). It can be shown [134] that

$$J = \tilde{f}'(\theta)^T \Upsilon^{-1} \tilde{f}'(\theta) \tag{6.25}$$

where

$$\tilde{f}'(\theta) = \left[\begin{array}{ccc} \frac{\partial\tilde{f}(r)}{\partial x} & \frac{\partial\tilde{f}(r)}{\partial z} & \frac{\partial\tilde{f}(r)}{\partial w} \end{array}\right]^T \tag{6.26}$$

$$= \left[\begin{array}{ccc} w\frac{\partial f(r)}{\partial x} & w\frac{\partial f(r)}{\partial z} & f(r) \end{array}\right]^T \tag{6.27}$$

In Appendix F, we present expressions for the required partial derivatives.

6.5 Experiment

6.5.1 Folate-indocyanine dye

A new folate-indocyanine dye suitable for use with 785 nm diode lasers was synthesized. The structural formula is shown in Figure 6.3. The folate-indocyanine dye synthesis procedure has been presented in detail elsewhere [45].

6.5.2 Animal Studies

To demonstrate the efficacy of the folate-indocyanine, we acquired images of a fluorescing tumor on a mouse subject. All animal procedures were carried out with approval from the Purdue Animal Care and Use Committee. Nu/nu mice were purchased from Harlan (Indianapolis, IN). The mice were at least eight weeks old when purchased and were immediately placed on folate-deficient chow (Harlan, Indianapolis, IN), and this occurred at least two weeks prior to imaging. Tumors were induced into the mice by injection of approximately 500000 L1210 cells subcutaneously. Imaging was performed approximately two weeks after tumor induction or when tumors reached approximately 1 cm in diameter.

The folate-indocyanine conjugate was injected into the femoral vein of the mice by making an incision in the leg to expose the vein. 100 μL of a PBS solution containing 20 μg of the folate-dye conjugate was then injected and images were collected two hours later. The wound was closed using Vetbond (The Butler Co., Indianapolis, IN), and the mice were euthanized at the specified times prior to imaging.

Figure 6.4 shows images of a nu/nu mouse with an L1210 tumor. The mouse was illuminated under an expanded laser diode beam at 785 nm, and images were acquired using a Roper PI-MAX intensified CCD camera. For the fluorescence measurements, a bandpass filter in the 820 nm range with 10 nm FWHM was placed over the camera. Figure 6.4(b) shows an image at the 785 nm laser excitation, and Figure 6.4(c) shows the fluorescence emitted from the mouse. From these images,

it is clear that the dye provides high contrast between tumors and the surrounding tissue.

6.5.3 Tumor Localization Measurement

Figure 6.5 schematically depicts an experiment to localize a fluorescent mouse tumor submerged under a lipid suspension. A Nu/nu mouse was induced to grow a lung tumor, injected intravenously with folate-indocyanine, and euthanized as described in Section 6.5.2. The tumor was excised and cut into two fragments. Each of the two tumor fragments was glued (using Vetbond) to the bottom of a cylindrical, 8.6 cm diameter, 1.8 cm deep plastic Petri dish. Each tumor fragment was approximately 5 mm in diameter and no thicker than 2 mm. The Petri dishes were then filled with a solid gel phantom, prepared from a suspension of 1% Intralipid and agarose [137]. In the 800 nm light wavelength range, 1% Intralipid has $\mu_a = 0.030$ cm^{-1} and $D = 0.033$ cm [57, 58], similar to the properties of human tissue. One of the Petri dishes was filled to a height of 0.79 cm, while the other was filled to a height of 1.20 cm. Assuming that the tumor centers were approximately 0.1 cm from the bottom of the Petri dishes, the true tumor depths were taken to be 0.69 cm and 1.1 cm.

To perform the fluorescence scan measurement, a previously described near-IR Raman imaging microscope (NIRIM) system [138, 139] was used. This system was originally designed for fast Raman spectral imaging using a 2-D square fiber bundle, a spectrograph, and a CCD detector. Rather than collecting a 2-D spatial grid of spectral data, we used the system to collect only a single fluorescence spectrum for each sample position. A diode laser (SDL-8630) operating at 785 nm was fiber-coupled into a microscope, illuminating the sample from above with about 100 mW of power. The same microscope simultaneously collected the fluorescence migrating from the obscured tumor to the top of the sample. The spectra were recorded by use of a spectrograph and a CCD detector (Princeton Instruments LN/CCD-1024 EHRB).

The sample was moved in one dimension with a motorized translation stage to 15 different positions, at a spacing of 0.2 cm, with the middle measurement position directly above the tumor. To minimize noise in our measurements, integration times of 20 seconds were used (although the emission could be clearly observed for shorter integration times). For each sample, a background spectrum was recorded over the Intralipid far from the tumor position. This measurement was used to subtract any effects due to dark current or background signal from the tumor fluorescence signal.

For each recorded spectrum, an intensity measurement was formed by integrating over wavelength from 800 nm to 840 nm. The resulting intensity measurement scans for both samples, with each scan normalized to arbitrary units, are shown in Figure 6.6. The localization procedure described in (6.24) was used to estimate the tumor's location, and the resulting best-fit diffusion model computations are plotted in Figure 6.6 for comparison purposes. The surface plots of Figure 6.7 show $c(r)$ in the vertical plane of points directly underneath the measurement scan computed for each of the two phantoms. The tumor coordinates were estimated to be at $(x, z) = (0.00, 0.68)$ cm and $(x, z) = (0.02, 0.95)$ cm, compared with the corresponding true coordinates $(x, z) = (0.00, 0.69)$ cm and $(x, z) = (0.00, 1.1)$ cm.

To compute the results of Figure 6.7, we used $[\Upsilon]_{ii} = |y_i|$, rather than (6.5), due to the fact that α was unknown. This results in an unknown scalar factor for the cost function $c(r)$, which has no effect on the optimization in (6.24). We address the question of determining α in Section 6.5.4.

6.5.4 Detection and localization performance bounds

To assess the diagnostic capabilities of the detection and localization methods, we considered the question of how deeply a realistic tumor could be embedded but still detected with high probability and localized with good precision. Due to the geometric constraints imposed by the microscope, we were unable to measure Intralipid-agarose samples substantially thicker than the ones described in Section 6.5.3. Hence,

we computed theoretical performance bounds using the methods of Sections 6.3 and 6.4.1. In order to develop a model for the measurement system, it was necessary to determine the noise scaling parameter α. From the data and fit depicted in Figure 6.6(b), we computed the ML estimate of α as [25]

$$\hat{\alpha} = \frac{1}{P}||y - \hat{w}f(\hat{r})||^2_{\tilde{\Upsilon}^{-1}}, \tag{6.28}$$

where $[\tilde{\Upsilon}]_{ii} = |y_i|$ and P is the number of measurements.

Using the model of the measurement system, we obtained bounds on the tumor detection probability and on the tumor localization error as described in Sections 6.3 and 6.4.1. To compute the probability of detection and the Cramèr-Rao bounds as a function of tumor position r for a realistic tumor, we used $\Upsilon = \hat{\alpha}\hat{w}\text{diag}|f(r)|$, where $\hat{w}$ was the estimate obtained in Section 6.5.3 using the data of Figure 6.6. This shot noise model is similar to the one described in (6.5), but with y_i replaced by its expected value $\hat{w}f_i(r)$. Figure 6.8(a) shows the probability of tumor detection as a function of z, for a false alarm rate of 0.03, computed using (6.18). This plot is an idealization due to the model simplifications described in Section 6.2.1 and due to the fact that the ROC is based on the LRT, rather than the GLRT. Nevertheless, it provides useful best-case information and shows that it is difficult to detect a tumor deeper than approximately 3.5 cm (the cutoff in Figure 6.8(a)) using this measurement device. Figure 6.8(b) and 6.8(c) shows lower bounds on the standard deviations σ_X and σ_Z of the position estimates of x and z, respectively, computed using (6.25). The assumption of a point source representation is considered to be valid for all positions in this example.

6.6 Conclusion

We have presented an approach for detecting and localizing an obscured, fluorescing tumor. We considered a 1-D measurement scan, with the excitation and collection at the same location above the sample. The method could be extended to 2-D planar measurement geometries for 3-D localization. In principle, a similar

geometry could be incorporated into a scanning fluorescence probe which is waved over a tissue region, or a small probe placed on a tissue surface. Such measurements could facilitate intraoperative detection and localization of tumors, allowing a surgeon to remove all tumors while minimizing damage to surrounding tissue.

The measurement system used to perform the experiment was not originally designed for this application, and several improvements could be made in subsequent work. The NIRIM system used in the experiment collects emitted light from an area which is on the order of 1 μm in diameter [138]. A larger-aperture detection system could be used to reduce the exposure times. Using time- or frequency-domain measurements, rather than only CW measurements, would potentially provide additional detection and localization criteria which are useful. In addition, all of the methods and analyses presented here can be extended to more sophisticated, multiple source/detector geometries which are typically used in FODT. Another possibility is that the dual-interfering source approach [102, 140] studied for 2-D tumor localization could be combined with our approach for estimating tumor depth, potentially resulting in a more accurate 3-D localization.

The fast detection and localization approach presented here could serve as a real-time complement to more computationally demanding, but more quantitative, FODT reconstruction algorithms. In addition, detection and localization of a fluorescing tumor might provide useful prior knowledge that can be incorporated into Bayesian FODT reconstruction approaches [42]. The ROC and the Cramèr-Rao bound results presented here could also be used to optimize an FODT measurement system geometry for the tasks of detection and localization. These possibilities, combined with the high tumor-to-background contrast offered by the folate-indocyanine imaging agent, could improve the performance in fluorescence-enhanced optical mammography [12] or other diagnostic imaging applications.

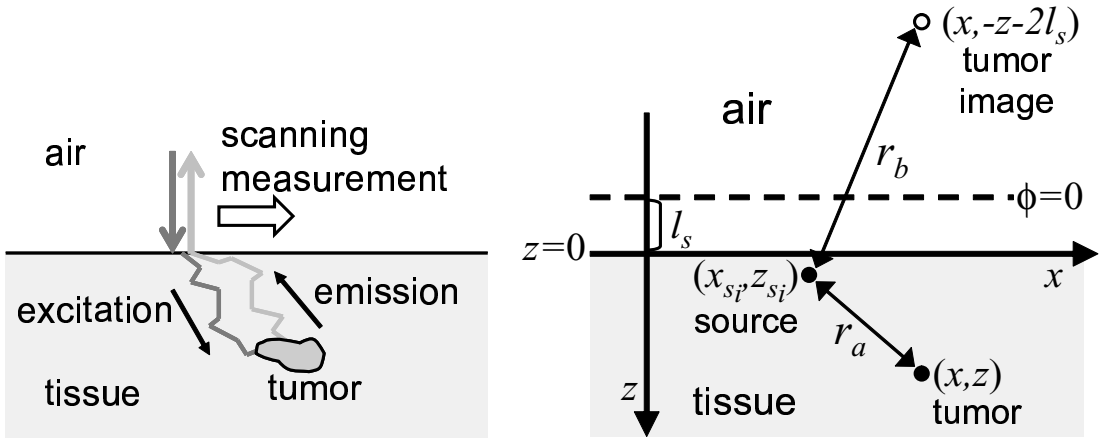


Fig. 6.1. (a) Illustration of fluorescence scan measurement, with photons migrating within the scattering tissue. (b) Semi-infinite geometry used to derive forward model. The method of images is used to insure that $\phi = 0$ for the boundary at a distance of l_s outside of the physical air-tissue interface.

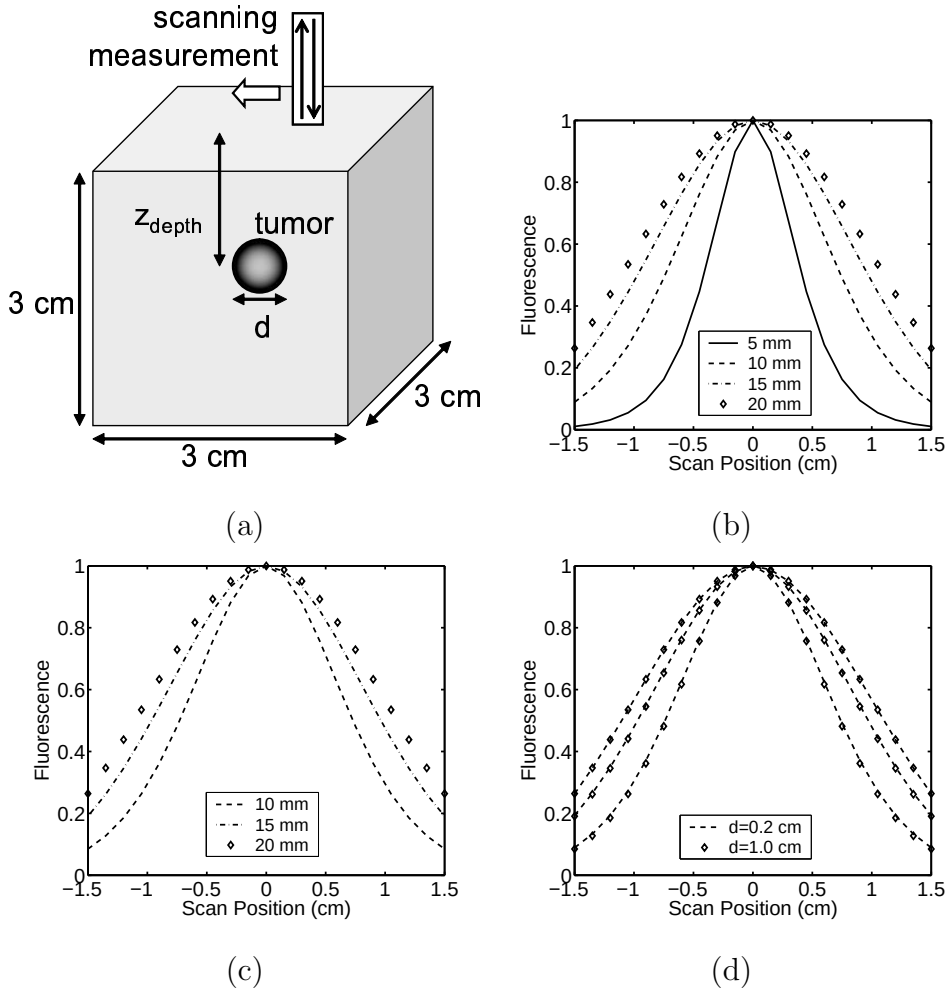


Fig. 6.2. Simulated measurement of tumor of diameter d at depth z_{depth}, with all datasets normalized to the maximum value. (a) Geometry. (b) Plot of simulated normalized intensity profile for a small tumor at different depths. (c) Plot of the intensity profile for a large tumor at different depths. (d) Superimposed plots from two different-sized tumors, showing the relative invariance to size.

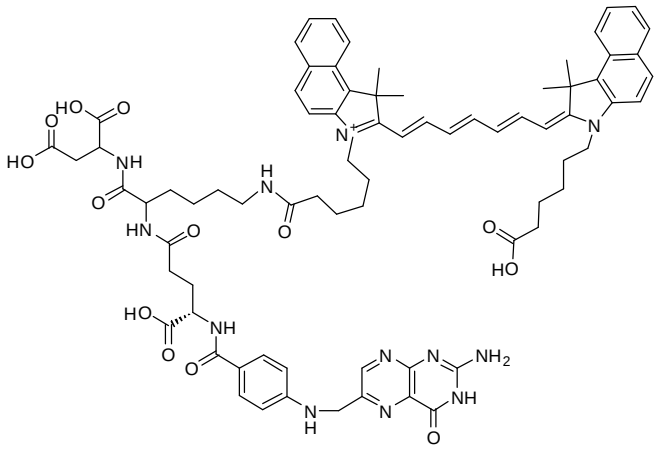
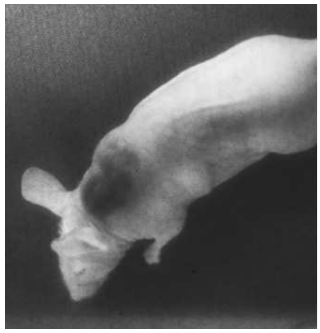


Fig. 6.3. Structural formula for folate-indocyanine

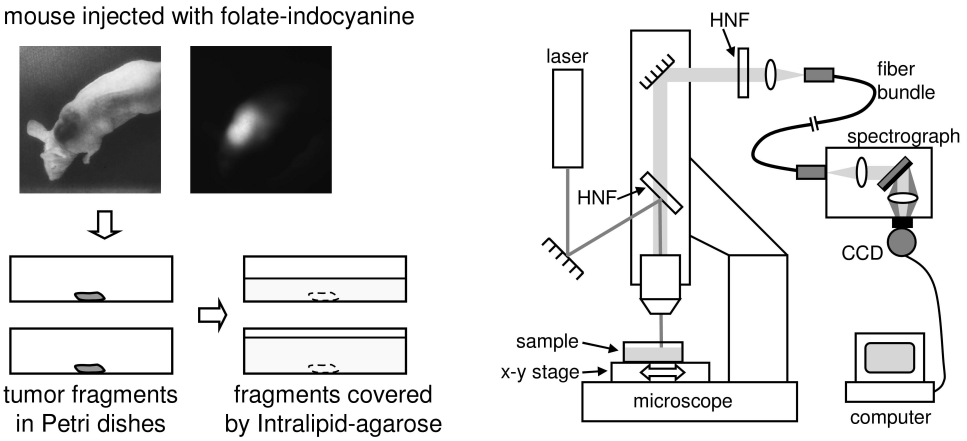


(a) Room light

(b) 785 nm excitation

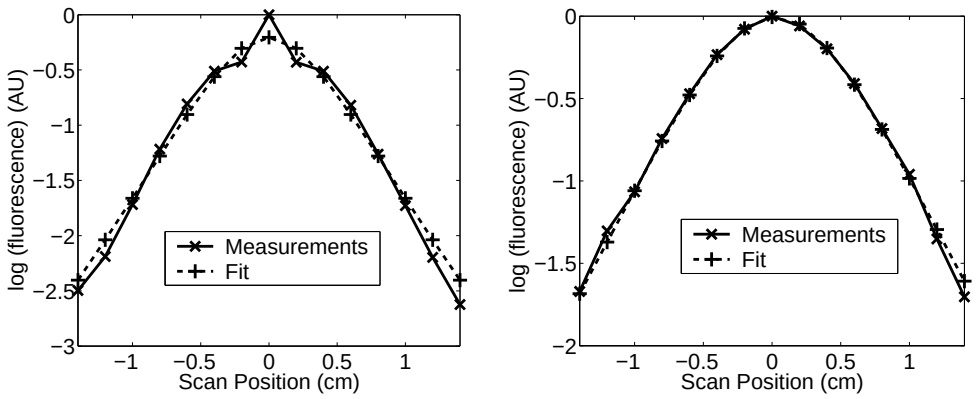
(c) 820 nm emission

Fig. 6.4. A nu/nu mouse injected with folate-indocyanine, which selectively targets folate receptors on tumors.



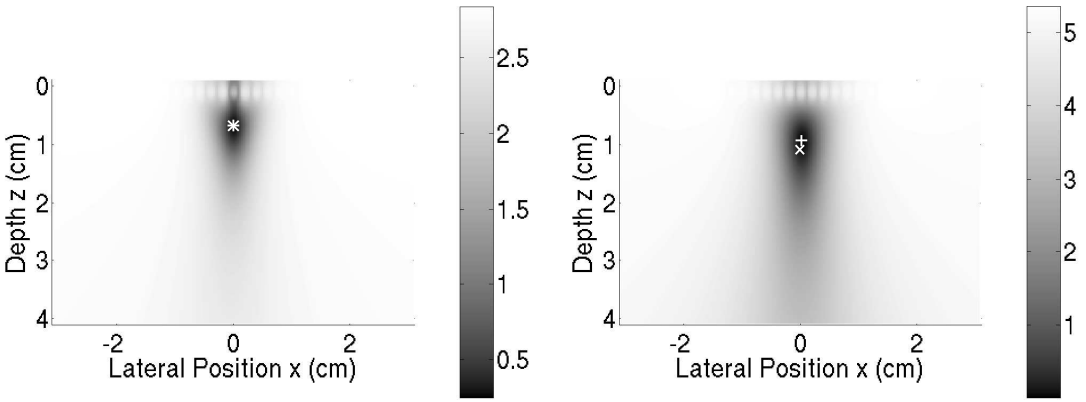
(a) (b)

Fig. 6.5. Schematic depictions of tumor localization experiment. (a) A tumor-bearing mouse is injected with folate-indocyanine, and excised tumor fragments are bonded to Petri dishes and covered with Intralipid-agarose. (b) The sample is scanned in a near-IR fluorescence microscope and measurements are recorded. HNF=Holographic Notch Filter



(a) (b)

Fig. 6.6. Normalized mouse tumor fluorescence intensity scans for two different Intralipid depths: (a) 0.69 cm deep (b) 1.1 cm deep. The dashed lines show the best fit to a diffusion model with a point fluorophore.



(a) (b)

Fig. 6.7. Cost function versus tumor position, for a mouse tumor obscured under (a) 0.69 cm and (b) 1.1 cm of Intralipid. The $\times$ symbol marks the true tumor location, while the $+$ symbol marks the estimated location.

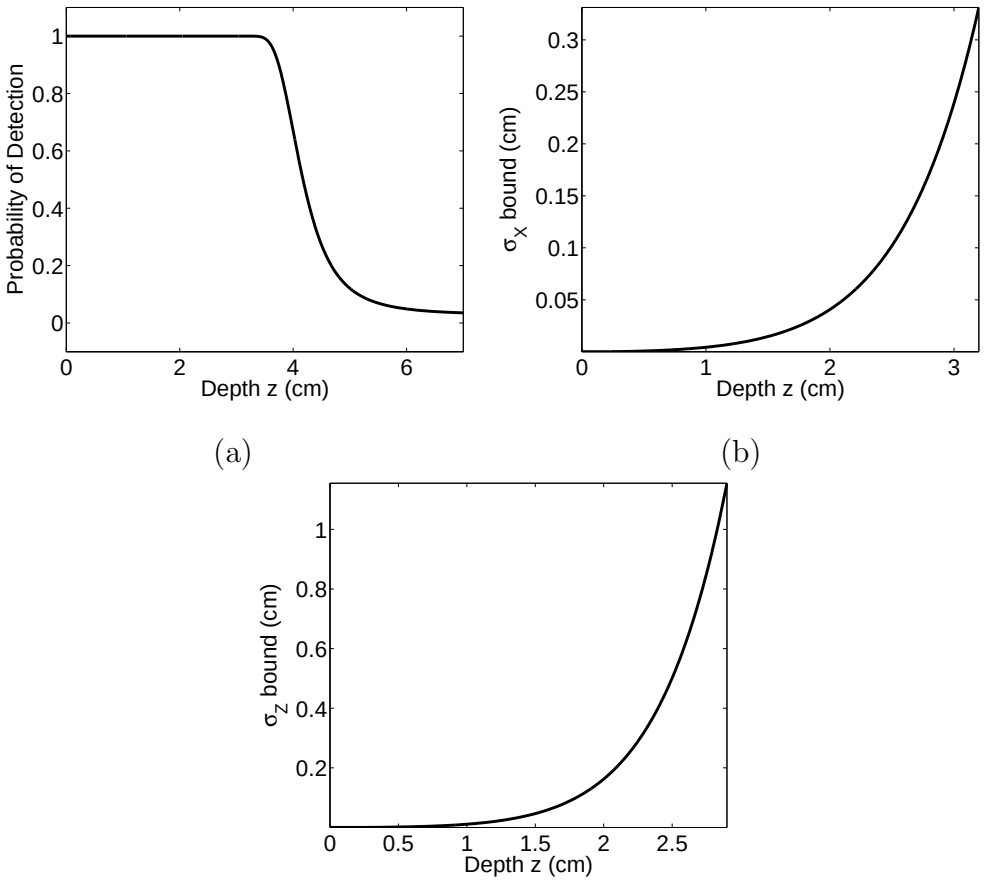


Fig. 6.8. Theoretical performance bounds for tumor measurement as a function of tumor depth. (a) Probability of detection for a false alarm rate of 0.03. (b) Cramèr-Rao bound for σ_X. (c) Cramèr-Rao bound for σ_Z.

7. OPTIMAL TIME-SEQUENTIAL MEASUREMENTS IN DYNAMIC OPTICAL DIFFUSION TOMOGRAPHY

7.1 Background

In the application of functional brain imaging, one of the principle advantages of ODT over other modalities such as fMRI is the potential for high temporal resolution [141]. Recently, researchers [86] have demonstrated the potential for reconstructing absorption changes in the brain due to hemodynamics, an indirect measure of brain activity. In addition, some recent work [142] has suggested that neuronal activity (a direct indication of brain activity) can produce observable changes in the scattering coefficient. However, these changes often take place on the order of seconds or fractions of a second, which is comparable to the acquisition time of most ODT instruments. As a result, new approaches must be developed for reconstructing an accurate time series of images.

Recently, Kaipio *et al.* [87,88] have presented a state-estimation approach for the dynamic ODT problem using Kalman filtering and Kalman smoothing techniques. While other dynamic models for inverse problems have been investigated and compared [89–91], we consider a state-space model similar to the one used by Kaipio *et al.*. Let the unknown μ_a and D image at time index i be denoted by the state vector X_i, and let the measurements at that time be denoted by Y_i. In addition, assume that X_i is a Gaussian, first-order Markov process, with a transition rule given by

$$X_{i+1} = X_i + \Delta X_i, \tag{7.1}$$

where ΔX_i is a Gaussian process which is uncorrelated from one time index to the next. Ignoring nonlinearities, assume that the measurements are described by

$$Y_i = J_i X_i + V_i \tag{7.2}$$

where V_i is independent Gaussian noise and J_i is the linear forward operator at time i. The Kalman filter is a recursive procedure for computing the estimate

$$\hat{x}_i = E[X_i|Y_i, Y_{i-1}, \dots]. \tag{7.3}$$

Using the recursive update equations, the estimate $\hat{x}_{i+1}$ is computed from Y_{i+1}, $\hat{x}_i$, and the covariance matrix

$$C_{i|i} = E[(X_i - \hat{x}_i)(X_i - \hat{x}_i)^H|Y_i, Y_{i-1}, \dots]. \tag{7.4}$$

One potential difficulty with the Kalman filter approach is the requirement for computing $C_{i|i}$, a non-sparse $N \times N$ matrix (where N is the size of X_i). For practical, 3-D imaging, storing and updating this entire matrix may be difficult. One way to circumvent this potential difficulty is to simultaneously optimize over all x_i, rather than sequentially estimating x_i from estimates at previous times. This method can be applied to both D and μ_a, and apply prior models such as the GMRF which are more sophisticated than those used to date. We also investigate a new, computationally tractable method for computing the optimal order in which the measurements should be made.

7.2 Bayesian Dynamic ODT

Suppose that measurements have been taken at time indices $i = 1, 2, \dots T$, and let

$$\begin{aligned} X &= \begin{bmatrix} X_1^T & X_2^T & \cdots & X_T^T \end{bmatrix}^T \\ Y &= \begin{bmatrix} Y_1^T & Y_2^T & \cdots & Y_T^T \end{bmatrix}^T \end{aligned}$$

One way to consider the dynamic imaging problem is to estimate each x_i sequentially, leaving the previous estimates $\hat{x}_{i-1}, \hat{x}_{i-2}, \cdots$ fixed:

$$\hat{x}_i = \arg\max_{x_i \geq 0} \{\ p(x_i|y_i, y_{i-1}, \dots)\ \}\ , \tag{7.5}$$

for each i. In effect, the Kalman filtering approach used by Kaipio *et al.* [87] is equivalent to (7.5). One difficulty with the sequential estimation approach is the appearance of the $N \times N$ covariance matrices for $p(x_i|y_{i-1}, y_{i-2}, \dots)$ (or their inverses). To demonstrate, we rewrite (7.5), with $T = 2$, in a more revealing form to obtain:

$$\begin{aligned}
\hat{x}_2 &= \arg\max_{x_2 \geq 0} \{p(x_2|y_2, y_1)\} && (7.6)\\
&= \arg\max_{x_2 \geq 0} \{p(x_2, y_2, y_1)\} && (7.7)\\
&= \arg\max_{x_2 \geq 0} \{p(y_2|x_2, y_1)p(x_2, y_1)\} && (7.8)\\
&= \arg\max_{x_2 \geq 0} \{p(y_2|x_2)p(x_2|y_1)\} && (7.9)
\end{aligned}$$

where we have used the fact that $p(y_2|x_2, y_1) = p(y_2|x_2)$, and thrown out functions of only y which do not affect the optimization. (We omit the subscript from the overloaded $p(\cdot)$ expression, for notational simplicity.) Note that the appearance of the $p(x_2|y_1)$ term requires us to consider its $N \times N$ covariance matrix (or its inverse). In general, this term may not have a computationally convenient form.

Another method of performing dynamic ODT is to jointly optimization over all of x, using all of y. In other words, one may compute

$$\hat{x} = \arg\max_{x \geq 0} \{\ p(x|y)\ \}\ , \tag{7.10}$$

For the $T = 2$ case, if we do the joint optimization over x_1 and x_2, we obtain

$$\begin{aligned}
\hat{x}_1, \hat{x}_2 &= \arg\max_{x_1, x_2 \geq 0} \{p(x_2, x_1|y_2, y_1)\} && (7.11)\\
&= \arg\max_{x_1, x_2 \geq 0} \{p(y_2, x_2, x_1, y_1)\} && (7.12)\\
&= \arg\max_{x_1, x_2 \geq 0} \{p(y_2|x_2, x_1, y_1)p(x_2, x_1, y_1)\} && (7.13)\\
&= \arg\max_{x_1, x_2 \geq 0} \{p(y_2|x_2)p(x_2, x_1, y_1)\} && (7.14)\\
&= \arg\max_{x_1, x_2 \geq 0} \{p(y_2|x_2)p(x_2|x_1, y_1)p(x_1, y_1)\} && (7.15)\\
&= \arg\max_{x_1, x_2 \geq 0} \{p(y_2|x_2)p(x_2|x_1)p(y_1|x_1)p(x_1)\}\ . && (7.16)
\end{aligned}$$

With appropriate models for $p(x_2|x_1)$ and $p(x_1)$ that have convenient forms (e.g., the GMRF), we do not have to deal explicitly with the problematic covariance matrices,

due to the lack of any terms of the form $p(x_i|y_{i-1}, y_{i-2}, \dots)$. The joint optimization can also be expected to yield better estimates than the sequential estimation approach. Due to the potentially large number of unknowns in (7.10), computational efficiency and robust convergence are essential. For very large T, we can reduce storage requirements by breaking up the entire time series into overlapping blocks of tractable size, and optimizing over each block separately.

7.3 Optimal Measurement Sequence

In their initial treatise, Kaipio *et al.* [87] observed that the order in which the dynamic ODT measurements were performed greatly affected the reconstruction quality. Intuitively, if measurements at consecutive time indices are localized to the same general area, then they are not very informative. Instead, consecutive measurements should sample different regions within the domain in order to get better spatial coverage. This observation is related to the concept of time-sequential sampling presented by Allebach [143]. Willis and Bresler [144] have studied the optimal scan for projection tomography applications. Here we examine the optimal order in which measurements are made in optical diffusion tomography. We optimize the mutual information $I(X;Y)$ over the measurement order. The order of the measurements is changed by permuting the indices i for the J_i terms in (7.2). In the static FODT problem, we have observed previously [42] that $I(X;Y)$ is tractable to compute because it requires only determinants of $P \times P$ matrices (where P is the number of measurements, typically much smaller than the number of voxels N). Our preliminary findings indicate that similar computational tractability exists for the dynamic problem as well.

Let $I(X;Y)$ be the mutual information between X and Y. Then

$$I(X;Y) \;=\; H(Y) - H(Y|X) \tag{7.17}$$

where $H(Y)$ and $H(Y|X)$ denote the differential entropy and the conditional differential entropy, respectively. Using the chain rule for differential entropy [65], we obtain

$$H(Y|X) = \sum_{i=1}^{T} H(Y_i|X, Y_1, \ldots, Y_{i-1}). \tag{7.18}$$

Note that Y_i is completely specified by X_i and the detector noise process V_i. When X_i is known, knowledge of Y_j or X_j for $j \neq i$ does not affect our uncertainty about Y_i. Hence,

$$H(Y_i|X, Y_1, \ldots, Y_{i-1}) = H(Y_i|X_i) \tag{7.19}$$

and

$$H(Y|X) \;\; = \;\; \sum_{i=1}^{T} H(Y_i|X_i). \tag{7.20}$$

Let Γ_i be the covariance for the detector noise V_i. Suppose we perform M measurements at each time index (*i.e.*, we use one source and all M detectors at each measurement time). Then, using the results of Appendix D,

$$H(Y_i|X_i) \;\; = \;\; \log(\pi)^M + M + \log|\Gamma_i| \tag{7.21}$$

and

$$H(Y|X) \;\; = \;\; \sum_{i=1}^{T} \log(\pi)^M + M + \log|\Gamma_i| \tag{7.22}$$

$$= \;\; T \times (\log(\pi)^M + M) + \sum_{i=1}^{T} \log|\Gamma_i|. \tag{7.23}$$

Suppose we change the measurement order by permuting the indices i for all of the J_i and Γ_i. It is clear that the summation (7.23) would still contain all of the same terms as before, appearing in a different order. Hence, $H(Y|X)$ is independent of the order in which the measurements are made. We may neglect $H(Y|X)$ as a constant in any optimization of $I(X;Y)$ over the measurement sequence, and consider only the $H(Y)$ component. Intuitively, this is reasonable, as maximization of information $I(X;Y)$ implies minimization of the statistical redundancy in the measurements.

For simplicity, we continue to assume a linearized forward operator for computing the mutual information. Let $\hat{x}_i = E[X_i|Y_j, Y_{j-1}, \dots]$ and $C_{i|j} = E[(X_i - \hat{x}_i)(X_i - \hat{x}_i)^H|Y_j, Y_{j-1}, \dots]$, and let $C_{\Delta x}$ be the covariance matrix for all ΔX_i. In addition, assume that $C_{i|j} = C_{i|j}^H$ and that $C_{\Delta x} = C_{\Delta x}^H$. Let Γ_i be the covariance for V_i. From standard optimal filtering results [145], it is possible to compute $C_{i|i}$ and $C_{i+1|i}$ from $C_{i-1|i}$ and $C_{\Delta x}$ using the following recursive equations:

$$C_{i|i} = C_{i|i-1} - C_{i|i-1}J_i^H(J_iC_{i|i-1}J_i^H + \Gamma_i)^{-1}J_iC_{i|i-1} \tag{7.24}$$

$$C_{i+1|i} = C_{i|i-1} - C_{i|i-1}J_i^H(J_iC_{i|i-1}J_i^H + \Gamma_i)^{-1}J_iC_{i|i-1} + C_{\Delta x} \tag{7.25}$$

Define

$$W_{ab}^i = J_aC_{i|i-1}J_b^H \tag{7.26}$$

$$W_{ab}^{\Delta x} = J_aC_{\Delta x}J_b^H. \tag{7.27}$$

By premultiplying (7.25) by J_a and postmultiplying by J_b^H, we obtain

$$W_{ab}^{i+1} = W_{ab}^i - W_{ai}^i(W_{ii}^i + \Gamma_i)^{-1}W_{ib}^i + W_{ab}^{\Delta x} \tag{7.28}$$

We can compute $H(Y)$ by taking determinants of terms of the form appearing in (7.28). Using the same procedure that is used to derive (D.8) in Appendix D, we obtain

$$H(Y_i|Y_1, \dots, Y_{i-1}) = \log(\pi)^M + M + \log|J_iC_{i|i-1}J_i^H + \Gamma_i| \tag{7.29}$$

$$= (\text{constant}) + \log|W_{ii}^i + \Gamma_i|. \tag{7.30}$$

Using the chain rule for conditional differential entropy

$$H(Y) = \sum_{j=1}^{T} H(Y_j|Y_1, \dots, Y_{j-1}) \tag{7.31}$$

and applying (7.30) the resulting expression for $H(Y)$ is given by

$$H(Y) = \sum_{i=1}^{T} \log|W_{ii}^i + \Gamma_i| + (\text{constant}). \tag{7.32}$$

Note that computing $H(Y)$ requires the determinants of several $M \times M$ matrices. To evaluate $H(Y)$, we can use (7.28) recursively to obtain all of the required W_{ab}^{i} matrix terms. A suitable recursive procedure for computing $H(Y)$ is given in Appendix G.

Suppose we wish to change the measurement order and reevaluate $H(Y)$. To change the order, we permute the indices i for all of the J_i and and Γ_i matrices. As a result, the W_{ab}^{i} matrix terms which are used to compute $H(Y)$ also must change accordingly. With the new indices i assigned to J_i and Γ_i, we run the algorithm in Appendix G to compute the new value of $H(Y)$. The optimization of $H(Y)$ over the measurement order may be performed by a direct search over several different measurement configurations.

8. ONGOING WORK TOWARD PRACTICAL IMAGING

In the previous Chapters, we have presented Bayesian methods for ODT, FODT, and kinetic model parameter imaging. We have also addressed the problem of fluorescent tumor localization and detection from a statistical framework. In principle, all of these methods can be applied to clinical diagnostic imaging of tumors to obtain high-quality images. We have also shown that the statistical framework, with reasonable models of the measurement system and the unknown image variability, can be used to obtain performance bounds which can be useful in determining and optimizing experimental design. The fast localization/detection methods and the full, quantitative ODT/FODT image reconstruction methods can be regarded as opposite ends of the entire spectrum of techniques that can be applied for tumor diagnosis. In particular, the kinetic model parameter imaging opens up new possibilities for clinical molecular imaging applications, where capillary permeability in the vicinity of tumors and the uptake kinetics of receptor-targeted dyes provide new possibilities for examining tumor physiology.

One of the major focuses of practical ODT research in the clinic has been optical breast imaging. To date, a number of preliminary clinical results have been reported [146–149], but useful results have been difficult to obtain on a consistent basis. To improve the diagnostic capabilities of optical breast imaging, one must consider all the ways a tumor may be distinguished from its surroundings. One possibility is its spectral characteristics. Blood oxygenation may be increased in the vicinity of a tumor due to increased metabolism. This phenomenon is observable in the optical absorption spectrum due to the different spectral peaks of oxy- and deoxyhemoglobin. Such effects have been observed in preliminary clinical imaging [147], and direct reconstruction of oxy- and deoxyhemoglobin from multiwavelength optical

measurements has been investigated [150]. Another possibility is the use of temporal information, perhaps through the use of periodic breathing exercises. Due to the expected higher concentration of blood in the vicinity of tumors, such a procedure might produce observable contrast and allow a self-calibrating difference imaging problem, similar to work which has been done in imaging of the human brain [86]. Still another possible means of distinguishing tumors from surrounding tissue is via the use of spatial resolution. Due to the need in the area of early tumor diagnosis, it is hoped that the tumor would be small and well localized compared with the more slowly-varying features of the background tissue. This suggests the possibility that multiresolution reconstruction algorithms employing wavelets may be advantageous. In addition, multiscale localization algorithms based on hypothesis testing [132], or generalizations of the methods of Chapter 6 utilizing a point-absorber model with an appropriate random field model for the background tissue could be applied. Such localization algorithms could provide prior information for more quantitative reconstruction algorithms, and they could directly characterize an instrument's diagnostic capabilities in terms of the ROC.

Recent investigations in the clinic have also made use of multimodality imaging [12, 151, 152]. ODT reconstructions have been performed simultaneously with other imaging modalities in order to validate the ODT results. In addition, sensor fusion algorithms have been studied for image reconstruction. Barnett *et al.* [152] have investigated Bayesian methods for estimating piecewise constant tissue property values using tissue type boundaries provided by segmented MRI images. This approach is related to earlier methods applied to PET image reconstruction [153]. Li *et al.* [146] have developed an algorithm for using X-ray images to divide the ODT reconstruction volume into regions with different Tikhonov regularization parameters. However, to the best of my knowledge, sensor fusion approaches which use Markov random fields as prior models have yet to be considered for ODT. Previously, Chen *et al.* [154] have presented a method for using MRI images as *a priori* information for emission tomography. The emission tomography images were reconstructed using

a Gibbs prior model to enforce smoothness and a line process [155] initialized with boundary information from the MRI images to model edges. A similar approach may be effective for sensor fusion involving ODT reconstruction.

LIST OF REFERENCES

LIST OF REFERENCES

[1] S. R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, 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 Proc. Magazine*, 18(6):57–75, Nov. 2001.

[3] V. G. Peters, D. R. Wyman, M. S. Patterson, and G. L. Frank. Optical properties of normal and diseased human breast tissues in the visible and near infrared. *Phys. Med. Biol.*, 35(9):1317–1334, September 1990.

[4] T. L. Troy, D. L. Page, and E. M. Sevick-Muraca. Optical properties of normal and diseased breast tissue: prognosis for optical mammography. *J. Biomed. Opt.*, 1(3):342–355, July 1996.

[5] 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(10):2529–2537, May 15 1991.

[6] B. Ballou, G. W. Fisher, T. R. Hakala, and D. L. Farkas. Tumor detection and visualization using cyanine fluorochrome-labeled antibodies. *Biotechnol. Prog.*, 13(5):649–658, 1997.

[7] 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.

[8] J. A. Reddy and P. S. Low. Folate-mediated targeting of therapeutic and imaging agents to cancers. *Critical Reviews in Therapeutic Drug Carrier Systems*, 15(6):587–627, 1998.

[9] 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.

[10] U. Mahmood, C. Tung, Jr A. Bogdanov, and R. Weissleder. Near-infrared optical imaging of protease activity for tumor detection. *Radiology*, 213(3):866–870, December 1999.

[11] 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(3):392–398, September 2000.

[12] 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*, 97(6):2767–2772, March 14 2000.

[13] 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. Biotech.*, 19:327–331, April 2001.

[14] 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(2):122–133, April 2001.

[15] M. S. Patterson and B. W. Pogue. Mathematical model for time-resolved and frequency-domain fluorescence spectroscopy in biological tissues. *Appl. Opt.*, 33(10):1963–1974, April 1994.

[16] 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(1):55–64, 1997.

[17] 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(13):2206–2215, May 2001.

[18] M. A. O'Leary, D. A. Boas, X. D. Li, B. Chance, and A. G. Yodh. Fluorescence lifetime imaging in turbid media. *Optics Letters*, 21(2):158–160, January 1996.

[19] D.Y. Paithankar, A.U. Chen, B.W. Pogue, M.S. Patterson, and E.M. Sevick-Muraca. Imaging of fluorescent yield and lifetime from multiply scattered light reemitted from random media. *Applied Optics*, 36(10):2260–2272, April 1997.

[20] H. Jiang. Frequency-domain fluorescent diffusion tomography: a finite-element-based algorithm and simulations. *Appl. Opt.*, 37(22):5337–5343, August 1998.

[21] J. Chang, H. L. Graber, and R. L. Barbour. Luminescence optical tomography of dense scattering media. *J. Opt. Soc. Am. A*, 14:288–299, January 1997.

[22] V. Ntziachristos and R. Weissleder. Experimental three-dimensional fluorescence reconstruction of diffuse media by use of a normalized Born approximation. *Opt. Lett.*, 26(12):893–895, June 2001.

[23] D. A. Boas. A fundamental limitation of linearized algorithms for diffuse optical tomography. *Opt. Express*, 1(13):404–413, December 22 1997.

[24] 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. Optical Society America A*, 16(10):2400–2412, October 1999.

[25] J. C. Ye, C. A. Bouman, K. J. Webb, and R. P. Millane. Nonlinear multigrid algorithms for Bayesian optical diffusion tomography. *IEEE Trans. on Image Processing*, 10(6):909–922, June 2001.

[26] J. C. Ye, K. J. Webb, R. P. Millane, and T. J. Downar. Modified distorted Born iterative method with an approximate Fréchet derivative for optical diffusion tomography. *J. Optical Society America A*, 16(7):1814–1826, July 1999.

[27] A. B. Milstein, S. Oh, J. S. Reynolds, K. J. Webb, C. A. Bouman, and R. P. Millane. Three-dimensional Bayesian optical diffusion tomography using experimental data. *Optics Letters*, 27:95–97, January 2002.

[28] 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. Optical Society America A*, 19(10):1983–1993, Oct. 2002.

[29] S. Oh, A. B. Milstein, K. J. Webb, C. A. Bouman, and R. P. Millane. Three-dimensional Bayesian optical diffusion tomography: source-detector calibration. In *SPIE Photonics West, Int. Symp. Optical Biopsy IV*, pages 174–182, San Jose, CA, USA, Jan. 2002.

[30] R. P. Millane, S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. A Bayesian image reconstruction algorithm for optical diffusion tomography incorporating source-detector calibration. In *Image and Vision Computing '01 New Zealand*, Dunedin, New Zealand, Nov. 26-28 2001.

[31] S. Oh, A. Milstein, J. S. Reynolds, K. J. Webb, C. A. Bouman, and R. P. Millane. Reconstructing optical diffusion images from frequency domain experimental data. In *OSA Annual Meeting*, Long Beach, USA, Oct. 2001.

[32] A. B. Milstein, S. Oh, K. J. Webb, and C. A. Bouman. Three-dimensional optical diffusion imaging in a Bayesian framework. In *IEEE/OSA Conf. Lasers and Electro-Optics 2001*, Baltimore, USA, May 2001.

[33] K. J. Webb, A. B. Milstein, S. Oh, C. A. Bouman, and R. P. Millane. Three-dimensional optical diffusion imaging with detector noise. In *European Conf. Biomed. Optic*, Munich, Germany, June 2001.

[34] S. Oh, A. Milstein, R. P. Millane, C. A. Bouman, and K. J. Webb. Three-dimensional Bayesian optical diffusion tomography with source-detector calibration. In *Signal Recovery and Synthesis*, Albuquerque, USA, Nov. 2001.

[35] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. A general framework for nonlinear multigrid inversion. *IEEE Trans. Image Processing*, *to appear*.

[36] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. Multigrid inversion algorithms with applications to optical diffusion tomography. In *Proceedings of the 36th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 3-6 2002.

[37] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. Multigrid algorithms for optimizations and inverse problems. In *2003 Electronic Imaging*, Santa Clara, CA, USA, Jan. 20-25 2003.

[38] R. P. Millane, A. B. Milstein, S. Oh, C. A. Bouman, and K. J. Webb. Image reconstruction in optical diffusion tomography. In *2002 SPIE Annual Meeting*, Seattle, WA, USA, July 7-11 2002.

[39] N. Iftimia and H. Jiang. Quantitative optical image reconstructions of turbid media by use of direct-current measurements. *Applied Optics*, 39(28):5256–5261, October 2000.

[40] D. Boas, T. Gaudette, and S. Arridge. Simultaneous imaging and optode calibration with diffuse optical tomography. *Opt. Express*, 8(5):263–270, February 2001.

[41] 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(16):3081–3094, June 2003.

[42] A. B. Milstein, J. J. Stott, S. Oh, D. A. Boas, R. P. Millane, C. A. Bouman, and K. J. Webb. Fluorescence optical diffusion tomography using multiple-frequency data. *J. Opt. Soc. Am. A*, 21(6):1035–1049, June 2004.

[43] A. B. Milstein, Seungseok Oh, K. J. Webb, and C. A. Bouman. Estimation of kinetic model parameters in optical diffusion tomography. In *Computational Imaging II, Proceedings of the SPIE Electronic Imaging*, volume 5299, San Jose, CA, Jan. 19 2004.

[44] K. J. Webb, A. B. Milstein, M. D. Kennedy, K. N. Jallad, C. A. Bouman, D. Ben-Amotz, and P. S. Low. Folate conjugate fluorescence labeling for tumor localization. In *Third Inter-Institute Workshop on Diagnostic Optical Imaging and Spectroscopy: The Clinical Adventure. Bethesda, MD, 2002, vol Poster Presentation*, Bethesda, MD, September 2002.

[45] A. B. Milstein, M. D. Kennedy, P. S. Low, C. A. Bouman, and K. J. Webb. Detection and localization of a fluorescent mouse tumor in a turbid medium. to be submitted, 2004.

[46] J. S. Reynolds, A. Przadka, S. Yeung, and K. J. Webb. Optical diffusion imaging: a comparative numerical and experimental study. *Applied Optics*, 35(19):3671–3679, July 1996.

[47] 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. Optical Society America A*, 11(10):2727–2741, October 1994.

[48] 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. *Applied Optics*, 36:2252–2259, April 1997.

[49] C. A. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. on Image Processing*, 2(3):296–310, July 1993.

[50] K. Sauer and C. A. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. on Signal Processing*, 41(2):534–548, February 1993.

[51] J. C. Adams. MUDPACK: Multigrid portable FORTRAN software for the efficient solution of linear elliptic partial differential equations. *Appl. Math. Comput.*, 34:113–146, 1989.

[52] J. J. Duderstadt and L. J. Hamilton. *Nuclear Reactor Analysis*. Wiley, New York, 1976.

[53] S. S. Saquib, C. A. Bouman, and K. Sauer. ML parameter estimation for Markov random fields with applications to Bayesian tomography. *IEEE Trans. on Image Processing*, 7(7):1029–1044, July 1998.

[54] K. Sauer and C. Bouman. Bayesian estimation of transmission tomograms using segmentation based optimization. *IEEE Trans. on Nuclear Science*, 39:1144–1152, 1992.

[55] Q. Zhang, T. J. Brukilacchio, T. Gaudett, L. Wang, A. Li, and D. A. Boas. Experimental comparison of using continuous-wave and frequency-domain diffuse optical imaging systems to detect heterogeneities. In *Proc. SPIE*, volume 4250, pages 219–238, June 2001.

[56] R. H. Mayer, J. S. Reynolds, and E. M. Sevick-Muraca. Measurement of the fluorescence lifetime in scattering media by frequency-domain photon migration. *Appl. Opt.*, 38:4930–4938, August 1999.

[57] H. J. van Staveren, C. J. M. Moes, J. van Marie, S. A. Prahl, and M. J. C. van Gemert. Light scattering in Intralipid-10% in the wavelength range of 400-1100 nm. *Appl. Opt.*, 30(31):4507–4514, November 1991.

[58] G. M. Hale and M. R. Querry. Optical constants of water in the 200-nm to 200-μm wavelength region. *Appl. Opt.*, 12(3):555–563, March 1973.

[59] R. C. Benson and H. A. Kues. Fluorescence properties of indocyanine green as related to angiography. *Phys. Med. Biol.*, 23(1):159–163, January 1978.

[60] 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(4):575–583, April 1976.

[61] A. D. Klose and A. H. Hielscher. A transport-theory based reconstruction algorithm for optical tomography. In *Optical Tomography and Spectroscopy of Tissue III*, volume Proc. 3597, pages 26–35, 1999.

[62] 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(4):778–791, Feb. 1 2002.

[63] V. Ntziachristos, J. Culver, M. Holboke, A. G. Yodh, and B. Chance. Optimal selection of frequencies for diffuse optical tomography. In *OSA Biomedical Topical Meetings (Optical Society of America)*, pages 475–477, Miami Beach, FL, Apr. 2000.

[64] C. E. Shannon. A mathematical theory of communication. *Bell. Syst. Tech. J.*, 27:379–423,623–656, July,October 1948.

[65] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.

[66] R. F. Wagner, D. G. Brown, and M. S. Pastel. Application of information theory to the assessment of computed tomography. *Med. Phys.*, 6(2):83–94, Mar./Apr. 1979.

[67] M. Fuderer. The information content of MR images. *IEEE Trans. Med. Imag.*, 7(4):368–380, Dec. 1988.

[68] 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. Imag.*, 8(4):322–336, Dec. 1989.

[69] A. O. Hero and L. Shao. Information analysis of single photon emission computed tomography with count losses. *IEEE Trans. Med. Imag.*, 9(2):117–127, June 1990.

[70] 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(10):701–703, May 15 2001.

[71] 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(1):102–110, Jan. 2003.

[72] J.J. Stott and D. A. Boas. A practical comparison between time-domain and frequency-domain diffusive optical imaging systems. In *OSA Biomedical Topical Meetings (Optical Society of America)*, pages 626–628, Miami Beach, FL, Apr. 2002.

[73] T. Berger. *Rate Distortion Theory: a Mathematical Basis for Data Compression*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1971.

[74] A. Ishimaru. *Wave Propagation and Scattering in Random Media*, volume 1. Academic Press, New York, New York, 1978.

[75] E. K. P. Chong and S. H. Zak. *An Introduction to Optimization*. John Wiley & Sons, Inc., New York, 1996.

[76] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging*. Institute of Physics Publishing, Philadelphia, 1998.

[77] A. D. Klose and A. H. Hielscher. Optical tomography using the time-independent equation of radiative transfer-part 2: inverse model. *JQSRT*, 72(5):715–732, March 2002.

[78] J. J. Stott, J. P. Culver, S. R. Arridge, and D. A. Boas. Optode positional calibration in diffuse optical tomography. *Appl. Opt.*, 42(16):3154–3162, June 2003.

[79] E. D. Morris, C. J. Endres, K. C. Schmidt, B. T. Christian, R. F. Muzic Jr., and R. E. Fisher. Kinetic modeling in pet. In M. Wernick and J. Aarsvold, editors, *Emission Tomography: the fundamentals of PET and SPECT*. Academic Press, April 2004.

[80] W. R. Potter, D. A. Bellnier, and T. J. Dougherty. Optical methods for *in-vivo* pharmacokinetics. In *Proc. SPIE: Optical Methods for Tumor Treatment and Detection: Mechanisms and Techniques in Photodynamic Therapy*, volume 1645, pages 166–170, Los Angeles, CA, Jan. 1992.

[81] R. A. Weersink, J. E. Hayward, K. R. Diamond, and M. S. Patterson. Accuracy of noninvasive *in vivo* measurements of photosensitizer uptake based on a diffusion model of reflectance spectroscopy. *Photochem. Photobiol.*, 66(3):326–335, Sept. 1997.

[82] M. Gurfinkel *et al.* Pharmacokinetics of ICG and HPPH-car for the detection of normal and tumor tissue using fluorescence, near-infrared reflectance imaging: a case study. *Photochem. Photobiol.*, 72(1):94–102, July 2000.

[83] R. Springett, Y. Sakata, and D. T. Delpy. Precise measurement of cerebral blood flow in newborn piglets from the bolus passage of indocyanine green. *Phys. Med. Biol.*, 46(8):2209–2225, July 2001.

[84] D. J. Cuccia, F. Bevilacqua, A. J. Durkin, S. Merritt, B. J. Tromberg, G. Gulsen, H. Yu, J. Wang, and O. Nalcioglu. *In vivo* quantification of optical contrast agent dynamics in rat tumors by use of diffuse optical spectroscopy with magnetic resonance imaging coregistration. *Appl. Opt.*, 42(16):2940–2950, June 2003.

[85] C. H. Schmitz, H. L. Graber, H. Luo, I. Arif, J. Hira, Y. Pei, A. Bluestone, S. Zhong, R. Andronica, I. Soller, N. Ramirez, S. S. Barbour, and R. L. Barbour. Instrumentation and calibration protocol for imaging dynamic features in dense-scattering media by optical tomography. *Appl. Opt.*, 39(34):6466–6486, December 1 2000.

[86] A. Y. Bluestone, G. Abdoulaev, C. H. Schmitz, R. L. Barbour, and A. H. Hielscher. Three-dimensional optical tomography of hemodynamics in the human head. *Opt. Express*, 9(6):272–286, Sept. 10 2001.

[87] V. Kolehmainen, S. Prince, S. R. Arridge, and J. P. Kaipio. State-estimation approach to the nonstationary optical tomography problem. *J. Opt. Soc. Am. A*, 20(5):876–889, May 2003.

[88] S. Prince, V. Kolehmainen, J. P. Kaipio, M. A. Franceschini, D. Boas, and S. R. Arridge. Time-series estimation of biological factors in optical diffusion tomography. *Phys. Med. Biol.*, 48(11):1491–1504, June 7 2003.

[89] U. Schmitt and A. K. Louis. Efficient algorithms for the regularization of dynamic inverse problems: I. theory. *Inverse Problems*, 18(3):645–658, June 2002.

[90] U. Schmitt, A. K. Louis, C. Wolters, and M. Vauhkonen. Efficient algorithms for the regularization of dynamic inverse problems: II. applications. *Inverse Problems*, 18(3):659–676, June 2002.

[91] Y. Zhang, A. Ghodrati, and D. H. Brooks. Analysis of spatial-temporal regularization methods for linear inverse problems from a common statistical framework. In *International Symposium on Biomedical Imaging*, pages 772–775, Arlington, VA, April 2004.

[92] M. N. Wernick, E. J. Infusino, and M. Milosevic. Fast spatio-temporal image reconstruction for dynamic PET. *IEEE Trans. Med. Imag.*, 18:185–195, March 1999.

[93] T. E. Nichols, J. Qi, E. Asma, and R. M. Leahy. Spatiotemporal reconstruction of list-mode PET data. *IEEE Trans. Med. Imag.*, 21(4):396–404, April 2002.

[94] B. W. Reutter, G. T. Gullberg, and R. H. Huesman. Direct least-squares estimation of spatiotemporal distributions from dynamic SPECT projections using a spatial segmentation and temporal B-splines. *IEEE Trans. Med. Imag.*, 19(5):434–450, May 2000.

[95] S. Ahn, J. A. Fessler, T. E. Nichols, and R. A. Koeppe. Covariance of kinetic parameter estimators based on time activity curve reconstructions: preliminary study on 1-D dynamic imaging. In *International Symposium on Biomedical Imaging: from Nano to Macro*, volume 2, pages 368–371, Arlington, VA, April 2004.

[96] J. A. Jacquez. *Compartmental analysis in biology and medicine*. The University of Michigan Press, Ann Arbor, 1985.

[97] R. E. Carson and K. Lange. The EM parametric image reconstruction algorithm. *J. Am. Stat. Assoc.*, 89(389):20–22, March 1985.

[98] J. Matthews, D. Bailey, P. Price, and V. Cunningham. The direct calculation of parametric images from dynamic PET data using maximum-likelihood iterative reconstruction. *Phys. Med. Biol.*, 42(6):1155–1173, June 1997.

[99] M. Kamasak, C. A. Bouman, E. D. Morris, and K. Sauer. Direct reconstruction of kinetic parameter images from dynamic PET data. In *Proceedings of 37th Asilomar Conference on Signals, Systems and Computers*, pages 1919–1923, Pacific Grove, CA, Nov. 9-12 2003.

[100] W. Jager, H. Feistel, E. M. Paterok, G. Ronay, A. H. Tulusan, F. Wolf, and N. Lang. Resection guided by antibodies (REGAJ): a diagnostic procedure during second-look operation in ovarian cancer patients. *British Journal of Cancer - Supplement*, 10:18–20, 1990.

[101] V. Ntziachristos, C. Bremer, and R. Weissleder. Fluorescence imaging with near-infrared light. *Eur. Radiol.*, 13:195–208, Jan. 2003.

[102] Y. Chen, G. Zheng, Z. H. Zhang, D. Blessington, M. Zhang, H. Li, Q. Liu, L. Zhou, X. Intes, S. Achilefu, and B. Chance. Metabolism-enhanced tumor localization by fluorescence imaging: *in vivo* animal studies. *Opt. Lett.*, 28(21):2070–2072, Nov. 1 2003.

[103] E. L. Hull, M. G. Nichols, and T. H. Foster. Localization of luminescent inhomogeneities in turbid media with spatially resolved measurements of cw diffuse luminescent emittance. *Appl. Opt.*, 37(13):2755–2765, May 1998.

[104] M. Pfister and B. Scholz. Localization of fluorescent spots with space-space MUSIC for mammographylike measurement system. *J. Biomed. Opt.*, 9(3):481–487, May/June 2004.

[105] E. Shives, Y. Xu, and H. Jiang. Fluorescence lifetime tomography of turbid media based on an oxygen-sensitive dye. *Opt. Express*, 10(26):1557–1562, Dec. 30 2002.

[106] G. Toffoli, C. Cernigoi, A. Russo, A. Gallo, M. Bagnoli, and M. Boiocchi. Overexpression of folate binding protein in ovarian cancers. *International Journal of Cancer*, 74:193–198, 1997.

[107] S. Miotti, S. Canevari, S. Menard, D. Mezzanzanica, G. Porro, S. M. Pupa, M. Regazzoni, E. Tagliabue, and M. I. Colnaghi. Characterization of human ovarian carcinoma-associated antigens defined by novel monoclonal antibodies with tumor-restricted specificity. *International Journal of Cancer*, 39:297–303, 1987.

[108] P. Garin-Chesa, I. Campbell, P. E. Saigo, J. L. Lewis, Jr., L. J. Old, and W. J. Rettig. Trophoblast and ovarian cancer antigen lk26. sensitivity and specificity in immunopathology and molecular identification as a folate-binding protein. *American Journal of Pathology*, 142:557–567, 1993.

[109] T. A. Patrick, D. M. Kranz, T. A. van Dyke, and E.J. Roy. Folate receptors as potential therapeutic targets in choroid plexus tumors of sv40 transgenic mice. *Journal of Neuro-Oncology*, 32:111–123, 1997.

[110] S. D. Weitman, A. G. Weinberg, L. R. Coney, V. R. Zurawski, D. S. Jennings, and B. A. Kamen. Cellular localization of the folate receptor: potential role in drug toxicity and folate homeostasis. *Cancer Research*, 52:6708–6711, 1992.

[111] W. J. Rettig, C. Cordon-Cardo, J. P. Koulos, J. L. Lewis, H. F. Oettgen, and L. J. Old. Cell surface antigens of human trophoblast and choriocarcinoma defined by monoclonal antibodies. *International Journal of Cancer*, 35:469–475, 1985.

[112] W. A. Franklin, M. Waintrub, D. Edwards, K. Christensen, P. Prendegrast, J. Woods, P. A. Bunn, and J. F. Kolhouse. New anti-lung-cancer antibody cluster 12 reacts with human folate receptors present on adenocarcinoma. *International Journal of Cancer - Supplement*, 8:89–95, 1994.

[113] J. Selhub and W. A. Franklin. The folate-binding protein of rat kidney. purification, properties, and cellular distribution. *Journal of Biological Chemistry*, 259:6601–6606, 1984.

[114] J. Selhub, D. Emmanouel, T. Stavropoulos, and R. Arnold. Renal folate absorption and the kidney folate binding protein. i. urinary clearance studies. *American Journal of Physiology*, 252:F750–F756, 1987.

[115] J. Selhub, S. Nakamura, and F. A. Carone. Renal folate absorption and the kidney folate binding protein. ii. microinfusion studies. *American Journal of Physiology*, 252:F757–F760, 1987.

[116] J. T. Hjelle, E. I. Christensen, F. A. Carone, and J. Selhub. Cell fractionation and electron microscope studies of kidney folate-binding protein. *American Journal of Physiology*, 260:C338–C346, 1991.

[117] H. Birn, S. Nielsen, and E. I. Christensen. Internalization and apical-to-basolateral transport of folate in rat kidney proximal tubule. *American Journal of Physiology*, 272:F70–F78, 1997.

[118] M. D. Kennedy, K. Jallad, J. Lu, P. S. Low, and D. Ben-Amotz. Evaluation of folate conjugate uptake and transport by the choroid plexus of mice. *Pharmaceutical Research*, 20:714–719, 2003.

[119] D. Wu and W. M. Pardridge. Blood-brain barrier transport of reduced folic acid. *Pharmaceutical Research*, 16:415–419, 1999.

[120] S. D. Weitman, K. M. Frazier, and B. A. Kamen. The folate receptor in central nervous system malignancies of childhood. *Journal of Neuro-Oncology*, 21:107–112, 1994.

[121] C. J. Mathias, S. Wang, R. J. Lee, D. J. Waters, P. S. Low, and M. A. Green. Tumor-selective radiopharmaceutical targeting via receptor-mediated endocytosis of gallium-67-deferoxamine-folate. *Journal of Nuclear Medicine*, 37:1003–1008, 1996.

[122] C. J. Mathias, S. Wang, D. J. Waters, J. J. Turek, P. S. Low, and M. A. Green. Indium-111-DTPA-folate as a potential folate-receptor-targeted radiopharmaceutical. *Journal of Nuclear Medicine*, 39:1579–1585, 1998.

[123] C. J. Mathias, S. Wang, P. S. Low, D. J. Waters, and M. A. Green. Receptor-mediated targeting of 67Ga-deferoxamine-folate to folate-receptor-positive human KB tumor xenografts. *Nuclear Medicine and Biology*, 26:23–25, 1999.

[124] S. D. Konda, M. Aref, M. Brechbiel, and E. C. Wiener. Development of a tumor-targeting MR contrast agent using the high-affinity folate receptor: work in progress. *Investigative Radiology*, 35:50–57, 2000.

[125] M. D. Kennedy, K. N. Jallad, D. H. Thompson, D. Ben-Amotz, and P. S. Low. Optical imaging of metastatic tumors using a folate-targeted fluorescent probe. *Journal of Biomedical Optics*, 8:636–641, 2003.

[126] C.-H. Tung, Y. Lin, W. K. Moon, and R. Weissleder. A receptor-targeted near-infrared fluorescence probe for *in vivo* tumor imaging. *ChemBioChem*, 3:784–786, 2002.

[127] A. Torricelli, A. Pifferi, P. Taroni, E. Giambattistelli, and R. Cubeddu. *In vivo* optical characterization of human tissues from 610 to 1010 nm by time-resolved reflectance spectroscopy. *Phys. Med. Biol.*, 46(8):2227–2237, August 2001.

[128] R. M. P. Doornbos, R. Lang, M. C. Aalders, F. W. Cross, and H. J. C. M. Sterenborg. The determination of *in vivo* human tissue optical properties and absolute chromophore concentrations using spatially resolved steady-state diffuse reflectance spectroscopy. *Phys. Med. Biol.*, 44(4):967–981, April 1999.

[129] A. J. Devaney and G. A. Tsihrintzis. Maximum likelihood estimation of object location in diffraction tomography. *IEEE Trans. Signal Processing*, 39(3):672–682, Mar. 1991.

[130] J. C. Ye, Y. Bresler, and P. Moulin. Cramèr-Rao bounds for 2-D target shape estimation in nonlinear inverse scattering problems with application to passive radar. 49(5):771–783, May 2001.

[131] G. Boverman. *Modeling and Nonlinear Inversion for Frequency Domain Diffuse Optical Tomography.* Master's thesis, Northeastern University, Boston, MA, June 2003.

[132] E. L Miller and A. S. Willsky. Multiscale, statistical anomaly detection analysis and algorithms for linearized inverse scattering problems. *Multidimensional Systems and Signal Processing*, 8:151–184, 1995.

[133] H. L. van Trees. *Detection, Estimation, and Modulation Theory, Part I.* John Wiley and Sons, New York, 1968.

[134] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley Publishing Co., New York, 1990.

[135] J. Luo, M.D. Smith, D. A. Lantrip, S. Wang, and P. L. Fuchs. Efficient synthesis of pyrofolic acid and pteroyl a azide, reagents for the production of carboxyl differentiated derivatives of folic acid. *J. Am. Chem. Soc.*, 119(42):10004–10013, 1997.

[136] S. Achilefu, R. Dorshow, J. Bugaj, and R. Rajagopalan. Novel receptor-targeted fluorescent contrast agents for *in vivo* tumor imaging. *Invest. Radiol.*, 35:479–485, 2000.

[137] G. Wagnières, S. Cheng, M. Zellweger, N. Utke, D. Braichotte, J. Ballini, and H. van der Bergh. An optical phantom with tissue-like properties in the visible for use in PDT and fluorescence spectroscopy. *Phys. Med. Biol.*, 42:1415–1426, July 1997.

[138] A. D. Gift, J. Ma, K. S. Haber, B. L. McClain, and D. Ben-Amotz. Near-infrared Raman imaging microscope based on fiber-bundle image compression. *J. Raman Spectroscopy*, 30:757–765, 1999.

[139] B. L. McClain, H. G. Hedderich, A. D. Gift, D. Zhang, K. N. Jallad, K. S. Haber, J. Ma, and D. Ben-Amotz. Fast chemical imaging: a rapid, noninvasive tool for medical, materials, and process analyses. *Spectroscopy*, 15(9):28–37, Sept. 2000.

[140] M. G. Erickson, J. S. Reynolds, and K. J. Webb. Comparison of sensitivity for single and dual interfering source configurations in optical diffusion imaging. *J. Optical Society America A*, 14(11):3080–3092, November 1997.

[141] G. Strangman, D. A. Boas, and J. P. Sutton. Non-invasive neuroimaging using near-infrared light. *Biol. Psychiatry*, 52:679–693, 2002.

[142] M. A. Franceschini and D. A. Boas. Noninvasive measurement of neuronal activity with near-infrared optical imaging. *NeuroImage*, 21(1):372–386, 2004.

[143] J. P. Allebach. Analysis of sampling-pattern dependence in time-sequential sampling of spatiotemporal signals. *J. Opt. Soc. Am. A*, 71(1):99–105, Jan. 1981.

[144] N. P. Willis and Y. Bresler. Optimal scan for time-varying tomography I: Theoretical analysis and fundamental limitations. *IEEE Trans. Image Processing*, 4(5):642–653, May 1995.

[145] B. D. O. Anderson. *Optimal filtering*. Prentice-Hall, Englewood Cliffs, NJ, 1979.

[146] A. Li, E. L. Miller, M. E. Kilmer, T. J. Brukilacchio, T. Chaves, J. Stott, Q. Zhang, Tao Wu, M. Chorlton, R. H. Moore, D. B. Kopans, and D. A. Boas. Tomographic optical breast imaging guided by three-dimensional mammography. *Appl. Opt.*, 42(25):5181–5190, Sept. 1 2003.

[147] H. Dehghani, B. W. Pogue, S. P. Poplack, and K. D. Paulsen. Multiwavelength three-dimensional near-infrared tomography of the breast: initial simulation, phantom, and clinical results. *Appl. Opt.*, 42(1):135–144, Jan. 1 2003.

[148] Y. Xu, X. Gu, L. L. Fajardo, and H. Jiang. *In vivo* breast imaging with diffuse optical tomography based on higher-order diffusion equations. *Appl. Opt.*, 42(16):3163–3169, June 1 2003.

[149] A. Pifferi, P. Taroni, A. Torricelli, F. Messina, and R. Cubeddu. Four-wavelength time-resolved optical mammography in the 680-980-nm range. *Opt. Lett.*, 28(13):1138–1140, July 1 2003.

[150] A. Li, Q. Zhang, J. P. Culver, E. L. Miller, and D. A. Boas. Reconstructing chromosphere concentration images directly by continuous-wave diffuse optical tomography. *Opt. Lett.*, 29(3):256–258, Feb. 1 2004.

[151] Q. Zhu, T. Durduran, V. Ntziachristos, M. Holboke, and A. G. Yodh. Imager that combines near-infrared diffusive light and ultrasound. *Opt. Lett.*, 24(15):1050–1052, Aug. 1 1999.

[152] A. H. Barnett, J. P. Culver, A. G. Sorensen, A. Dale, and D. A. Boas. Robust inference of baselinie optical properties of the human head with three-dimensional segmentation from magnetic resonance imaging. *Appl. Opt.*, 42(16):3095–3108, June 1 2003.

[153] S. Sastry and R. E. Carson. Multimodality Bayesian algorithm for image reconstruction in positron emission tomography: a tissue composition model. *IEEE Trans. Med. Imag.*, 16(6):750–761, Dec. 1997.

[154] C.T. Chen, X. Ouyang, W.H. Wong, X. Hu, V.E. Johnson, C. Ordonez, and C.E. Metz. Sensor fusion in image reconstruction. *IEEE Trans. on Nuclear Science*, 38:687–692, 1991.

[155] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, November 1984.

[156] S. R. Arridge. Photon-measurement density functions. Part 1: Analytical forms. *Applied Optics*, 34(31):7395–7409, November 1995.

[157] W. C. Chew. *Waves and Fields in Inhomogeneous Media*. Van Nostrand Reinhold, New York, 1990.

APPENDICES

APPENDICES

Appendix A: Fréchet Derivatives

Here we describe the computation of the Fréchet derivatives of the forward operators used in this study. Let $g(r_{src}, r_{obs}; x)$ be the diffusion equation Green's function for the problem domain computed using the image vector x and a numerical forward solver, with r_{src} as the source location and r_{obs} as the observation point. In addition, suppose that for a particular image x there are K sources and M detectors, and a total of $P = KM$ measurements. Let r_{s_k} represent the position of the k^{th} source and let $r_{d_{m'}}$ represent the position of the $\{m'\}^{th}$ detector. (Here, we use the letter m to denote detector number, as in our previous publications [24, 25, 27, 28], but with a prime mark to avoid confusion with the fluorescence emission subscript.) It follows that the computed data vector $f(x)$ is given by

$$f(x) = \begin{bmatrix} g(r_{s_1}, r_{d_1}; x) \\ g(r_{s_1}, r_{d_2}; x) \\ \vdots \\ g(r_{s_1}, r_{d_M}; x) \\ g(r_{s_2}, r_{d_1}; x) \\ \vdots \\ g(r_{s_K}, r_{d_M}; x) \end{bmatrix}. \tag{A.1}$$

For image vectors x of size $2N$, the Fréchet derivative is the $P \times 2N$ complex matrix given by

$$f'(x) = \begin{bmatrix} \frac{\partial g(r_{s_1}, r_{d_1}; x)}{\partial x_1} & \frac{\partial g(r_{s_1}, r_{d_1}; x)}{\partial x_2} & \cdots & \frac{\partial g(r_{s_1}, r_{d_1}; x)}{\partial x_{2N-1}} & \frac{\partial g(r_{s_1}, r_{d_1}; x)}{\partial x_{2N}} \\ \frac{\partial g(r_{s_1}, r_{d_2}; x)}{\partial x_1} & \frac{\partial g(r_{s_1}, r_{d_2}; x)}{\partial x_2} & \cdots & \frac{\partial g(r_{s_1}, r_{d_2}; x)}{\partial x_{2N-1}} & \frac{\partial g(r_{s_1}, r_{d_2}; x)}{\partial x_{2N}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial g(r_{s_1}, r_{d_M}; x)}{\partial x_1} & \frac{\partial g(r_{s_1}, r_{d_M}; x)}{\partial x_2} & \cdots & \frac{\partial g(r_{s_1}, r_{d_M}; x)}{\partial x_{2N-1}} & \frac{\partial g(r_{s_1}, r_{d_M}; x)}{\partial x_{2N}} \\ \frac{\partial g(r_{s_2}, r_{d_1}; x)}{\partial x_1} & \frac{\partial g(r_{s_2}, r_{d_1}; x)}{\partial x_2} & \cdots & \frac{\partial g(r_{s_2}, r_{d_1}; x)}{\partial x_{2N-1}} & \frac{\partial g(r_{s_2}, r_{d_1}; x)}{\partial x_{2N}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial g(r_{s_K}, r_{d_M}; x)}{\partial x_1} & \frac{\partial g(r_{s_K}, r_{d_M}; x)}{\partial x_2} & \cdots & \frac{\partial g(r_{s_K}, r_{d_M}; x)}{\partial x_{2N-1}} & \frac{\partial g(r_{s_K}, r_{d_M}; x)}{\partial x_{2N}} \end{bmatrix} \quad \text{(A.2)}$$

For the absorption and scattering coefficients, the discrete representations of the Fréchet derivative matrix elements have been derived and reported previously [26, 156] as

$$\frac{\partial g(r_{s_k}, r_{d_{m'}}; x)}{\partial \mu_a(r_i)} \simeq -g(r_{d_{m'}}, r_i; x) g(r_{s_k}, r_i; x) V \quad \text{(A.3)}$$

$$\frac{\partial g(r_{s_k}, r_{d_{m'}}; x)}{\partial D(r_i)} \simeq -\nabla g(r_{d_{m'}}, r_i; x) \cdot \nabla g(r_{s_k}, r_i; x) V \ , \quad \text{(A.4)}$$

where $\simeq$ is used due to domain discretization errors, V is the voxel volume, r_i is the position of the i^{th} voxel, and reciprocity [157] (which allows replacement of $g(r_{src}, r_{obs}; x)$ with $g(r_{obs}, r_{src}; x)$) has been used to reduce computation. Here, ∇ is the spatial gradient operator, which, in our computations, is evaluated numerically as a symmetric first difference. The separability of (A.3) and (A.4) with respect to source index and detector index enables additional savings in computation and in storage [28]. Rather than creating the entire $KM \times 2N$ matrix, it suffices to

initially compute and store two Green's function matrices of sizes $K \times N$ and $M \times N$, respectively:

$$G^{(s)} = \begin{bmatrix} g(r_{s_1}, r_1; x) & \cdots & g(r_{s_1}, r_N; x) \\ \vdots & \ddots & \vdots \\ g(r_{s_K}, r_1; x) & \cdots & g(r_{s_K}, r_N; x) \end{bmatrix} \tag{A.5}$$

$$G^{(d)} = \begin{bmatrix} g(r_{d_1}, r_1; x) & \cdots & g(r_{d_1}, r_N; x) \\ \vdots & \ddots & \vdots \\ g(r_{d_M}, r_1; x) & \cdots & g(r_{d_1}, r_N; x) \end{bmatrix}. \tag{A.6}$$

During the ICD scan, when the i^{th} voxel of x is to be modified, the i^{th} column of $f'(x)$ can be formed from (A.5) and (A.6).

For the fluorescence problem, more specific notation is needed. We define the expression $g_x(r_{src}, r_{obs}; x_x)$ to denote the λ_x Green's function obtained by solving (3.1), and let $g_m(r_{src}, r_{obs}; x_m)$ denote the λ_m Green's function obtained by solving (3.3). We denote the Green's function matrices accordingly:

$$G_x^{(s)} = \begin{bmatrix} g_x(r_{s_1}, r_1; x_x) & \cdots & g_x(r_{s_1}, r_N; x_x) \\ \vdots & \ddots & \vdots \\ g_x(r_{s_K}, r_1; x_x) & \cdots & g_x(r_{s_K}, r_N; x_x) \end{bmatrix} \tag{A.7}$$

$$G_x^{(d)} = \begin{bmatrix} g_x(r_{d_1}, r_1; x_x) & \cdots & g_x(r_{d_1}, r_N; x_x) \\ \vdots & \ddots & \vdots \\ g_x(r_{d_M}, r_1; x_x) & \cdots & g_x(r_{d_M}, r_N; x_x) \end{bmatrix} \tag{A.8}$$

$$G_m^{(s)} = \begin{bmatrix} g_m(r_{s_1}, r_1; x_m) & \cdots & g_m(r_{s_1}, r_N; x_m) \\ \vdots & \ddots & \vdots \\ g_m(r_{s_K}, r_1; x_m) & \cdots & g_m(r_{s_K}, r_N; x_m) \end{bmatrix} \tag{A.9}$$

$$G_m^{(d)} = \begin{bmatrix} g_m(r_{d_1}, r_1; x_m) & \cdots & g_m(r_{d_1}, r_N; x_m) \\ \vdots & \ddots & \vdots \\ g_m(r_{d_M}, r_1; x_m) & \cdots & g_m(r_{d_M}, r_N; x_m) \end{bmatrix}. \tag{A.10}$$

Consider one reparameterization of the right hand side of (3.2):

$$\eta \mu_{a_f}(r) \frac{1 - j\omega\tau(r)}{1 + [\omega\tau(r)]^2} = \beta_R(r) - j\beta_I(r). \tag{A.11}$$

It follows immediately that the inverse problem for β_R and β_I is linear. We define the expression $g_f(r_{src}, r_{obs}; x_x, x_m)$ to denote the fluorescence observed at r_{obs} emitted in response to excitation at r_{src}. The Fréchet derivatives for β_I and β_R are given by

$$\frac{\partial g_f(r_{s_k}, r_{d_{m'}}; x_x, x_m)}{\partial \beta_R(r_i)} \simeq g_m(r_{d_{m'}}, r_i; x_m) g_x(r_{s_k}, r_i; x_x) V \quad \text{(A.12)}$$

$$\frac{\partial g_f(r_{s_k}, r_{d_{m'}}; x_x, x_m)}{\partial \beta_I(r_i)} \simeq -j g_m(r_{d_{m'}}, r_i; x_m) g_x(r_{s_k}, r_i; x_x) V. \quad \text{(A.13)}$$

It is possible to solve the fluorescence inverse problem using this parameterization, and then convert the result into the physical parameters $\eta\mu_{a_f}$ and τ. However, the computation of τ requires a division of β_I by β_R, an operation which could result in large noise artifacts in regions where β_R is small. To circumvent this problem, we use the γ and τ parameterization of (3.8), permitting us to apply regularization directly to τ. In our sequential update scheme, τ is assumed constant while updates of γ are performed, and vice versa. As a result, we use the following Fréchet derivative expressions:

$$\frac{\partial g_f(r_{s_k}, r_{d_{m'}}; x_x, x_m)}{\partial \gamma(r_i)} \simeq g_m(r_{d_{m'}}, r_i; x_m) g_x(r_{s_k}, r_i; x_x)(1 - j\omega\hat{\tau}(r_i)) V \quad \text{(A.14)}$$

$$\frac{\partial g_f(r_{s_k}, r_{d_{m'}}; x_x, x_m)}{\partial \tau(r_i)} \simeq -j\omega\hat{\gamma}(r_i) g_m(r_{d_{m'}}, r_i; x_m) g_x(r_{s_k}, r_i; x_x) V. \quad \text{(A.15)}$$

After the reconstructions $\hat{x}_x$ and $\hat{x}_m$ are obtained, $G_x(r_s, r; x_x)$ and $G_m(r_d, r; x_m)$ have already been stored, and the Green's functions of (A.14) and (A.15) need not be recomputed. As the estimates $\hat{\gamma}$ and $\hat{\tau}$ are updated, they are incorporated into the derivative expressions.

Appendix B: Pseudocode for Fluorescence Optical Tomography Inversion Algorithm

main {

1. Initialize $\hat{x}_x$, $\hat{x}_m$, and $\hat{x}_f$ with background estimates.
2. Repeat until converged: {
   (a) $\hat{\alpha}_x \leftarrow \frac{1}{P_x} \; || \; y_x - f_x(\hat{x}_x) \; ||^2_{\Lambda_x}$
   (b) For $k = 1 : K$ {
   Compute $g_x(r_{s_k}, r; \hat{x}_x)$ by solving Eq. (3.1) with source at r_{s_k}
   }
   (c) For $m' = 1 : M$ {
   Compute $g_x(r_{d_{m'}}, r; \hat{x}_x)$ by solving Eq. (3.1) with source at $r_{d_{m'}}$
   }
   (d) Form $G_x^{(s)}$ and $G_x^{(d)}$ using Eq. (A.7) and Eq. (A.8)
   (e) $\hat{x}_x \leftarrow$ ICD_update$(\hat{x}_x, \hat{\alpha}_x, G_x^{(s)}, G_x^{(d)})$
   }
3. Repeat until converged: {
   (a) $\hat{\alpha}_m \leftarrow \frac{1}{P_m} \; || \; y_m - f_m(\hat{x}_m) \; ||^2_{\Lambda_m}$
   (b) For $k = 1 : K$ {
   Compute $g_m(r_{s_k}, r; \hat{x}_m)$ by solving Eq. (3.3) with source at r_{s_k}
   }
   (c) For $m' = 1 : M$ {
   Compute $g_m(r_{d_{m'}}, r; \hat{x}_m)$ by solving Eq. (3.3) with source at $r_{d_{m'}}$
   }
   (d) Form $G_m^{(s)}$ and $G_m^{(d)}$ using Eq. (A.9) and Eq. (A.10)
   (e) $\hat{x}_m \leftarrow$ ICD_update$(\hat{x}_m, \hat{\alpha}_m, G_m^{(s)}, G_m^{(d)})$
4. Repeat until converged: {
   (a) $\hat{\alpha}_f \leftarrow \frac{1}{P_f} \; || \; y_f - f_f(\hat{x}_f, \hat{x}_x, \hat{x}_m) \; ||^2_{\Lambda_f}$
   (b) $\hat{x}_f \leftarrow$ ICD_update$(\hat{x}_f, \hat{\alpha}_f, G_x^{(s)}, G_m^{(d)})$
   }

}

$\hat{x} \leftarrow$ ICD_update$(\hat{x}, \hat{\alpha}, G^{(s)}, G^{(d)}; x))$ {

1. For $i = 1, \ldots, N$ (in random order), {
   (a) Compute $[f'(\hat{x})]_{*(i)}$, as described in Appendix A
   (b) Update x_i, as described by Ye *et al.* [24]:

$$\hat{x}_i \leftarrow \arg\min_{x_i \geq 0} \left\{ \frac{1}{\hat{\alpha}} \left|\left| y - f(\hat{x}) - [F'(\hat{x})]_{*(i)} (x_i - \hat{x}_i) \right|\right|_{\Lambda}^2 + \frac{1}{p_a {\sigma_a}^{p_a}} \sum_{j \in \mathcal{N}_i} b_{i-j} |x_i - \hat{x}_i|^{p_a} \right\}$$

   }
2. For $i = N+1, \ldots, 2N$ (in random order), {
   (a) Compute $[f'(\hat{x})]_{*(i)}$, as described in Appendix A
   (b) Update x_i, as described by Ye *et al.* [24]:

$$\hat{x}_i \leftarrow \arg\min_{x_i \geq 0} \left\{ \frac{1}{\hat{\alpha}} \left|\left| y - f(\hat{x}) - [F'(\hat{x})]_{*(i)} (x_i - \hat{x}_i) \right|\right|_{\Lambda}^2 + \frac{1}{p_b {\sigma_b}^{p_b}} \sum_{j \in \mathcal{N}_i} b_{i-j} |x_i - \hat{x}_i|^{p_b} \right\}$$

   }
3. Return $\hat{x}$.

}

Appendix C: Pseudocode for Multifrequency FODT Inversion Algorithm

main {

1. Form $G_x^{(s)}$ and $G_m^{(d)}$
2. Repeat until converged: {
   (a) $\hat{\alpha}_f \leftarrow \frac{1}{P_f} \; || \; y_f - f_f(\hat{x}_f, \hat{x}_x, \hat{x}_m) \; ||^2_{\Lambda_f}$
   (b) $\hat{x}_f \leftarrow \text{ICD\_update}(\hat{x}_f, \hat{\alpha}_f, G_x^{(s)}, G_m^{(d)})$
   }

}

$\hat{x} \leftarrow \text{ICD\_update}(\hat{x}, \hat{\alpha}, G^{(s)}, G^{(d)}; x))$ {

1. For $q = 1, \ldots, Q$ {
   (a) $z_{\omega_q} \leftarrow y_{\omega_q} - f_{\omega_q}(\hat{x})$
   }
2. For $i = 1, \ldots, N$ (in random order), {
   (a) $\tilde{x}_i \leftarrow \hat{x}_i$
   (b) For $q = 1, \ldots, Q$ {
      i. Compute $[J_{\omega_q}]_{*(i)}$, by taking the i^{th} column in Eq. (4.12)
      ii. $\theta_{1,\omega_q} \leftarrow -2\text{Re}\left\{[J_{\omega_q}]_{*i}^H \Lambda_{\omega_q} z_{\omega_q}\right\}$
      iii. $\theta_{2,\omega_q} \leftarrow 2[J_{\omega_q}]_{*i}^H \Lambda_{\omega_q} [J_{\omega_q}]_{*i}$
      }
   (c) $\hat{x}_i \leftarrow \arg\min_{x_i \geq 0} \{$
   $\frac{1}{\hat{\alpha}} \sum_{q=1}^{Q} \Big(\theta_{1,\omega_q}[h(x, r_i, \omega_q) - h(\tilde{x}, r_i, \omega_q)]$
   $+\frac{\theta_{2,\omega_q}}{2}[h(x, r_i, \omega_q) - h(\tilde{x}, r_i, \omega_q)]^2\Big)$
   $+\frac{1}{\rho_A \sigma_A{}^{\rho_A}} \sum_{j \in \mathcal{N}_i} b_{i-j} \, |x_i - \tilde{x}_j|^{\rho_A} \Big\}$
   (d) For $q = 1, \ldots, Q$ {
      i. $z_{\omega_q} \leftarrow z_{\omega_q} + [J_{\omega_q}]_{*i}[h(\hat{x}, r_i, \omega_q) - h(\tilde{x}, r_i, \omega_q)]$

}

}

3. For $i = N+1, \ldots, 2N$ (in random order), {

   (a) $\tilde{x}_i \leftarrow \hat{x}_i$

   (b) For $q = 1, \ldots, Q$ {

      i. Compute $[J_{\omega_q}]_{*(i-N)}$, by taking the $(i-N)^{th}$ column in Eq. (4.12)

      ii. $\theta_{1,\omega_q} \leftarrow -2\mathrm{Re}\left\{[J_{\omega_q}]^H_{*(i-N)} \Lambda_{\omega_q} z_{\omega_q}\right\}$

      iii. $\theta_{2,\omega_q} \leftarrow 2[J_{\omega_q}]^H_{*(i-N)} \Lambda_{\omega_q} [J_{\omega_q}]_{*(i-N)}$

      }

   (c) $\hat{x}_i \leftarrow \arg\min_{x_i \geq 0}$ {

   $\frac{1}{\hat{\alpha}} \sum_{q=1}^{Q} \Big(\theta_{1,\omega_q}[h(x, r_{i-N}, \omega_q) - h(\tilde{x}, r_{i-N}, \omega_q)]$

   $+\frac{\theta_{2,\omega_q}}{2}[h(x, r_{i-N}, \omega_q) - h(\tilde{x}, r_{i-N}, \omega_q)]^2\Big)$

   $+\frac{1}{\rho_B \sigma_B{}^{\rho_B}} \sum_{j \in \mathcal{N}_i} b_{i-j} \left|x_i - \tilde{x}_j\right|^{\rho_B}\Big\}$

   (d) For $q = 1, \ldots, Q$ {

      i. $z_{\omega_q} \leftarrow z_{\omega_q} + [J_{\omega_q}]_{*(i-N)}[h(\hat{x}, r_{i-N}, \omega_q) - h(\tilde{x}, r_{i-N}, \omega_q)]$

      }

   }

}

Appendix D: Mutual Information Derivation

Here, we derive the expression for the mutual information which is is given in Eq. (4.26). The mutual information given by Eq. (4.21) 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. (3.10),(3.13),(4.24), and (4.25). Define the shot noise process $Z = Y - JX$, and assume that Z is independent of X. Then $Y = JX + Z$ is Gaussian, and

$$E[Y] = E[E[Y|X]] = E[JX] = JE[X] = 0. \tag{D.1}$$

In addition,

$$\begin{aligned} E[YY^H] &= E[(JX+Z)(JX+Z)^H] && \text{(D.2)} \\ &= E[ZZ^H] + E[(JX)Z^H] + E[Z(JX)^H] + E[JXX^HJ^H] && \text{(D.3)} \\ &= E[ZZ^H] + JE[XX^H]J^H && \text{(D.4)} \\ &= \alpha\Lambda^{-1} + \sigma^2 JC^{-1}J^H. && \text{(D.5)} \end{aligned}$$

Hence,

$$p_Y(y) = \frac{1}{(\pi)^P|\Upsilon|}\exp\left\{-||y||^2_{\Upsilon^{-1}}\right\}. \tag{D.6}$$

where

$$\Upsilon = \alpha\Lambda^{-1} + \sigma^2 JC^{-1}J^H. \tag{D.7}$$

Using Eqs. (4.22) and (D.6), $H(Y)$ is given by [65]

$$H(Y) = \log(\pi)^P + P + \log|\Upsilon|. \tag{D.8}$$

Similarly, using Eq. (4.23),

$$H(Y|X) = \log(\pi)^P + P + \log|\alpha\Lambda^{-1}|. \tag{D.9}$$

Substituting Eqs. (D.8) and (D.9) into Eq. (4.21)

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= \log \frac{|\Upsilon|}{|\alpha\Lambda^{-1}|} \\ &= \log \frac{|\alpha\Lambda^{-1} + \sigma^2 JC^{-1}J^H|}{|\alpha\Lambda^{-1}|} \\ &= \log \left| I + \frac{\sigma^2}{\alpha}\Lambda JC^{-1}J^H \right| \end{aligned} \tag{D.10}$$

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

Appendix E: Pseudocode for Kinetic Model Inversion Algorithm

main {

1. Form $G_x^{(s)}$ and $G_m^{(d)}$
2. Repeat until converged: {
   (a) $\hat{\alpha}_f \leftarrow \frac{1}{P_f} \; || \; y_f - f_f(\hat{x}_f, \hat{x}_x, \hat{x}_m) \; ||^2_{\Lambda_f}$
   (b) $\hat{x}_f \leftarrow \text{ICD\_update}(\hat{x}_f, \hat{\alpha}_f, G_x^{(s)}, G_m^{(d)})$
   }

}

$\hat{x} \leftarrow \text{ICD\_update}(\hat{x}, \hat{\alpha}, G^{(s)}, G^{(d)}; x))$ {

1. For $c = 1, \ldots, C$ {
   (a) For $q = 1, \ldots, Q$ {
      i. $z_{\omega_q, t_c} \leftarrow y_{\omega_q, t_c} - f_{\omega_q, t_c}(\hat{x})$
      }
   }
2. For $i = 1, \ldots, N$ (in random order), {
   (a) $\tilde{x}_i \leftarrow \hat{x}_i$
   (b) For $c = 1, \ldots, C$ {
      i. For $q = 1, \ldots, Q$ {
         A. Compute $[J_{\omega_q, t_c}]_{*(i)}$, by taking the i^{th} column in Eq. (4.12)
         B. $\theta_{1,\omega_q,t_c} \leftarrow -2\text{Re}\left\{[J_{\omega_q,t_c}]^H_{*i} \Lambda_{\omega_q,t_c} z_{\omega_q,t_c}\right\}$
         C. $\theta_{2,\omega_q,t_c} \leftarrow 2[J_{\omega_q,t_c}]^H_{*i} \Lambda_{\omega_q,t_c} [J_{\omega_q,t_c}]_{*i}$
         }
      }
   (c) $\hat{\tau} \leftarrow \arg\min_{x_{(0),i} \geq 0} \{$
   $\frac{1}{\hat{\alpha}} \sum_{c=1}^{C} \sum_{q=1}^{Q} \left(\theta_{1,\omega_q,t_c}[h(x_{(*),i}, \omega_q, t_c) - h(\tilde{x}_{(*),i}, \omega_q, t_c)]\right.$

$$+\frac{\theta_{2,\omega_q,t_c}}{2}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]^2\Big)$$
$$+\frac{1}{\rho_{(0)}\sigma_{(0)}{}^{\rho_{(0)}}}\sum_{j\in\mathcal{N}_i}b_{i-j}\left|x_{(0),i}-\tilde{x}_{(0),j}\right|^{\rho_{(0)}}\Big\}$$

(d) $\hat{\gamma}_1 \leftarrow \arg\min_{x_{(1),i}\geq 0}\{$
$$\tfrac{1}{\hat{\alpha}}\sum_{c=1}^{C}\sum_{q=1}^{Q}\Big(\theta_{1,\omega_q,t_c}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]$$
$$+\frac{\theta_{2,\omega_q,t_c}}{2}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]^2\Big)$$
$$+\frac{1}{\rho_{(1)}\sigma_{(1)}{}^{\rho_{(1)}}}\sum_{j\in\mathcal{N}_i}b_{i-j}\left|x_{(1),i}-\tilde{x}_{(1),j}\right|^{\rho_{(1)}}\Big\}$$

(e) $\hat{\gamma}_2 \leftarrow \arg\min_{x_{(2),i}\geq 0}\{$
$$\tfrac{1}{\hat{\alpha}}\sum_{c=1}^{C}\sum_{q=1}^{Q}\Big(\theta_{1,\omega_q,t_c}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]$$
$$+\frac{\theta_{2,\omega_q,t_c}}{2}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]^2\Big)$$
$$+\frac{1}{\rho_{(2)}\sigma_{(2)}{}^{\rho_{(2)}}}\sum_{j\in\mathcal{N}_i}b_{i-j}\left|x_{(2),i}-\tilde{x}_{(2),j}\right|^{\rho_{(2)}}\Big\}$$

(f) if $(\hat{\gamma}_2 > \hat{\gamma}_1)\{$

$$\hat{\gamma}_1,\hat{\gamma}_2 \leftarrow \arg\min_{x_{(1),i}=x_{(2),i}\geq 0}\{$$
$$\tfrac{1}{\hat{\alpha}}\sum_{c=1}^{C}\sum_{q=1}^{Q}\Big(\theta_{1,\omega_q,t_c}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]$$
$$+\frac{\theta_{2,\omega_q,t_c}}{2}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]^2\Big)$$
$$+\frac{1}{\rho_{(1)}\sigma_{(1)}{}^{\rho_{(1)}}}\sum_{j\in\mathcal{N}_i}b_{i-j}\left|x_{(1),i}-\tilde{x}_{(1),j}\right|^{\rho_{(1)}}$$
$$+\frac{1}{\rho_{(2)}\sigma_{(2)}{}^{\rho_{(2)}}}\sum_{j\in\mathcal{N}_i}b_{i-j}\left|x_{(2),i}-\tilde{x}_{(2),j}\right|^{\rho_{(2)}}\Big\}$$

}

(g) $\hat{\gamma}_3 \leftarrow \arg\min_{x_{(3),i}\geq 0}\{$
$$\tfrac{1}{\hat{\alpha}}\sum_{c=1}^{C}\sum_{q=1}^{Q}\Big(\theta_{1,\omega_q,t_c}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]$$
$$+\frac{\theta_{2,\omega_q,t_c}}{2}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]^2\Big)$$
$$+\frac{1}{\rho_{(3)}\sigma_{(3)}{}^{\rho_{(3)}}}\sum_{j\in\mathcal{N}_i}b_{i-j}\left|x_{(3),i}-\tilde{x}_{(3),j}\right|^{\rho_3}\Big\}$$

(h) $\hat{\gamma}_4 \leftarrow \arg\min_{x_{(4),i}\geq 0}\{$
$$\tfrac{1}{\hat{\alpha}}\sum_{c=1}^{C}\sum_{q=1}^{Q}\Big(\theta_{1,\omega_q,t_c}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]$$

$$+\frac{\theta_{2,\omega_q,t_c}}{2}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]^2\Big)$$
$$+\frac{1}{\rho_{(4)}\sigma_{(4)}{}^{\rho_{(4)}}}\sum_{j\in\mathcal{N}_i}b_{i-j}\left|x_{(4),i}-\tilde{x}_{(4),j}\right|^{\rho_{(4)}}\Big\}$$

(i) if $(\hat{\gamma}_3 > \hat{\gamma}_4)\{$

$$\hat{\gamma}_3,\hat{\gamma}_4 \leftarrow \arg\min_{x_{(3),i}=x_{(4),i}\geq 0}\{$$
$$\tfrac{1}{\hat{\alpha}}\sum_{c=1}^{C}\sum_{q=1}^{Q}\left(\theta_{1,\omega_q,t_c}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]\right.$$
$$\left.+\frac{\theta_{2,\omega_q,t_c}}{2}[h(x_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]^2\right)$$
$$+\frac{1}{\rho_{(3)}\sigma_{(3)}{}^{\rho_{(3)}}}\sum_{j\in\mathcal{N}_i}b_{i-j}\left|x_{(3),i}-\tilde{x}_{(3),j}\right|^{\rho_{(3)}}$$
$$+\frac{1}{\rho_{(4)}\sigma_{(4)}{}^{\rho_{(4)}}}\sum_{j\in\mathcal{N}_i}b_{i-j}\left|x_{(4),i}-\tilde{x}_{(4),j}\right|^{\rho_{(4)}}\Big\}$$

}

(j) $[\hat{x}_{(0),i},\hat{x}_{(2),i},\hat{x}_{(3),i},\hat{x}_{(4),i}] \leftarrow [\hat{\tau},\hat{\gamma}_1,\hat{\gamma}_2,\hat{\gamma}_3]$

(k) For $c = 1,\ldots,C$ {

i. For $q = 1,\ldots,Q$ {

A. $z_{\omega_q,t_c} \leftarrow z_{\omega_q,t_c} + [J_{\omega_q,t_c}]_{*i}[h(\hat{x}_{(*),i},\omega_q,t_c)-h(\tilde{x}_{(*),i},\omega_q,t_c)]$

}

}

}

}

Appendix F: Partial Derivatives for Cramèr-Rao Bound

Here, we compute the partial derivatives required in (6.25). We first rewrite $f_i(r)$:

$$f_i(r) = [a_i(r) - b_i(r)]^2, \tag{F.1}$$

where a_i and b_i are the first and second terms in the brackets of (6.3), respectively. The partial derivatives are as follows:

$$\frac{\partial f_i}{\partial x} = 2\left[a_i - b_i\right]\left[\frac{\partial a_i}{\partial x} - \frac{\partial b_i}{\partial x}\right] \tag{F.2}$$

$$\frac{\partial f_i}{\partial z} = 2\left[a_i - b_i\right]\left[\frac{\partial a_i}{\partial z} - \frac{\partial b_i}{\partial z}\right] \tag{F.3}$$

$$\frac{\partial a_i}{\partial x} = -(x - x_{s_i})a_i\left[r_a^{-2} + kr_a^{-1}\right] \tag{F.4}$$

$$\frac{\partial a_i}{\partial z} = -(z - z_{s_i})a_i\left[r_a^{-2} + kr_a^{-1}\right] \tag{F.5}$$

$$\frac{\partial b_i}{\partial x} = -(x - x_{s_i})b_i\left[r_b^{-2} + kr_b^{-1}\right] \tag{F.6}$$

$$\frac{\partial b_i}{\partial z} = -(z + z_{s_i} + 2l_s)b_i\left[r_b^{-2} + kr_b^{-1}\right] \tag{F.7}$$

Appendix G: Pseudocode for Computing $H(Y)$ in Dynamic Optical Diffusion Tomography

For $a = 1, \ldots, T$ {

  For $b = 1, \ldots, a$ {

    $W_{ab}^{\Delta x} \leftarrow J_a C_{\Delta x} J_b^H$

    $W_{ba}^{\Delta x} \leftarrow \left(W_{ab}^{\Delta x}\right)^H$

    $W_{ab}^{1} \leftarrow J_a C_{1|0} J_b^H$

    $W_{ba}^{1} \leftarrow \left(W_{ab}^{1}\right)^H$

  } /* end b */

} /* end a */

$\tilde{H} \leftarrow \log |W_{11}^1 + \Gamma_1|$

For $i = 2, \ldots, T$ {

  For $a = i, \ldots, T$ {

    For $b = i, \ldots, a$ {

      $W_{ab}^{i} \leftarrow W_{ab}^{i-1} - W_{a(i-1)}^{i-1} (W_{(i-1)(i-1)}^{(i-1)} + \Gamma_{i-1})^{-1} W_{(i-1)b}^{i-1} + W_{ab}^{\Delta x}$

      $W_{ba}^{i} \leftarrow \left(W_{ab}^{i}\right)^H$

    } /* end b */

  } /* end a */

$\tilde{H} \leftarrow \tilde{H} + \log |W_{ii}^i + \Gamma_i|$

} /* end i */

VITA

VITA

Adam Milstein began his undergraduate studies in Electrical Engineering at Purdue University in Fall of 1996. He worked five semesters as a cooperative education student at the Johnson Space Center in Houston, TX. In May 2000, he received his BSEE degree with highest distinction. In Fall 2000, he began his PhD graduate studies at Purdue University in the field of Electrical Engineering.