Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 201 (2004) 771–797
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

A fast numerical method for time-resolved photon diffusion in general stratified turbid media

Alex H. Barnett *

Courant Institute of Mathematical Sciences, New York University, 251 Mercer St, New York, NY 10012, USA

Received 24 October 2003; received in revised form 17 June 2004; accepted 30 June 2004
Available online 13 August 2004

Abstract

Efficient forward models of photon migration in complex geometries are important for noninvasive imaging of tissue in vivo with diffuse optical tomography (DOT). In particular, solving the inverse problem requires multiple solutions of the forward model and is therefore computationally intensive. We present a numerical algorithm for the rapid solution of the time-dependent diffusion equation in a semi-infinite inhomogeneous medium whose scattering and absorption coefficients are arbitrary functions of depth, given a point source impulsive excitation. Such stratified media are biomedically important. A transverse modal representation leads to a series of one-dimensional diffusion problems which are solved via finite-difference methods. A novel time-stepping scheme allows effort to scale independently of total time (for fixed system size). Tayloring to the DOT application gives run times of order 0.1 s. We study convergence, computational effort, and validate against known solutions in the case of 2-layer media. The method will be useful for other forward and inverse diffusion problems, such as heat conduction and conductivity measurement.

MSC: 65M06; 65M12; 78A70; 80M20; 92C55
Keywords: Parabolic; Diffusion; Photon migration; Diffuse optical tomography; Time-resolved; Inverse problem; Stratified

1. Introduction

Diffuse optical tomography (DOT) [1,9] is receiving growing attention as a noninvasive method for functional imaging of living tissue. Near-infrared light propagates diffusively inside tissue for long distances, allowing surface measurements to be sensitive to absorption at depths of order centimeters. In this wavelength range the small absorption is dominated by the spectra of oxy- and deoxy-hemoglobin. The large

* Corresponding author. Tel.: +1-212-998-3296.
E-mail address: barnett@cims.nyu.edu.

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2004.06.017

difference between the two spectra allows reconstruction of independent spatial maps of oxy- and deoxy-hemoglobin, and therefore of total blood volume and oxygenation, from absorption images at multiple wavelengths. Metabolic activity and hemodynamic response give a window on tissue function, providing information not accessible to other imaging modalities. This, combined with the ability to map the scattering coefficient and a sensitivity to optical contrast agents such as fluorescent dyes [2], results in a wealth of applications. On the clinical side they include breast tumor screening [4], neonatal cerebral monitoring [3] and arthritis assessment [6]. On the functional side they include neuroimaging [2,1] and muscle oximetry [5].

Reconstructing absorption and scattering coefficients from surface flux measurements is an inverse problem [9]. When optical inhomogeneity is large, or when baseline measurements are desired, model-based nonlinear methods [10,13,20,15,36,14] are optimal. They require many solutions to the forward problem, namely the simulation of flux measurements at certain distances from the sources given an optical model of the underlying tissue with certain parameter values. These parameter values are then adjusted iteratively [20,15,36] or sampled in a Bayesian statistical fashion [14] to solve the inverse problem. This can require hundreds or more forward model solutions. Surface light flux measurements may be continuous-wave (unmodulated), radio-frequency modulated, or time-resolved. Of these three measurement types, the latter is considered to provide the most information [1]. The resulting reflectance signal measured at various distances from each impulsive point source is called a *temporal point spread function* (TPSF). Therefore, to simulate such signals an efficient time-dependent forward solution is needed.

In this paper, we provide such a solution for the diffusion approximation (DA) to photon transport in a layered medium. Layered media are ubiquitous in the biomedical field – for instance they approximate the local geometry of scalp, skull, and brain, or skin, fat, and muscle – and recently a growing need for fast layered solvers has been identified [32–34]. The fact that DOT makes use of point-like sources and detectors at various separations means that the problem has cylindrical symmetry rather than being simply one-dimensional (1D). The symmetry allows a much more rapid solution than is possible in general tissue geometries where 3D finite element [9] or finite difference [20,15,14] methods are required. The DA, whose validity is reviewed in Section 2, is orders of magnitude faster to simulate than the more physically accurate radiative transfer equation (RTE). We note that gains due to the symmetry also extend to the RTE, for instance using layered Monte Carlo [18,19], however its simulation remains an intensive task.

Our framework takes advantage of the symmetry of the medium (Fig. 1): using separation of variables into planar coordinate $\boldsymbol{\rho} \equiv (x, y)$ and depth coordinate z, general fluence solutions to the DA can be built up from solutions with a definite in-plane 2D wavevector $\mathbf{k}$. Furthermore, the cylindrical symmetry of the

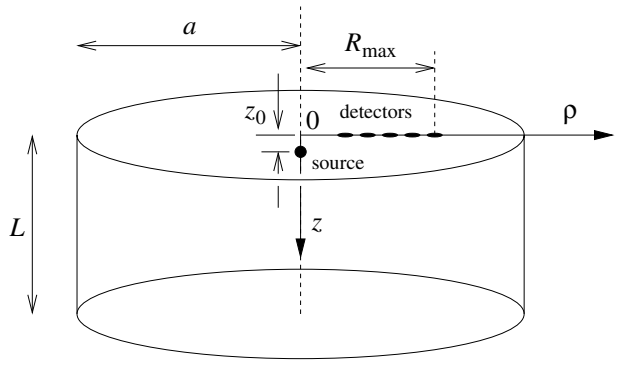

Fig. 1. Coordinate system, and geometry of pill-box used to model the semi-infinite medium. The source fiber (not shown) lies along the negative z-axis; the effective DA point source is shown as a large dot. Detector fibers (not shown) lie parallel to the negative z-axis at the various radii R shown. As explained in Section 4, the pill-box is chosen to be large enough (~100 mm) that the resulting signals differ negligibly from those due to a semi-infinite medium (infinite depth and radius).

source term means that only the 1-parameter family of radially-symmetric solutions is needed (namely the $J_0(k\rho)$ Bessel function parametrized by wavenumber $k = |\mathbf{k}|$, where the radius is $\rho = |\boldsymbol{\rho}|$). At each k there remains to be solved a diffusive 1D partial differential equation (PDE) in z and time. If the system is of infinite extent in the radial coordinate ρ then k is a continuous parameter; integrating the contributions from different k then corresponds to a Hankel transform from k to ρ. This idea is well-known in the literature on multilayer (piecewise homogeneous) stratified media. For instance in the 2-layer medium it is used by Dayan et al. [30] to derive approximate solutions, and, in the case of steady-state solutions, by Kienle et al. [32] who evaluate the Hankel transform via numerical quadrature (the 1D steady-state PDE having an analytic solution).

Our approach, which as far as we are aware is new, involved first truncating the system's radius to a finite value large enough to induce negligible error (because of the nature of the diffusion equation this error can be made exponentially small). Therefore, the integral over k becomes a sum over discrete (quantized) wavenumbers k_m. The choice of method for the 1D evolution of each mode m is very flexible: we choose finite-difference methods. Computational efficiency is achieved (a) by carefully considering the errors and requirements of the time-resolved DOT application, (b) by introducing a novel timestepping scheme appropriate for impulsively-excited parabolic problems (this contrasts the uniform timestepping currently used both in the DOT field [13,20,24,14] and more generally), and (c) by minimizing the length of time for which each mode m is evolved. The last of these relies on the fact that errors due to truncation in the Fourier domain are also exponentially small (see Fig. 3(b)). The desired TPSF signals are then extracted in the usual way from the resulting solution at the surface.

We now categorize some of the different choices made by other authors. In our method sketched above the fluence function $\phi(\rho, z, t)$ is written as a sum of modes $\psi_m(\rho)u_m(z, t)$, which we could call a *transverse mode approach*. The other possibility is a sum of modes $\xi_l(\rho, t)v_l(z)$, or *longitudinal mode approach*, where the roles of ρ and z have been reversed. The latter has been used for steady-state solutions in 2- and 3-layer media [28,29], and time-resolved solutions [34] based upon them. The longitudinal eigenvalues are found by numerically solving a transcendental equation (root-finding), and the $\xi_l(\rho, t)$ obey the analytically-solvable 2D homogeneous diffusion equation. However, a finite total depth is needed to ensure that all eigenvalues are discrete, and computational effort increases with eigenvalue density which is in turn proportional to total depth. Therefore, the approach is ill-adapted to infinitely-thick media. Quantization in both $\boldsymbol{\rho}$ and z is used in the eigenfunction method of Martelli et al. [35], which handles finite media but requires triple summation.

Other approaches include the remarkable analytic expansion of Tualle et al. [33], which relies on a volume source representation of the solution due to a planar medium interface to generate very fast 2-layer solutions. We also note that integral equation techniques lend themselves naturally to the solution of PDEs in piecewise-homogeneous media – in particular since the problem of history-dependence of time-dependent diffusion has been elegantly overcome [47,48] – however we know of no biomedical applications of these promising techniques.

In contrast to all the above layered approaches which are based on boundary matching, our choice of 1D evolution method allows us to model media with arbitrary depth-dependent absorption and scattering coefficients, hence we can tackle an arbitrary number of layers, and continuously-varying coefficients, with no extra complexity. This may be important for tissues such as the skull which appears to have scattering properties which vary continuously with depth.

The remainder of this paper is structured as follows. In Section 2, we present the well-known DA model and discuss its validity. In Section 3, we bring together the requirements placed on our method by the DOT application; this will prove important for maximizing numerical efficiency. Emulation of the infinite system by one of finite size is presented, along with numerical convergence tests in Section 4. The core of the method, which solves the DA in this finite-sized system, is explained in Section 5. Section 5.5 provides an algorithm summary. Numerical convergence of the finite-difference method, and validation against an existing

method, is presented in Section 6. The scaling of computational effort is analyzed in detail in Section 7. We discuss some issues, limitations, and future directions in Section 8, and conclude in Section 9.

2. The diffusion approximation model

Incoherent mono-energetic photon passage through tissue is described by a Boltzmann-type RTE giving the time-evolution of the radiance $\Phi(\mathbf{r}, \boldsymbol{\Omega}, t)$ (photon density in the 5-dimensional phase-space comprising position $\mathbf{r} \equiv (x, y, z)$ and solid angle $\boldsymbol{\Omega}$). By retaining only the first (dipole) term in the angular dependence, and assuming this dependence is small [11,9], we get the DA,

$$\frac{1}{c}\frac{\partial \phi}{\partial t} = \nabla \cdot (\kappa \nabla \phi) - \mu_a \phi + q, \tag{1}$$

where $\phi(\mathbf{r}, t)$ is the fluence (the angular integral of radiance), $q(\mathbf{r}, t)$ is the source term, $\kappa(z)$ is the local diffusion constant (in units of length), $\mu_a(z)$ is the local absorption (in units of inverse length), and c is the speed of light in the medium, defined as the speed in vacuo divided by n the refractive index. κ is related to reduced scattering coefficient μ_s' by the good approximation [16,9] $\kappa = 1/(3\mu_s')$. It is the reduced dimensionality (fluence is a function of 3 spatial dimensions as opposed to 5 for radiance) that makes the DA easier to solve than the RTE. We remind the reader of the coordinate system used for what follows: $z = 0$ defines the air-tissue interface with z increasing with depth inside the medium, and $\boldsymbol{\rho} \equiv (x, y)$ is the in-plane coordinate whose length is the (cylindrical) radial distance ρ (see Fig. 1). We reserve the symbol R to refer to distance of a detector from the source. Note that in this work we have spatial dependence of κ and μ_a but not of n (allowing this would require only simple changes to the finite difference scheme).

Given a source–detector separation R, the DA is a good approximation to the RTE when several conditions are satisfied [39,9,16]: that $\mu_a \ll \mu_s'$ everywhere, that $1/\mu_s'$ is much smaller than both R and any significant geometrical features, that times much greater than R/c are considered, and that appropriate boundary conditions and optode models, discussed below, are used. The size of the error involved in using the DA is a topic of ongoing investigation: it has found to be acceptably small for time-resolved reflectance of layered media [32], although biomedical situations where the steady-state fluence is inaccurate may be found [16]. In particular, the DA breaks down in the clear cerebro-spinal fluid (CSF) layer between the skull and brain, however there exist ways [21,40] to handle this within the DA, and there is recent evidence that the effective μ_s' of CSF is only a few times less than that of surrounding tissue [22].

The best choice of boundary condition for an air-tissue interface appropriate for the time-domain setting is a topic of recent research [38,12,25,31,37]. The extrapolated boundary conditions

$$\phi(\mathbf{r}, t)|_{z=-z_b} = 0, \quad \forall \boldsymbol{\rho}, t, \tag{2}$$

have proved to be reasonable [37]. With $n \approx 1.4$ for tissue [41] the extrapolation distance, taking into account Fresnel reflection [25], is $z_b \approx 1.96 z_0$. Here the scattering depth is $z_0 \equiv 1/\mu_s'(0) = 3\kappa_0$, written in terms of the diffusion constant in the neighborhood of the surface $\kappa_0 \equiv \lim_{z \to 0^+} \kappa(z)$.

The impulsive source q is only nonzero at $t = 0$ (in this work we ignore the finite width of the impulse response of optical fibers and instrumentation since this can be modeled easily by convolution in time). Therefore, we can treat Eq. (1) as an initial-boundary value problem with no source term ($q = 0$). We take the initial conditions appropriate for a narrow normally-incident optical fiber source in contact with the skin to be

$$\phi(\mathbf{r}, 0) = \delta(x)\delta(y)\delta(z - z_0), \tag{3}$$

the commonly-used isotropic DA point source model [25] (see Fig. 1). We note that, alternatively, a gradient source at $z = 0$ (that is, an inhomogeneous Neumann boundary condition) would also be easy to implement within our scheme, requiring only an appropriate change to Eq. (19).

We define $I(R, t)$ to be the detected signal (TPSF) at a source–detector distance R and time t. The best way to extract this from the fluence solution in the time-domain has also recently been discussed [25,31,37]. The general linear combination,

$$I(R,t) = \kappa_0 \left[A \frac{\phi(R,t)}{z_b} + (1-A) \frac{\partial \phi}{\partial z}(R,t) \right] \Bigg|_{z=0}, \tag{4}$$

with $A \in [0, 1]$, covers all possibilities and, by varying A, can account for arbitrary detector fiber numerical aperture. The weighting is chosen so that in the long-time limit the signal is independent of A. Unless stated we will choose $A = 1$, since this is known to perform well (in [37] this choice is named 'R3'). Other choices of boundary condition and measurement type are easy to implement within our method.

3. Requirements of our numerical method

In order to maximize numerical efficiency, it is important to list the requirements which arise in the biomedical DOT application, which we do in this section. Our method is currently designed and used for fitting signals collected by time-correlated single photon counting (TCSPC) apparatus [7,8,1], a popular choice for time-resolved DOT. The goal is fast computation of $I(R, t)$ for each source–detector separation R, to an accuracy sufficient for the fitting process. The first four requirements listed, involving the system parameters, arise generally in time-resolved biomedical DOT. We will show later that requirements (ii) and (iv) have quite a strong influence on the total effort needed for computation; see Eq. (50). The last requirement, namely the choice of acceptable error level, also has a strong effect on total effort, thus we have matched this error level in more detail to our TCSPC fitting application. All five requirements are expected to have analogues in other diffusion applications, such as heat conduction.

(i) *Source-detector separation range*: At each R, we define the time-resolved peak signal $I_{\text{peak}}(R) \equiv \max_t I(R,t)$. The decay of $I_{\text{peak}}(R)$ with R is rapid: for example in a homogeneous medium with zero absorption, $I_{\text{peak}}(R) \sim R^{-5}$. Thermal safety considerations, fiber area and detector sensitivity limit useful source–detector distances of interest to $R_{\max}$, which is of order 40 mm. It transpires that our numerical scheme is insensitive to any minimum R, so we set $R \in [0, R_{\max}]$.
(ii) *Time range of interest*: At early times the DA does a poor job of approximating the RTE [39]. Therefore, there is a time $t_{\min} \sim 100$ ps [17,23,31] before which it is not appropriate [43] to compute the signal. After a time $t_{\max}$ of order a few thousand ps, the signal drops below detectable levels. Therefore, $I(R, t)$ is needed only for $t \in [t_{\min}, t_{\max}]$.
(iii) *Maximum depth of medium*: We are interested in both finite-thickness and semi-infinite media. In this work, we tackle the semi-infinite case, which is more challenging because it requires a decision about how to truncate the medium's infinite spatial extent while keeping errors acceptable (see Section 4).
(iv) *Optical parameter ranges*: Given a medium we define the inhomogeneity ratio $\alpha \equiv \kappa_{\max}/\kappa_{\min} = \mu'_{s,\max}/\mu'_{s,\min}$, where $\kappa_{\max}$ and $\kappa_{\min}$ are the maximum and minimum of $\kappa(z)$. For biomedical applications α rarely exceeds 10, and is usually much less. However, α is not small enough that perturbative methods of solution are possible. For instance, the range of μ'_s reported in in vivo human head tissues, excluding CSF, is 0.6–2.5 mm^{-1} [14,16,42,44]. There is recent evidence [22] that even the clear CSF has an effective μ'_s of 0.16–0.32 mm^{-1}, in which case including it in the DA model causes α hardly to exceed the above limit. Because $\mu_a \ll \mu'_s$ within the DA it will transpire that there are effectively no further constraints on absorption values.
(v) *Acceptable error as a function of time*: The basic principle we apply is that forward model errors (numerical and physical) should not exceed a small fraction of the detection apparatus measurement errors [45]. Given a signal $I(t)$ we define an acceptable absolute error level $I_{\text{err}}(t)$. If absolute numerical errors in

calculating $I(t)$ are below this baseline for all $t \in [t_{\min}, t_{\max}]$, then we define this to be acceptable accuracy. If relative error alone defined acceptability, this would correspond to $I_{\text{err}}(t) = \epsilon I(t)$, where ϵ sets the relative error. However, when signals are very small, such as at early times and large R, in practice relative measurement errors grow (or the signal is below detection levels), and this criterion becomes too stringent. We therefore propose a modified criterion which includes the Poisson shot-noise statistics of single photon counting [7], namely

$$I_{\text{err}}(t;\gamma,\epsilon) \equiv \max\left[\frac{I_{\text{peak}}}{\gamma}, \sqrt{\frac{I_{\text{peak}} I(t)}{\gamma}}, \epsilon I(t)\right], \tag{5}$$

where the max function returns the maximum of its three arguments. Note that for $I(t)$ sufficiently large this expression equals $\epsilon I(t)$. The first and second arguments together approximate the form of the standard deviation of the Poisson distribution ($\sigma \sim \sqrt{n}$ for n detected photons). I_{peak} is relevant because at a given R, in practice the apparatus gain is adjusted to prevent saturation at I_{peak}. The dynamic range of interest is set by γ. In our context, with $\sim 10^5$ photons detected per source–detector pair [14], and a dark count of $\sim 10^{-3} I_{\text{peak}}$, we believe $\gamma \sim 10^4$ or 10^5 is appropriate. The third argument is the original relative error criterion; typically we need $\epsilon \sim 10^{-2}$. Clearly larger γ or smaller ϵ demand smaller numerical errors; this will result in a slower forward solution. In Fig. 2, we show a typical signal $I(t)$ and the resulting acceptable error level $I_{\text{err}}(t)$.

4. Truncation of the spatial domain

We want to solve the initial-boundary value problem given by the sourceless version of Eq. (1), namely

$$\frac{1}{c}\frac{\partial \phi}{\partial t} = \nabla \cdot (\kappa \nabla \phi) - \mu_{\text{a}} \phi, \tag{6}$$

with boundary condition equation (2) and initial condition equation (3), in the semi-infinite slab $\rho \in [0, \infty]$, $z \in [-z_{\text{b}}, \infty]$. Note that the slight extension of the domain into negative z results from the extrapolated boundary condition. Here we give a recipe for truncating this domain while causing negligible changes to the desired signals $I(R, t)$. Although the problems of depth truncation and radial truncation can be treated independently due to the system's separability, they are formally similar so we treat them both here.

We truncate both spatial coordinates to create a finite cylindrical box $\rho \in [0, a]$, $z \in [-z_{\text{b}}, L]$ (see Fig. 1). We impose numerically-convenient (we choose Dirichlet) boundary conditions on the newly-created boundaries:

$$\phi(a, z, t) = 0 \quad \forall z, t, \tag{7}$$

$$\phi(\rho, L, t) = 0 \quad \forall \rho, t. \tag{8}$$

The key idea is that because the form of the time-dependent Green's function has Gaussian tails, the errors introduced by the box die exponentially with increasing box size, at fixed time. The spreading increases with time, therefore we should use $t = t_{\max}$ when bounding the errors. A one-dimensional schematic is given in Fig. 3(a).

We now formalize this more precisely in the similar case of a homogeneous medium of scattering coefficient κ and absorption coefficient μ_{a}, inside the *cuboid* box $x \in [-a, a]$, $y \in [-a, a]$, $z \in [-z_{\text{b}}, L]$. Because of exponential convergence, the general result will carry over to the cylindrical box case; we will also verify this numerically. The exact solution in semi-infinite space is the sum of a heat kernel and its image charge,

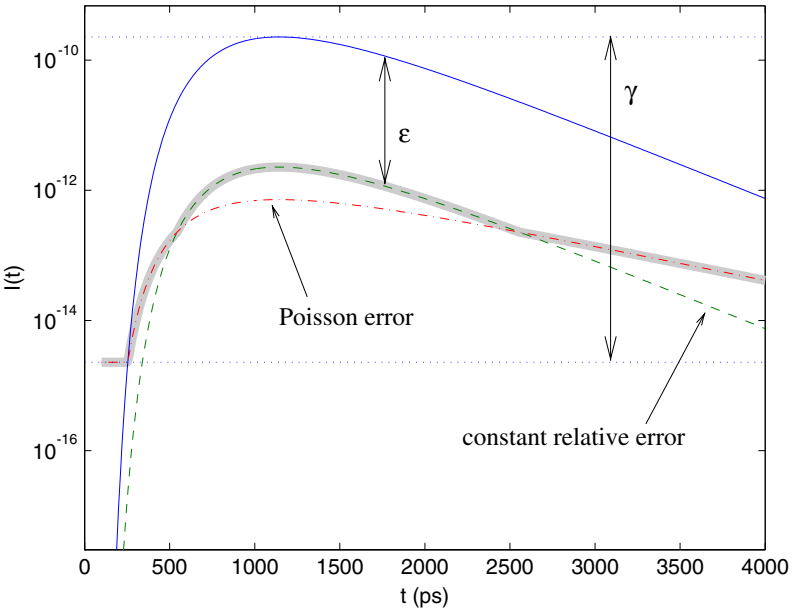

Fig. 2. Typical signal $I(t)$ (thin solid line), and its corresponding acceptable error level $I_{\mathrm{err}}(t)$ (fat grey solid line); see Section 3. $I_{\mathrm{err}}(t)$ is the larger of the Poisson error level (dash-dotted) and a constant relative error level (dashed). Here the relative error level dominates from roughly 500–2500 ps. The horizontal dotted lines show I_{peak} and I_{peak}/γ. The error parameters are a dynamic range of $\gamma = 10^5$ and relative error of $\epsilon = 10^{-2}$. This signal resulted from a homogeneous medium with $\mu_s' = 1$ mm^{-1}, $\mu_a = 0.01$ mm^{-1}, $R = 40$ mm.

$$\phi(\rho,z,t) = \frac{\mathrm{e}^{-\frac{\rho^2}{4\kappa ct}-\mu_a ct}}{(4\pi\kappa ct)^{3/2}}\left[\mathrm{e}^{-\frac{(z-z_0)^2}{4\kappa ct}} - \mathrm{e}^{-\frac{(z+2z_b+z_0)^2}{4\kappa ct}}\right]. \tag{9}$$

For large t we can Taylor expand the term in square brackets. The first nonzero term gives the dipole approximation

$$\phi_d(\rho,z,t;z_d) \approx d\frac{z-z_d}{t^{5/2}}\mathrm{e}^{-\frac{(z-z_d)^2+\rho^2}{4\kappa ct}-\mu_a ct}, \tag{10}$$

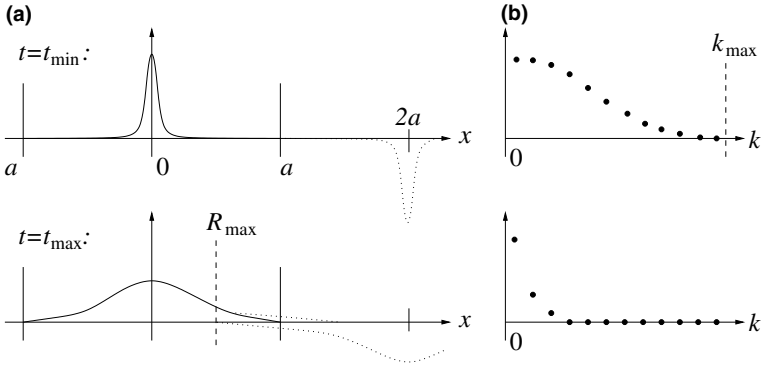

Fig. 3. One-dimensional example of real-space (a) and Fourier space (b) representation of the diffusion equation Green's function, or heat kernel, in a box. As the Gaussian width grows in real-space, it shrinks in Fourier space.

where the dipole strength is $d = (z_b + z_0)/(4\pi)^{3/2}(\kappa c)^{5/2}$ and its vertical location $z_d = -z_b$. We use Eq. (4) with $A = 0$ (the A-dependence is negligible) to give the signal detected at distance R_d in the xy-plane from a dipole vertically located at z_d,

$$I(R_d, t; z_d) = \kappa \frac{\partial \phi_d(R_d, z, t; z_d)}{\partial z}\bigg|_{z=0}. \tag{11}$$

The cuboid box case can be solved using the method of images [46], using an infinite periodic rectangular 3D lattice of image dipole sources. We are interested in the case where the relative errors are small; this corresponds to a heat kernel width $\sqrt{2\kappa c t}$ which is smaller than a or L. Therefore, the error size will be well approximated by considering only the nearest set of image dipoles (the remaining infinite sum contributing exponentially less, as can be proved by bounding the sum by an exact Gaussian integral [47]).

In the xy-plane the four closest image dipoles have distance $2a$ from the origin, and therefore distance $R_d \geqslant 2a - R$ from the detector. Since R, $R_d \gg z_b$, we can use $z_d \approx 0$. The relative signal error due to these image dipoles is bounded by

$$\epsilon_a \leqslant 4\frac{I(2a - R, t; 0)}{I(R, t; 0)} = 4e^{-\frac{a(a-R)}{\kappa c t}} \leqslant 4e^{-\frac{(a-R)^2}{\kappa c t}}. \tag{12}$$

Within our domain of interest this is largest at $R = R_{max}$ and $t = t_{max}$. Given a desired ϵ_a our choice of a is therefore

$$a \approx R_{max} + \sqrt{\kappa c t_{max} \ln(4/\epsilon_a)}. \tag{13}$$

We have found that the difference between the convergence of ϵ_a with a in the cuboid and cylindrical cases is slight. Therefore, Eq. (13) is also a reasonable choice of radius for the cylindrical box.

Similarly, along the z-axis the two closest image dipoles are at $z_d \approx \pm 2L$ and $R_d = R$, where we again have used $z_b \approx 0$. The relative signal error due to these two dipoles is

$$\epsilon_L \approx 2\frac{I(R, t; 2L)}{I(R, t; 0)} = 2\left(1 + \frac{2L^2}{\kappa c t}\right)e^{-\frac{L^2}{\kappa c t}}. \tag{14}$$

Again using $R = R_{max}$ and $t = t_{max}$, an approximate solution for L is

$$L \approx \sqrt{\kappa c t_{max} \ln[C(\epsilon_L)/\epsilon_L]}, \tag{15}$$

where $C(\epsilon_L) = 2[1 + 2\ln(1/\epsilon_L)]$.

Returning to our inhomogenous system, where no closed form for $\phi(\rho, z, t)$ is known, we need a recipe to choose the smallest safe values of a and L. We use the fact that the spatial spread is bounded (in the xy-plane) by a Gaussian of width $\sqrt{2\kappa_{max} c t}$. This leads us to use Eq. (13) with the crude substitution $\kappa = \kappa_{max}$. A similar consideration leads to Eq. (15) with the same substitution. In other words, we are using a homogeneous system of diffusion coefficient κ_{max} to bound the box size needed for an inhomogeneous system. When the inhomogeneity ratio α is large, these choices may be less than optimal, but because the Gaussian width scales like $\kappa^{1/2}$, the factor by which both a and L could be suboptimal is only $\alpha^{1/2}$ or at most about 3.

No rigorous analytic proof of the validity of the above box size bounds for the inhomogeneous case are known to the author, however Fig. 4 shows verification that these estimates are valid for various inhomogeneous two-layered systems with large inhomogeneity $\alpha = 10$. In practise the exponential convergence ensures that we can set ϵ_a and ϵ_L to values much smaller than the desired relative error ϵ with very little penalty: for the rest of this work we use $\epsilon_a = \epsilon_L = 10^{-6}$. In our application, the typical resulting box size is $a \sim L \sim 10^2$ mm.

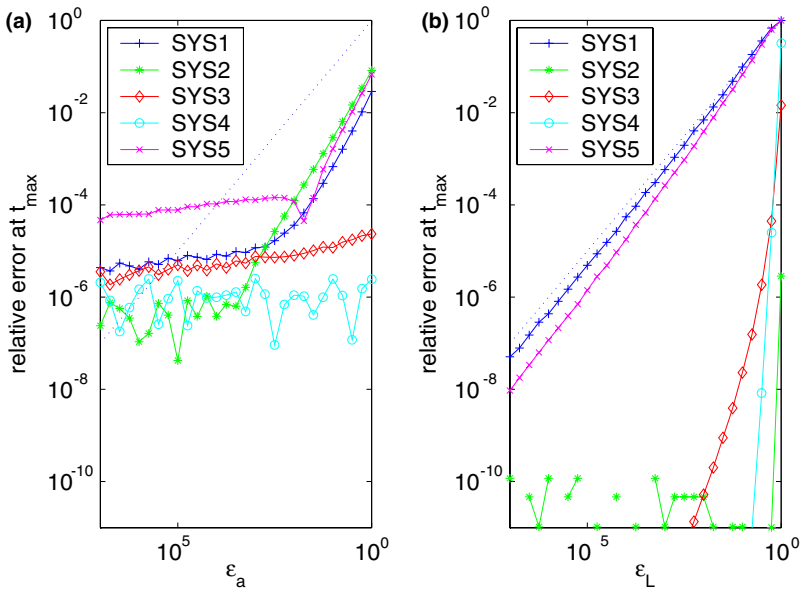

Fig. 4. Test of convergence of relative error at $R = 40$ mm (a typical $R_{\max}$), and $t = 5$ ns (a typical $t_{\max}$) due to box truncation. The five systems are: homogeneous (SYS1) and four inhomogeneous 2-layer systems with maximal range of κ (SYS2–5), as described in Table 1. (a) shows relative error using a given by Eq. (13) while $\epsilon_L = 10^{-15}$, and (b) shows the same using L given by Eq. (15) while $\epsilon_a = 10^{-15}$. The dotted lines show expected behavior if these estimates were correct. The systems are compared against a calculation with $\epsilon_a = \epsilon_L = 10^{-15}$. In (a) the bottoming out at a relative error of about 10^{-4} reflects the overall error of the method; because a changes it cannot be removed. In (b) λ_{shift} was fixed to allow comparison beyond this overall error level. Finite difference parameters were $h = 0.5$ mm, $\beta = 0.05$ (see Sections 5.2 and 5.3).

5. The method

5.1. Transverse mode representation

All solutions of Eq. (6) with the boundary condition equations (2), (7) and (8) can be written

$$\phi(\rho, z, t) = \sum_{m=1}^{M} B_m \psi_m(\rho) u_m(z, t), \tag{16}$$

which is exact in the limit $M \to \infty$, where the radial functions are normalized eigenmodes of the Laplacian in the disc of radius a in the xy-plane,

$$\psi_m(\rho) = \frac{1}{\sqrt{\pi} a J_1(\chi_m)} J_0(k_m \rho). \tag{17}$$

The transverse wavenumbers are given by $k_m = \chi_m/a$, where χ_m is the argument of the mth zero of the J_0 Bessel function. The coefficents corresponding to the correct initial condition equation (3) are $B_m = [\sqrt{\pi} a J_1(\chi_m)]^{-1}$, which can easily be shown by using $\delta(x)\delta(y) = \sum_{m=1}^{\infty} \psi_m(x, y)\psi_m(0, 0)$, a special case of the eigenfunction closure relation. We remind the reader that it is the symmetry of the initial condition that causes only radially-symmetric modes to contribute.

For each m the evolution of $u_m(z, t)$ obeys

$$\frac{1}{c}\frac{\partial u_m}{\partial t} = \frac{\partial}{\partial z}\left(\kappa \frac{\partial u_m}{\partial z}\right) - \mu_{a,m} u_m, \tag{18}$$

$$u_m(z,0) = \delta(z - z_0), \tag{19}$$

$$u_m(-z_b, t) = u_m(L,t) = 0, \quad \forall t, \tag{20}$$

with the effective m-dependent absorption coefficient

$$\mu_{a,m}(z) = \mu_a(z) + k_m^2 \kappa(z). \tag{21}$$

The signal is extracted using Eq. (4) applied to Eq. (16),

$$I(R,t) = \kappa_0 \sum_{m=1}^{M} B_m \psi_m(R) \left[A \frac{u_m(0,t)}{z_b} + (1-A) \frac{\partial u_m}{\partial z}(0,t) \right]. \tag{22}$$

We discuss choice of the number of modes M in Section 5.4.

5.2. Spatial finite-difference discretization

Currently the 1D problem $u_m(z, t)$ is discretized spatially using a finite-difference lattice of cell size h. A uniform lattice is used out of convenience; undoubtedly improved efficiency will result with a nonuniform lattice. Values of fluence $u_m^{(i)}$ are based at nodes (i.e. junctions between cells), labelled by $i = 0, 1, \ldots, n_z$. The extrapolated boundary conditions require that the first cell begins at $z = -z_b$, with the $i = 0$ node fluence value held at zero. A minimum number of cells n_z is chosen such that the last cell ends beyond the box depth $z = L$ (recall that L has been chosen in Section 4 such that the results are not sensitive to L). The $i = n_z$ node fluence value is held at zero to model the box boundary condition. In Eq. (18) the right-hand side is spatially discretized using a 3-point template to give the continuous time-evolution (often called the 'Method of Lines' [26]),

$$\frac{1}{c} \frac{\partial u_m^{(i)}}{\partial t} = \frac{\kappa^{(i-\frac{1}{2})}}{h^2} u_m^{(i-1)} - \left(\frac{\kappa^{(i-\frac{1}{2})} + \kappa^{(i+\frac{1}{2})}}{h^2} + \frac{\mu_{a,m}^{(i-\frac{1}{2})} + \mu_{a,m}^{(i+\frac{1}{2})}}{2} \right) u_m^{(i)} + \frac{\kappa^{(i+\frac{1}{2})}}{h^2} u_m^{(i+1)}, \quad i = 1, 2, \ldots, n_z - 1. \tag{23}$$

Here $\kappa^{(i-\frac{1}{2})}(\kappa^{(i+\frac{1}{2})})$ signifies a diffusion coefficient, and $\mu_{a,m}^{(i-\frac{1}{2})}(\mu_{a,m}^{(i+\frac{1}{2})})$ an effective absorption coefficient, within the cell to the left (right) of node i.

The cell-averaging is not entirely obvious. By considering the steady-state flux conservation law and Fick's law across a cell split into two homogeneous pieces with different κ, it is easy to show that μ_s', or the inverse of κ, is the correct quantity to average. On similar physical grounds μ_a, and therefore for each mode the effective $\mu_{a,m}$, is the correct quantity to average. Therefore, making use of Eq. (21),

Table 1
Homogeneous and 2-layer systems used for numerical tests

Name	Upper layer			Lower layer		
	μ_s'	μ_a	Thickness	μ_s'	μ_a	Thickness
SYS1	1.0	0.01	∞			
SYS2	0.3	0	10	3	0.03	∞
SYS3	3	0	10	0.3	0.03	∞
SYS4	0.3	0.03	10	3	0	∞
SYS5	3	0.03	10	0.3	0	∞
SYS6	0.5	0.01	10	2.0	0.003	∞

Absorption μ_a and reduced scattering coefficient μ_s' are in units of mm^{-1}, while thickness is in units of mm. The first four 2-layer systems cover the four possibilities of large and small absorption and diffusion for each layer, with a large choice of diffusion inhomogeneity $\alpha = 10$. All refractive indices are $n = 1.4$.

$$(\kappa^{(i+\frac{1}{2})})^{-1} \equiv \frac{1}{h}\int_{z_i}^{z_i+h} \kappa(z)^{-1}\,\mathrm{d}z, \tag{24}$$

$$\mu_{\mathrm{a},m}^{(i+\frac{1}{2})} \equiv \frac{1}{h}\int_{z_i}^{z_i+h} [\mu_{\mathrm{a}}(z) + k_m^2\kappa(z)]\mathrm{d}z, \tag{25}$$

where $z_i = \mathrm{i}h - z_{\mathrm{b}}$ is the depth of node i. Note the peculiarity that κ^{-1} is averaged for the diffusion term, but κ itself is averaged within the effective absorption term. The scheme is second-order accurate in h, as demonstrated in Section 6.1.

We can express the single mode evolution equation (23) using the column vector $\mathbf{u}_m \equiv \{u_m^{(i)} : i = 1, 2, \ldots, n_z - 1\}$ as the ordinary differential equation

$$\frac{\partial \mathbf{u}_m}{\partial t} = -S_m\mathbf{u}_m(t), \tag{26}$$

$$\mathbf{u}_m(0) = \mathbf{u}_0. \tag{27}$$

The initial condition is given by a standard linear spatial interpolation of $\delta(z - z_0)$ appearing in Eq. (19) onto the lattice. Similarly, the signal is extracted using linear spatial interpolations of $u_m(0, t)$ and $(\partial u_m/\partial z)(0, t)$ in Eq. (22) on the lattice. These linear interpolations preserve the overall $\mathrm{O}(h^2)$ accuracy.

Furthermore, we may express the whole discretized system as the N-component vector $\mathbf{U} \equiv [\mathbf{u}_1^{\mathrm{T}} \ldots \mathbf{u}_{\mathrm{M}}^{\mathrm{T}}]^{\mathrm{T}}$ which is simply the vectors $\mathbf{u}_m$ stacked vertically. Its evolution (that is, the Method of Lines for the whole system) is then

$$\frac{\partial \mathbf{U}}{\partial t} = -S\mathbf{U}(t), \tag{28}$$

$$\mathbf{U}(0) = \mathbf{U}_0. \tag{29}$$

The $N \times N$ matrix S is block diagonal, with tridiagonal blocks S_m. Signal extraction by Eq. (22) can now be summarized in the form of a projection onto a time-independent vector $\mathbf{P}(R)$,

$$I(R, t) = \mathbf{P}(R)^{\mathrm{T}}\mathbf{U}(t). \tag{30}$$

5.3. Fast time-stepping scheme

The exact solution to Eqs. (28) and (29) is

$$\mathbf{U}(t) = \mathrm{e}^{-St}\mathbf{U}_0 = V\mathrm{diag}(\mathrm{e}^{-\lambda_j t})V^{\mathrm{T}}\mathbf{U}_0, \tag{31}$$

where the matrix exponential has been written in terms of the spectral decomposition of S: the eigenvalues are λ_j and the eigenvector matrix is $V = (\mathbf{v}_1 \cdots \mathbf{v}_N)$, that is, eigenvectors $\mathbf{v}_j$ arranged in columns. Evaluating the evolution exactly in this manner would require computing the full decomposition. It is the job of any time-stepping scheme to approximate the important contributions to this evolution with much less numerical effort.

Traditionally in the DOT field, if a finite-difference method is used then the time-stepping has been uniform, using either forward Euler [14] or Crank–Nicolson (CN) [13,20,24,9] schemes. The reasoning that CN allows a longer time step Δt than forward Euler because it is guaranteed stable for arbitrarily large Δt is often repeated [9] but can be unwittingly abused. What is not commonly understood is that in the limit of stiff lattice modes (high decay rates λ) then CN only approaches *marginal* stability: in Fig. 5(a) the CN curve is asymptotic to -1. Technically CN is 'A-stable' but not 'L-stable' [26]. This manifests itself as

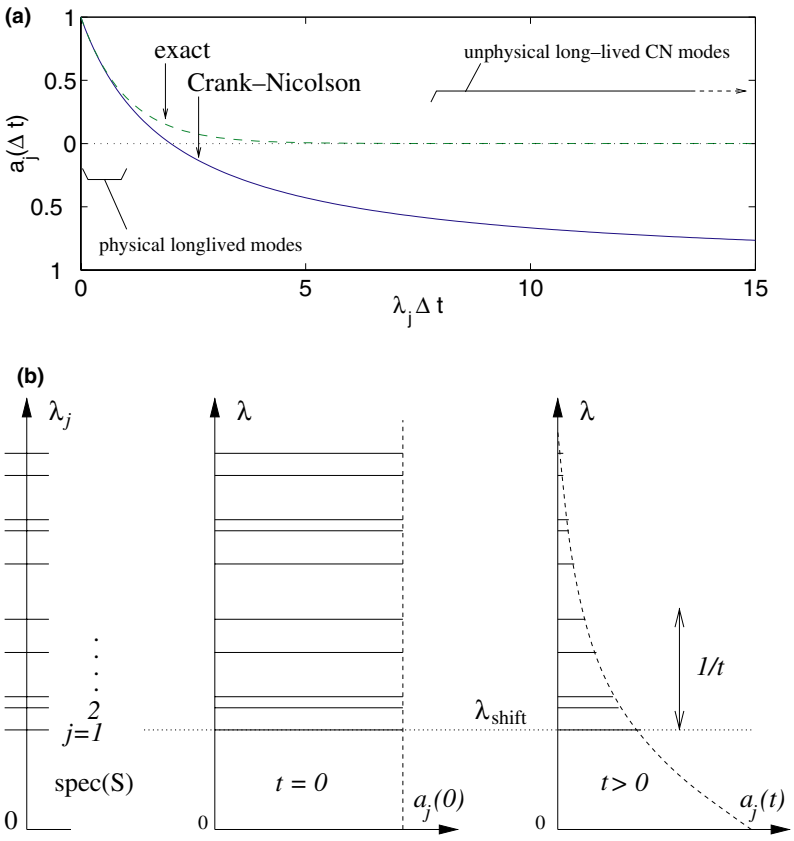

Fig. 5. (a) Decay of amplitude $a_j(\Delta t)$ of a mode j of the system due to a single Crank–Nicolson time-step Δt (solid line), compared to its exact decay $e^{-\lambda_j \Delta t}$ (dotted line). For convenience we took $a_j(0) = 1$. (b) Spectrum of system matrix S. Amplitude coefficients $a_j(0)$ at $t = 0$ are assumed to have similar sizes. At a later time t they have decayed as an exponential function of their eigenvalue λ_j.

unphysical extremely long decay times of the most oscillatory modes of the lattice, which can decay more slowly than the physical low-frequency modes, thereby causing an exponentially erroneous solution. (Contrary to one claim [51], these modes cannot easily be damped out without reducing accuracy or requiring a large increase in numerical effort). These oscillatory modes are strongly excited by a spatially-discontinuous initial condition such as Eq. (19). This fundamental problem has not been addressed in the DOT numerical literature. The scheme we present overcomes this problem while also vastly reducing computational effort.

Eq. (31) states that $\mathbf{U}(t)$ is the sum of eigenvectors with exponentially-decaying coefficients,

$$\mathbf{U}(t) = \sum_{j=1}^{N} a_j(t)\mathbf{v}_j. \tag{32}$$

The coefficients $a_j(t) = e^{-\lambda_j t} a_j(0)$ have initial values $a_j(0) = \mathbf{v}_j^{\mathrm{T}}\mathbf{U}_0$. Consider the spectrum of the system matrix S and the coefficient decay illustrated in Fig. 5(b). We assume (i) that the density of eigenvalues is approximately uniform along the λ axis (this is to be expected since it is the generic behavior for a wave eigenproblem in 2D [49]), and (ii) that all the $a_j(0)$ are of the same order of magnitude. At a fixed time $t > 0$ the decay $e^{-\lambda t}$ is exponential in λ. Therefore, at this time only the coefficients corresponding to eigenvalues

lying in the range from a minimum of λ_1 to a maximum of $\lambda_{\max}(t) \approx \lambda_1 + \mathrm{O}(1/t)$ are needed; dropping coefficients corresponding to λ above this maximum would result in exponentially small error. Thus we see that as t increases, a shrinking piece of the spectrum of size $\mathrm{O}(1/t)$ is relevant.

A single CN timestep of duration Δt performs approximate time-evolution via the formal replacement

$$\mathrm{e}^{-S\Delta t} = \frac{1 - \frac{1}{2}S\Delta t}{1 + \frac{1}{2}S\Delta t} + \mathrm{O}(\Delta t^3) + \cdots, \tag{33}$$

where $\mathrm{O}(\Delta t^3)$ and higher terms are ignored. In practice the matrix division is handled by solution of the tridiagonal linear system of equations, which requires $\mathrm{O}(N)$ computational effort [27]. The CN approximation in the eigenbasis can be written in terms the coefficients thus

$$a_j(\Delta t) \approx a_j^{\mathrm{CN}}(\Delta t) = \frac{1 - \frac{1}{2}\lambda_j\Delta t}{1 + \frac{1}{2}\lambda_j\Delta t} a_j(0). \tag{34}$$

This is plotted in Fig. 5(a): the approximation is good when $\lambda\Delta t \ll 1$, where the dominant error term is $-(\lambda\Delta t)^3/12$. For $\lambda\Delta t \gg 1$ the approximation is very poor. However, if we choose Δt such that $\lambda_{\max}(t)\Delta t < \mathrm{O}(1)$, then all the relevant coefficients will be evolved accurately. Coefficients corresponding to larger eigenvalues will have totally incorrect evolution, however we have already established above that this causes negligible error. This suggests a t-dependent (i.e. nonuniform) timestep $\Delta t(t) \sim [\lambda_1 + \mathrm{O}(1/t)]^{-1}$.

The long-time limit of $\Delta t(t)$ presented above is a constant, $1/\lambda_1$. This situation can easily be improved so that there is no upper limit. We shift the spectrum of S by defining

$$S' \equiv S - \lambda_{\mathrm{shift}} I, \tag{35}$$

with shift $\lambda_{\mathrm{shift}} = \lambda_1$. Now the minimum eigenvalue λ'_1 of S' is zero, and the relevant spectral range is from zero to $\lambda'_{\max}(t) \sim \mathrm{O}(1/t)$. Thus we have the modified scheme

$$\Delta t(t) = \beta t, \tag{36}$$

where the choice of constant $\beta \ll 1$ is discussed in Section 6.2. We call this 'logarithmic timestepping' since it is uniform in $\ln(t)$. We use CN with this timestepping scheme to approximate the evolution

$$\mathbf{U}'(t) = \mathrm{e}^{-S't}\mathbf{U}_O \tag{37}$$

in the usual way (namely Eq. (33) with the substitution S' for S). It then requires very little numerical effort to undo the shift:

$$\mathbf{U}(t) = \mathrm{e}^{-\lambda_{\mathrm{shift}}t}\mathbf{U}'(t). \tag{38}$$

In practice λ_{shift} can be computed (using the LAPACK routine DSTEVX [50]) in negligible $\mathrm{O}(n_z)$ time since λ_1 of S is simply the lowest eigenvalue of the matrix S_m for the first mode $m = 1$. For each CN step the tridiagonal system of equations is solved using LAPACK [50], separately for each block S_m comprising S.

This logarithmic scheme evolves modes of a given decay constant accurately during the time that they contribute, then inaccurately (without increasing their amplitudes) once they reach a negligible (exponentially small) contribution. However, we need a recipe to initialize the method, since Eq. (36) as it stands would require infinite numerical effort to handle small times $t \to 0$. One solution would be to start with an analytic solution for $\mathbf{U}(t)$ at some small $t = t_0$, and evolve from there. This would require simple forms for $\kappa(z)$ and $\mu_{\mathrm{a}}(z)$ in the uppermost part of the medium. Instead we decided to use an initial CN phase from $t = 0$ to $t = t_0$, with constant step size Δt_0 sufficiently small that even the highest decay-rate modes of the lattice decay to safe amplitudes. We call this phase X (see Fig. 6(b)). A reliable recipe for the choice of t_0 and Δt_0 is given in Appendix A.

This timestepping scheme is second-order accurate in β, as verified in Section 6.2. The exponential decay of modes means that either assumption (i) or (ii) would have to be very strongly violated to invalidate the

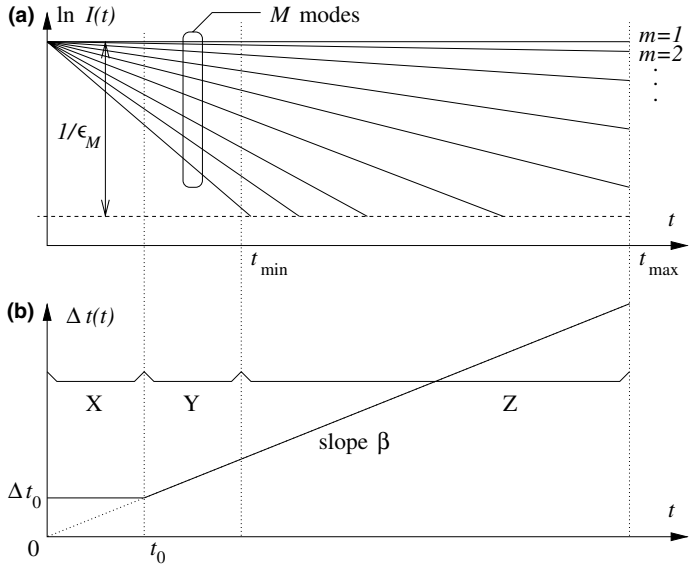

Fig. 6. (a) Schematic of mode decay and error parameter ϵ_{M}. A constant M modes are evolved from $t = 0$ to $t = t_{\mathrm{min}}$; after this they are successively 'killed' when their contribution falls below the error level ϵ_{M}. (b) CN timestep $\Delta t(t)$ as a function of current time t. Assuming that $t_0 \leqslant t_{\mathrm{min}}$, the evolution falls into three phases labelled X, Y, and Z.

above method. However, once the above-calculated $\mathbf{U}(t)$ is substituted into Eq. (30), we note a third assumption becomes required: (iii) that after the projection equation (30), coefficients contribute approximately equally to the signal without significant cancellation. Unfortunately we will see in Section 6.2 that for a certain corner of the (R, t) plane assumption (iii) is not valid and therefore larger errors result.

5.4. Time-dependent truncation of the mode sum

We truncate the number of terms in the sum Eq. (16) to a fixed finite number M as we evolve from $t = 0$ to $t = t_{\mathrm{min}}$ (phases X and Y). We then successively shrink this number for the remaining segment of time $t \in [t_{\mathrm{min}}, t_{\mathrm{max}}]$ (phase Z). This is sketched in Fig. 6(a). This latter truncation in phase Z results in a significant improvement in the scaling of numerical effort (from $\log t_{\mathrm{max}}$ to independent of t_{max}).

The truncation can be performed with exponentially small errors. This is evident from Fig. 7(a) where it is clear that at each fixed time t, the mode contributions die exponentially with m, and that as t increases, fewer and fewer modes are needed to give an accurate answer. This exponential convergence of the mode sum can be understood as the result of the growth with m of the effective absorption coefficient $\mu_{a,m}$ in Eq. (21). Alternatively it can be viewed as the result of the exponential tails of the transverse Fourier representation of the heat kernel [47] as it contracts in time, as sketched in Fig. 3(b) and explained in Appendix B.

We define a desired truncation error level ϵ_{M}. Either by considering the decay rate equation (21) in the top layer of the medium, or using the reasoning in Appendix B, we have an estimate for the transverse wavenumber at which truncation can occur, at a given time t,

$$k_{\mathrm{max}}(t) = \sqrt{\frac{\ln(1/\epsilon_{\mathrm{M}})}{c\kappa_0 t}}. \tag{39}$$

Notice that it decays with time like $t^{-1/2}$. Using the asympotic form for the locations of Bessel zeros, we have a linear relation between wavenumber and mode number,

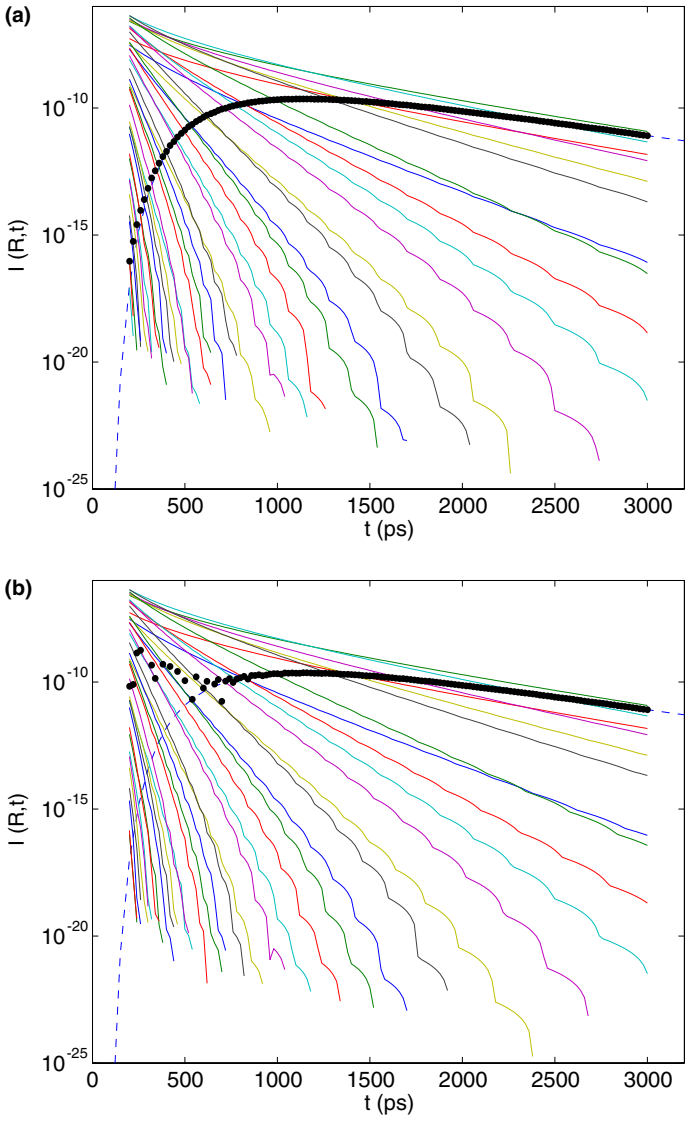

Fig. 7. Contribution due to each mode m (shown as lines) which are summed to give the computed signal $I(R, t)$ (shows as dots), for a homogeneous medium (SYS1), at $R = 40$ mm, $t_{\min} = 200$ ps, giving $M = 38$. The dashed line shows the exact analytic answer. Note the logarithmic vertical scale. (a) Uses consistent time-stepping scheme presented in Section 5.3, and (b) uses a very slight modification which results in inconsistency and therefore huge relative errors at early times where cancellation is important (see Section 8.1).

$$M \approx \frac{k_{\max} a}{\pi}. \tag{40}$$

Therefore, we compute M using Eqs. (39) and (40) with the substitution $t = t_{\min}$, finally rounding M up to the nearest integer. We also decide when to 'kill' each mode m in phase Z by rearranging Eq. (39) to give

$$t_{\text{kill},m} = \frac{\ln(1/\epsilon_{\text{M}})}{c\kappa_0 k_m^2}, \tag{41}$$

where instead of $k_{\max}$ the transverse wavenumber k_m appropriate for each mode has been used.

Our estimates in this section have been based on a homogeneous system with κ equal to its surface value in the inhomogeneous system. This will only cause inaccuracies if $\kappa(z)$ changes substantially over a z range equal to the kernel width $\sqrt{2\kappa_0 c t_{\min}}$. In practise we have found no problems in inhomogeneous systems of interest. We ensure that the error level is set many orders of magnitude lower than the dynamic range γ; we typically use $\epsilon_M = 10^{-12}$, giving in our setting $M_0 \sim 10^2$.

It could also be possible to avoid the above estimation formulae by adding modes successively until convergence of $I(R, t)$ is achieved within the entire desired domain of the (R, t) plane, however we have not implemented this.

5.5. Algorithm summary

We are given the functions $\kappa(z)$ and $\mu_a(z)$ defining the medium, the parameters $t_{\min}$, $t_{\max}$ and $R_{\max}$ defining the region of interest in the (R, t) plane, the detector weighting A, and the error parameters ϵ_a, ϵ_L, ϵ_M, and ϵ_X whose values have all been discussed earlier. The finite-difference parameters h and β are then chosen as discussed in Section 6. The box sizes a and L are chosen via Eqs. (13) and (15). The maximum mode number M is chosen as in Section 5.4. We then compute the wavenumbers k_m and the Bessel coefficients $B_m\psi_m(R)$ appearing in Eq. (22), using lookup tables of χ_m and $J_1(\chi_m)$. The 1D system tridiagonal matrices S_m are then set up as in Section 5.2, and λ_{shift} is found as in Section 5.3. The phase X parameters t_0 and Δt_0 are found as in Appendix A. Each mode m is then initialized and evolved as in Section 5.3 until its time to be 'killed' given by Eq. (41), or $t_{\max}$, whichever comes sooner. We note that, until it is 'killed', every mode undergoes an identical CN timestepping scheme. The signals from each mode are summed using Eq. (22).

The signal output $I(R, t)$ is computed for a requested set of R values, on a uniform time grid. Fig. 11(a) shows representative resulting $I(R, t)$. Linear interpolation is used to resample from the CN timesteps to the uniform time grid. This resampling is $O(\beta^2)$ accurate, so preserves the basic scaling with β (see Section 6.2).

A software implementation (C callable from a MATLAB interface) is freely available online at http://www.cims.nyu.edu/~barnett/software.html.

6. Convergence and accuracy

In this section, we demonstrate convergence with respect to h (finite-difference lattice spacing) and β (timestepping parameter), and validate against a known solution. For convergence with respect to h (Section 6.1) we use relative signal errors, however for the remaining tests we introduce an error measure more relevant for our application. We define the error measure

$$E(t) \equiv \frac{I(t) - I_{\text{exact}}(t)}{I_{\text{err}}(t)}, \tag{42}$$

where $I_{\text{exact}}(t)$ is some 'exact' (or reference) signal. If $E(t) \leqslant 1$ for all t we define this to be acceptable error. Note that the denominator is the acceptable error level $I_{\text{err}}(t)$ defined in Eq. (5) and is controlled by two error parameters: we will use the choices $\epsilon = 10^{-2}$ and $\gamma = 10^5$ throughout. Therefore, for large signals, $E(t)$ is a measure of % relative error.

6.1. Lattice spacing h

The finite-difference scheme is second-order accurate in h, as demonstrated by the 2-layer results for relative error shown in Fig. 8. The convergence prefactor depends strongly on time and system optical properties. For these results, the system with the largest errors of all six systems of Table 1 has been chosen.

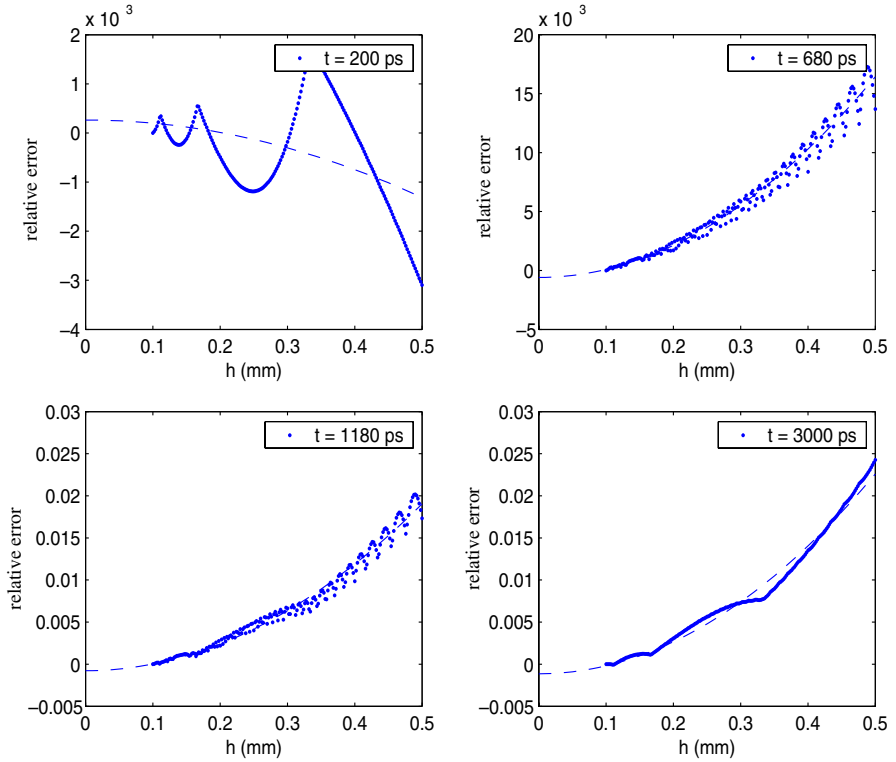

Fig. 8. Convergence of computed signals $I(R, t)$ with finite-difference lattice spacing h, at four choices of t, at $R = 20$ mm. The system was SYS5, which had the largest convergence errors of all the systems from Table 1. The time-stepping scheme was held fixed ($\beta = 0.05$, $\Delta t_0 = 0.04$ ps, see Section 5.3). Normalized changes in signal relative to the value for $h = 0.1$ mm are shown. The dashed curves show best-fit $O(h^2)$ convergence, and are intended only as a guide to the eye.

Rather than having smoothly-varying optical properties, as finite-difference methods are usually intended to handle, these 2-layer systems have a maximal discontinuity in both optical coefficients at the interface between the upper and lower layer. Therefore, we consider this to be a stringent test of the method. Given a desired relative error $\epsilon = 10^{-2}$, this indicates $h = 0.3$ mm is sufficient in this worse-case scenario.

Notice that there is sometimes rapid variation of error as a function of h, resulting in quasi-random scatter, which also dies like $O(h^2)$. This results from the fixed depth of the interface falling at different fractions across a cell as the lattice changes size. Correct cell-averaging (Eqs. (24) and (25)) is needed to ensure this effect does not spoil the $O(h^2)$ convergence. Note that the choice of absorption diagonal term in Eq. (23), namely a $\mu_{a,m}$ that has been averaged across two cells, may also not be optimal.

6.2. Timestepping parameter β

Our CN logarithmic timestepping scheme is second-order accurate in β, as demonstrated by the results in Fig. 9 for a homogeneous system. We postpone the theoretical analysis of this result to a future publication. We only note that the result is reasonable since the constant-Δt CN method is $O(\Delta t^2)$ accurate over a fixed interval of time, and in any small window of time our method resembles this method. It is also important to note that β can be held fixed as the lattice spacing h is reduced: timestepping error at fixed β, since it is related to the physical modal structure, is independent of h. In the figure, oscillations in error are visible,

due to the linear interpolation between the CN timesteps. An improved interpolation scheme (for instance, cubic) could improve the error prefactor but not its basic scaling.

The error $E(t)$ has a peak at early times, which becomes more dominant and wide as R grows (see Fig. 9). This peak is due to breakdown of assumption (iii) from Section 5.3. As shown in Fig. 7(a), the signal at larger R takes some time to grow from very small values; this is simply because the diffusing light has not yet reached that radius (we are in the exponential tail of the spreading fluence distribution). Therefore, at large R and small t a large number of modes m cancel each other almost exactly, allowing the contribution of rapidly-decaying modes to be important. These rapidly-decaying modes were the ones inaccurately evolved in our timestepping scheme, causing the larger errors. Note that at later times the cancellation disappears and higher accuracy is achieved. It is this early peak which sets our acceptable choice of β. Given an $R_{\max}$ of 40 mm and typical tissue optical parameters, this peak is only an order of magnitude larger than later errors, and $\beta = 0.04$ is acceptable. However, if $R_{\max}$ becomes much larger then the necessary β becomes smaller rendering the scheme less efficient. Our form of acceptable error level defined in Eq. (5), as well as our choice of ϵ and γ, have significant effects on the largest acceptable β. For instance, decreasing γ to 10^4 would allow $\beta = 0.07$ to be acceptable, with a corresponding increase in calculation speed.

Fixing $\beta = 0.04$, we have verified that errors are acceptable in a variety of 2-layer systems, as shown in Fig. 10. Because of the large inhomogeneity $\alpha = 10$ tested here, we believe that the errors will also be acceptable for the less extreme systems usually arising in the biomedical application. Notice that, as with the homogeneous system, the initial peak in $E(t)$ is responsible for the maximum error levels.

6.3. Verification in 2-layer system

We have independently verified the accuracy of our method in the case of 2-layer systems, by comparing against the real-space analytic power series method of Tualle et al. [33,52] (with summation parameters

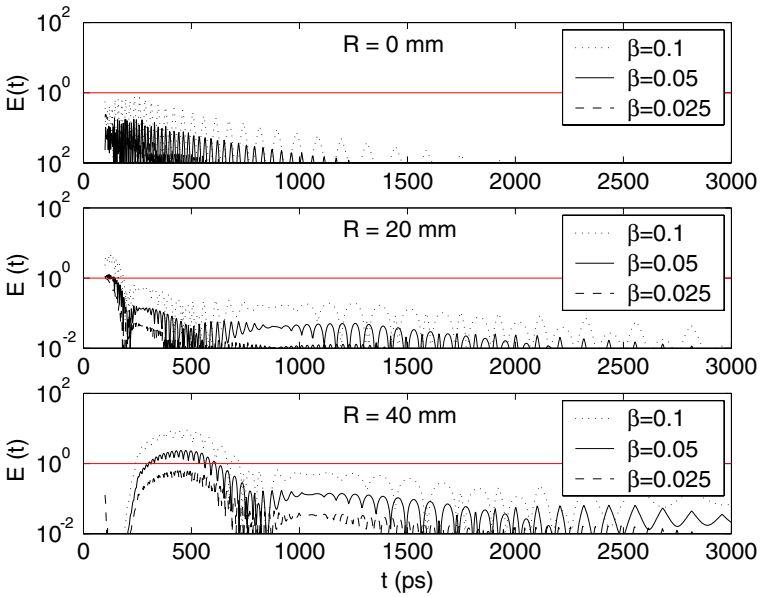

Fig. 9. Convergence of timestepping scheme: signal error measure $E(t)$ as a function of time t, for three timestepping choices of β, at three separations R. The system was homogeneous (SYS1 from Table 1). Other parameters were $h = 0.5$ mm. Error level $E(t)$, which for large signals measures % relative error, is defined in Eq. (42), using the known analytical formula for $I_{\text{exact}}(t)$.

$n = 2$, $l = 20$; see [33]). Fig. 11 illustrates the agreement between the two methods for an example 2-layer system with $\alpha = 4$. The agreement is excellent, typically at the 0.5% level. Notice that the fact that the peak values agree demonstrates the accuracy of the *relative* signals at different separations R (the peaks have not been normalized to have the same values). In terms of the error level $E(t)$, we see that choices of method parameter values h and β derived from the above convergence experiments result in acceptable error. Because the method of Tualle et al. uses the detection of fluence gradient (that is, $A = 0$ in Eq. (4)), our comparison was made using this choice; in any case the variation with A is very slight. Finally we note that our method is stable over much wider variety of system parameters than that of Tualle et al. [52]: we have found that for 2-layer systems where the lower layer is significantly less scattering than the upper, their method can give unusable (nonconvergent) results.

7. Scaling of computational effort

We would like to know how the computational effort scales with the various system and numerical parameters. Nearly all the effort is spent in timestepping the finite-difference method. The cost of increasing the number of desired separations R or interpolating onto more time points t is negligible.

Defining f to be the cost of a single CN step for a single mode m (i.e. a finite-difference system of size n_z), we can estimate total effort with the help of the schematic Fig. 6. Phase X requires $1/\beta$ steps, and phase Y requires $\ln(t_{\min}/t_0)/\beta$ steps, both involving all M modes. In phase Z the number of modes involved at time t is $M\sqrt{t_{\min}/t}$. The total effort is therefore

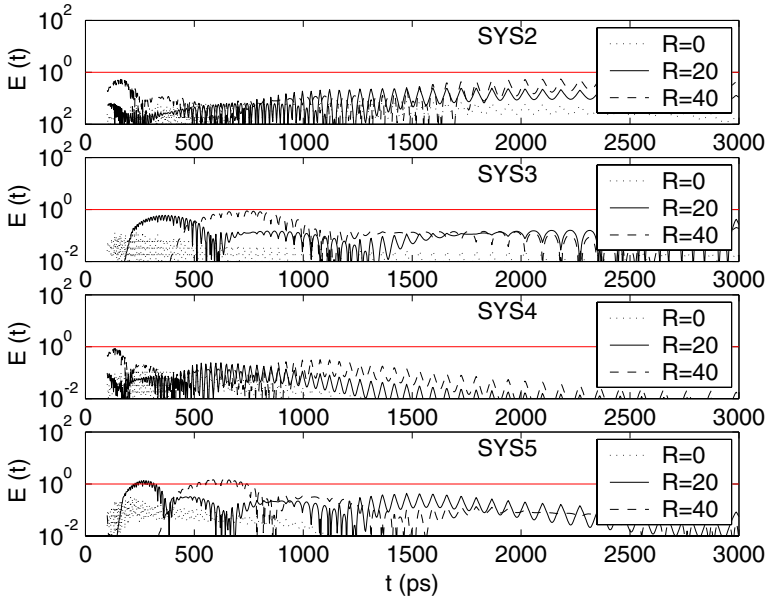

Fig. 10. Verification that timestepping errors are at acceptable levels for the four highly-inhomogeneous systems (SYS2–SYS5) from Table 1, at three source–detector separations R (in mm). Timestepping was fixed at $\beta = 0.04$, with $h = 0.5$ mm. Error measure $E(t)$ is defined in Eq. (42), with $I_{\text{exact}}(t)$ found via uniform CN timestepping at $\Delta t = 0.25$ ps.

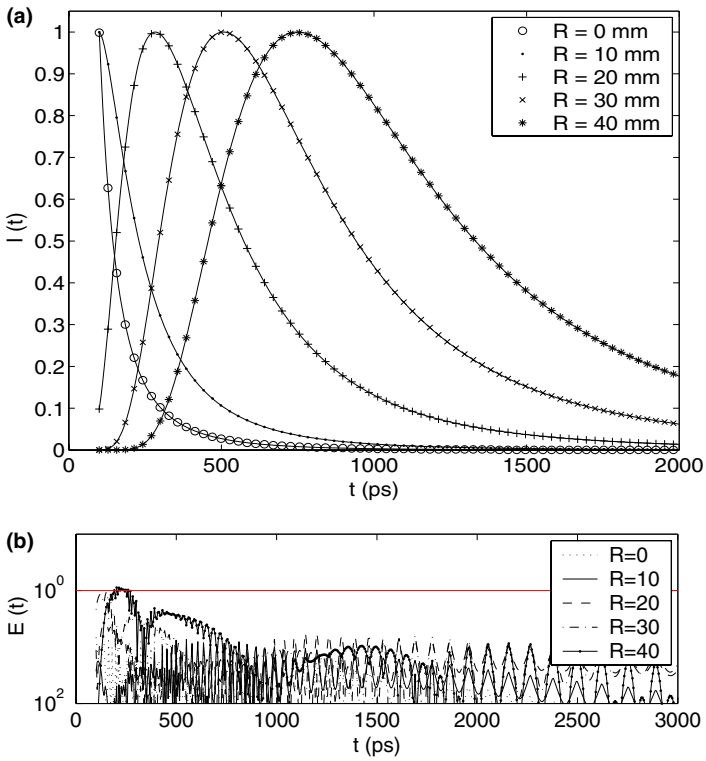

Fig. 11. Verification of method accuracy for the 2-layer system SYS6 (defined in Table 1). (a) Shows computed signals $I(t)$ using our method (symbols) against the signals $I_{\text{exact}}(t)$ computed with the method of Tualle et al. (solid lines). The five source–detector distances R are as labelled. A single multiplicative prefactor was fitted, then used to compare the two sets of signals for all values of R. At each R, $I_{\text{exact}}(t)$ is normalized to have peak value 1 in the time range of interest, and $I(t)$ is normalized by the same factor. (b) Shows error measure $E(t)$, defined in Eq. (42), resulting from comparing the two methods. Timestepping was fixed at $\beta = 0.04$, with $h = 0.5$ mm and $t_{\min} = 100$ ps.

$$F = \frac{Mf}{\beta}\left[1 + \ln\left(\frac{t_{\min}}{t_0}\right)\right] + \int_{t_{\min}}^{t_{\max}} \mathrm{d}t \frac{f}{\beta t} M \sqrt{\frac{t_{\min}}{t}} \tag{43}$$

$$= \frac{Mf}{\beta}\left[3 + \ln\left(\frac{t_{\min}}{t_0}\right) - 2\sqrt{\frac{t_{\min}}{t_{\max}}}\right] \tag{44}$$

$$\approx \frac{qMf}{\beta}, \tag{45}$$

where we have used the fact that $t_{\max} \gg {}_{\min}$. The number q is of order 3–7. The choice of β depends on the required relative accuracy (choice of ϵ); for $\epsilon = 10^{-2}$ appropriate values were discussed in Section 6.2. In the remain required We treat β as a given constant, which is chosen (as discussed in Section 6.2).

7.1. Finite system

We now ask, given the finite cylindrical system of fixed depth a and radius L, how the total effort F scales. We make use of $f = pn_z = pL/h$, where p is the effort per lattice node per timestep in the CN scheme.

Since p is of order a few floating-point operations (flops), this gives F units of flops. Substituting for M using $t = t_{\min}$ in Eqs. (39) and (40), we get

$$F(a,L) \approx \frac{qp\sqrt{\ln(1/\epsilon_M)}}{\pi\beta} \cdot \frac{aL}{h\sqrt{c\kappa_0 t_{\min}}}. \tag{46}$$

The first factor is a number of order 10^3–10^4. The second factor is a ratio of two areas: the system's cross-sectional area aL, and the product of lattice size and the heat kernel spreading size by time $t_{\min}$. Furthermore, although we postpone the derivation and tests to future work, the h required for accuracy ϵ_h at early times $t = t_{\min}$ can be estimated (using a homogeneous system) as

$$h \approx \sqrt{\epsilon_h c\kappa_0 t_{\min}}, \tag{47}$$

which follows from the kernel spreading size and the $O(h^2)$ lattice convergence. We can set $\epsilon_h \sim \epsilon \sim 10^{-2}$. Substituting this gives

$$F(a,L) \approx \frac{qp\sqrt{\ln(1/\epsilon_M)}}{\pi\beta\sqrt{\epsilon_h}} \cdot \frac{aL}{c\kappa_0 t_{\min}}, \tag{48}$$

where the first factor is of order 10^4–10^5, and the second factor ratio is the fraction of the cross-sectional area occupied by the spreading kernel by time $t_{\min}$. Notice that $t_{\max}$ has become irrelevant: the limit $t_{\max} \to \infty$ can be taken with almost no extra effort. This contrasts common constant-Δt schemes where the effort scales linearly with $t_{\max}$. This surprising result is due to our logarithmic timestepping and successive 'killing' of irrelevant modes in phase Z.

7.2. Infinite system

In Section 4, we found the a and L necessary for approximation of the infinite system by a finite one. Using Eqs. (13) and (15), assuming $R_{\max} \ll a$ and $\epsilon_L = \epsilon_a$ we can estimate

$$aL \approx c\kappa_{\max} t_{\max} \ln(1/\epsilon_a). \tag{49}$$

Thus area grows linearly in time, as it must in a diffusion process. Combining this with Eq. (48) gives the effort for the semi-infinite problem

$$F \approx \frac{qp[\ln(1/\epsilon_a)]^{3/2}}{\pi\beta\epsilon_h^{1/2}} \cdot \frac{\kappa_{\max}}{\kappa_0} \cdot \frac{t_{\max}}{t_{\min}}, \tag{50}$$

where we have assumed $\epsilon_M = \epsilon_a$ for convenience. The first factor is of order 10^5–10^6, the second is a ratio bounded by α, and the third is of order 50 in the biomedical application. This gives a typical $F \sim 10^7$ flops. This is consistent with typical runtimes of 0.1–0.3 s which we observe on a 1 GHz Pentium III processor (whose memory-limited speed is less than 10^8 flops per second), with $t_{\min} = 100$ ps and $t_{\max} = 5000$ ps. Thus we are nearly an order of magnitude faster than the ~1 s run time quoted by Martelli et al. [34,35], and, when a large grid of R and t samples is needed, by Tualle et al. [33].

In summary, with error parameters fixed, our infinite system method scales linearly with $t_{\max}/t_{\min}$, at most linearly with α, and inversely with β.

8. Discussion

We have described a method well-adapted for cylindrical systems of finite radius and depth. We have applied it to the case of infinite radius and depth by using error estimates derived from a homogeneous

medium. Because the diffusion inhomogeneity α is not too large, these estimates are not excessively wrong (Section 4). However, the cross-sectional area of the required finite medium grows linearly in time.

There are two potential improvements in effort scaling for the infinite system: (1) Use of a nonuniform finite-difference lattice, where h increases with depth (this could be extended to be adaptive in time, or furthermore, to have a maximum depth L which increases in time). (2) Non-uniform representation in k-space, in the manner of Greengard and Lin [48]. This latter approach is specifically designed to represent the infinite-space heat kernel efficiently, that is, using as few Fourier modes as possible. These modes are logarithmically spaced in wavenumber. The rigorous convergence proofs in their work would need to be relaxed to handle the inhomogeneous medium case. Since currently a large fraction of effort is spent in high-m modes, the logarithmic spacing would be beneficial. The use of these two changes together could change the effort scaling from $O(t_{max}/t_{min})$ to $O(\ln[t_{max}/t_{min}])$, although since in the DOT application t_{max}/t_{min} is only of order 50 this may be limited to an order of magnitude improvement. In other applications this could be more important.

In general, there is much future analytic work that could be done to improve the error estimates in inhomogeneous systems. For instance, can tighter estimates for a and L be found? (A speed-up factor of up to α is available here.) Can rigorous bounds be placed on the error due to truncation to a given a and L? Can better estimates for M be found? Similarly, we have not analysed in detail how the acceptable h depends on t_{min} and the optical properties (the scaling Eq. (47) need not be correct for inhomogeneous systems).

It is worth mentioning why a 2D finite-difference approach, that is, a cylindrical lattice in (ρ, z), is likely to be less successful than the method we have presented. The main problem with a 2D system is that higher Fourier modes cannot be 'killed' in the manner of Section 5.4, so efficiency would be lower. The simplicity of the CN implementation and small dataset size (which can fit within a CPU cache) in 1D also tend to favor our choice. However, it would also be interesting to explore the use of our logarithmic timestepping scheme for parabolic 3D finite-difference applications without symmetry.

Currently our method is limited by the fact that our logarithmic timestepping scheme performs poorly when there is a lot of cancellation between mode contributions; this occurs at large R and early times. This is a fundamental problem for the CN scheme that it cannot handle a large range of decay rates accurately (large decay rates are relevant when there is cancellation; small rates are relevant at later times when cancellation disappears). One potential solution is to change λ_{shift} as a function of m in a consistent fashion (e.g. Eq. (51) below). A more radical approach, which may be competitive in speed, is to solve for (some subset of) eigenvalues and eigenvectors of the matrices S_m and perform an exact Method of Lines evolution, bypassing timestepping schemes altogether.

8.1. Consistency and cancellation at early times

Here, we discuss a detail which can be skipped on first reading. By grouping together the M 1D finite-difference problems into a single ordinary differential equation (28), we have simplified our discussion of timestepping. However, initially in our research each 1D evolution problem equation (26) was handled using a separate time-stepping scheme. This was deemed necessary to preserve accuracy: the rapidly-decaying (high-m) finite-difference problems could either be given smaller timesteps Δt or a different λ_{shift} adapted to handling their large decay rates more accurately. A typical resulting signal at large R is shown in Fig. 7(b): here for illustrative purposes we changed the value of Δt_0 during phase X very slightly for each mode m. The evolution of individual mode contributions shows almost no perceptible difference from Fig. 7(a) (their accuracy remains the same), however their sum is radically different. The resulting signal is utterly wrong for the first 700 ps. The accuracy achievable is limited to a constant fraction (here about 10^{-3}) of the *largest terms* in the sum, which here are many orders of magnitude larger than the true signal.

Therefore, there is a consistency condition. It is not how accurately we can evolve each mode but rather whether the scheme is consistent between modes which controls the signal accuracy when there is a high degree of cancellation. Cancellation at large R and small t is present for the basic reason that the Fourier representation of the Gaussian kernel does a poor job of representing the (exponentially-smaller) tails to high *relative* accuracy. It is a pitfall that we cannot avoid in a transverse mode representation. When a scheme is consistent, then the only limit on the size of signal that can be accurately calculated is the machine precision (in Fig. 7(a) if R is further increased, the minimum achievable accurate signal is $\sim 10^{-15}$ of the maximum term in the sum, due to the double precision arithmetic).

We aim to analyze the consistency condition theoretically in future work. We note that consistency requires that a single timestepping scheme is used (that is, all modes m share a common choice of $\Delta t(t)$, as in Section 5.3). We have also found empirically that consistency is preserved by a m-dependent λ_{shift} when it takes the form

$$\lambda_{\text{shift},m} = \lambda_{\text{shift}}(k_m) = c_1 + c_2 k_m^2, \tag{51}$$

where c_1 and c_2 do not depend on m. There is potential to improve the method by using a λ_{shift} of this form.

9. Conclusions

We have described and tested an efficient new method for solution of the time-dependent diffusion equation (6) in stratified media with arbitrary absorption and diffusion coefficient depth profiles. Our method is geared towards the DA to photon transport forward and inverse problems arising in biomedical imaging, in particular, DOT. In this application a typical run time on a 1 GHz CPU is 0.2 s, which is several times faster than quoted for current methods for less-flexible 2- or 3-layer systems. Our algorithm is reliable and includes automatic choice of all numerical parameters (apart from β discussed in Section 6.2); this is important for unsupervised use in iterative inverse problems.

The basic method applies to a finite cylindrical system, using a sum over transverse modes each of which requires solution of a 1D diffusion problem. A novel logarithmic timestepping scheme (one which is generally applicable to sourceless parabolic problems) allows efficient solution of the 1D problems using finite-difference methods. In order to extend the method to the semi-infinite system case, we derived and tested expressions for the finite system size needed to accurately emulate the infinite system.

We have analyzed numerical errors and convergence in depth, using an error measure acceptable for the DOT application. A recurring theme is that because the diffusion inhomogeneity α arising in our application is not too large, we can use convergence estimates derived for a homogeneous system. In particular, the fact that the heat kernel has exponential tails in position space is used for the finite system size error estimates, and the fact that the same kernel has exponential tails in Fourier space is used for truncating the mode sum. Both therefore give exponential convergence. The finite-difference evolution is $O(h^2)$ and $O(\beta^2)$ accurate. Our tests involved a variety of highly-inhomogeneous 2-layer media, although we emphasize that the method handles arbitrary depth-dependent media. The conclusion that the effort scales like $1/t_{\min}$ for the finite system, and $t_{\max}/t_{\min}$ for the infinite system, means that for speed in DOT applications it is important to increase the desired $t_{\min}$ to as large a value as possible.

Within our general framework there is much flexibility. We have presented the case of reflection geometry, however the modification to transmission through a finite-thickness slab is trivial. We expect that major speed gains could result from implementing a nonuniform finite-difference lattice, and a nonuniform representation of the sum over transverse wavenumber. Other than optical tomography, our method can also be of use in impulsive stratified heat conduction problems, and the inverse problem of determining thermal conductivity depth profiles.

Acknowledgement

In completing this work the author has benefited greatly from discussions with Jonathan Goodman, Mike Shelley, Leslie Greengard, Robert Kohn, Marsha Berger, Donna Calhoun, Dwight Barkley, David Boas, Joe Culver, and Ross Barnett. We also thank Jean-Michel Tualle for providing computer code for his 2-layer method, and the comments of the anonymous reviewers. The author is supported by the Courant Institute at New York University.

Appendix A. Parameter choice for initial Crank–Nicolson phase

The logarithmic CN scheme of Eq. (36) relies on starting at a finite time t_0 with all unphysical high decay-rate modes exponentially damped. We do this via an initial phase of CN with constant time-step Δt_0. We call this phase X (see Fig. 6(b)). In order for damping to be reliable, we need to consider two cases separately: t_0 either less than or greater than $t_{\min}$.

We first assume $t_0 \leqslant t_{\min}$. We consider it a good choice to impose $\Delta t_0 = \beta t_0$, in which case $n \approx 1/\beta$ steps will be involved. The highest decay rate of the finite-difference lattice system S defined in Section 5.2 can be estimated as

$$\lambda_h \approx c\kappa_{\max}\left(\frac{4}{h^2} + k_{\mathrm{M}}^2\right) + c\mu_{\mathrm{a,max}}. \tag{A.1}$$

This follows from considering Eq. (23) for the lattice 'optical mode' (alternating values ± 1), and $m = M$. The CN amplitude damping coefficient for this mode is

$$a_h^{\mathrm{CN}}(\Delta t_0) = \frac{1 - \frac{1}{2}\lambda_h \Delta t_0}{1 + \frac{1}{2}\lambda_h \Delta t_0} \approx -1 + \frac{4}{\lambda_h \Delta t_0} - \cdots \tag{A.2}$$

We truncate the series expansion in $1/\lambda_h \Delta t_0$ at the term shown. After $n = 1/\beta$ such steps, we require the damping to have reached the small ratio ϵ_{X}, which we typically set to 10^{-9}. Therefore

$$\epsilon_{\mathrm{X}} = \left(\left|a_h^{\mathrm{CN}}(\Delta t_0)\right|\right)^n \approx \left(1 - \frac{4}{\lambda_h \Delta t_0}\right)^n \approx \mathrm{e}^{-\frac{4n}{\lambda_h \Delta t_0}}. \tag{A.3}$$

Thus choosing $n = 1/\beta$ gives our estimate

$$\Delta t_0 = \frac{4}{\beta \lambda_h \ln(1/\epsilon_{\mathrm{X}})}. \tag{A.4}$$

The total time is then set by $t_0 = \Delta t_0/\beta$.

If the resulting t_0 turns out to be greater than $t_{\min}$, then we cannot guarantee that the damping has been sufficient by the time signals are needed at $t = t_{\min}$. Therefore in this case we fix $t_0 = t_{\min}$, which removes phase Y (see Fig. 6(b)). We then find the Δt_0 needed to perform damping by this time, using $n = t_{\min}/\Delta t_0$ steps, which will differ from $1/\beta$. Substituting this n into Eq. (A.3) gives in this case

$$\Delta t_0 = \sqrt{\frac{4 t_{\min}}{\lambda_h \ln(1/\epsilon_{\mathrm{X}})}}. \tag{A.5}$$

Appendix B. Truncation of heat kernel in Fourier space

We can use a one-dimensional model to bound the error due to truncation in the Fourier domain of the heat kernel at a given time $t > 0$. See Fig. 3(a) for a schematic in the discrete mode case. The error in 2D or

3D is essentially the same. The result is well known (our presentation parallels the discrete wavenumber case presented in Refs. [47,48]). The fluence in a medium with homogenous properties κ, μ_a, with the single spatial coordinate x obeys

$$\frac{1}{c}\frac{\partial\phi}{\partial t} = \kappa\frac{\partial^2\phi}{\partial x^2} - \mu_a\phi, \tag{B.1}$$

which has the fundamental solution

$$\phi(x,t) = \frac{1}{\sqrt{4\pi\kappa ct}}\mathrm{e}^{-\frac{x^2}{4\kappa ct}-\mu_a ct}. \tag{B.2}$$

This has spatial Fourier transform

$$\tilde{\phi}(k,t) \equiv \int_{-\infty}^{\infty}\phi(x,t)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x = \mathrm{e}^{-(k^2\kappa+\mu_a)ct}. \tag{B.3}$$

We call $\phi_{\mathrm{trunc}}(x,t)$ the inverse Fourier transform of $\tilde{\phi}(k,t)$ using a truncated wavenumber range $k\in[-k_{\max},k_{\max}]$. The resulting error is

$$\begin{aligned}\Delta\phi(x,t) &\equiv |\phi_{\mathrm{trunc}}(x,t)-\phi(x,t)| = \left|\frac{1}{2\pi}\left(\int_{-\infty}^{-k_{\max}}+\int_{k_{\max}}^{\infty}\right)\mathrm{e}^{-(k^2\kappa+\mu_a)ct}\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k\right| \\ &\leqslant \frac{1}{2\pi}\left(\int_{-\infty}^{-k_{\max}}+\int_{k_{\max}}^{\infty}\right)\mathrm{e}^{-(k^2\kappa+\mu_a)ct}\,\mathrm{d}k \\ &\leqslant \mathrm{e}^{-k_{\max}^2\kappa_0 ct}\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-(k^2\kappa+\mu_a)ct}\,\mathrm{d}k = \mathrm{e}^{-k_{\max}^2\kappa ct}\phi(0,t).\end{aligned} \tag{B.4}$$

The second inequality can be shown by shifting the integration variable by $\pm k_{\max}$ as appropriate, and dropping the cross term from the resulting $(k\pm k_{\max})^2$ factor in the exponent. Therefore, the relative error $\Delta\phi(x,t)/\phi(0,t)$ (error relative to the spatial peak) at time t is bounded by $\mathrm{e}^{-k_{\max}^2\kappa ct_{\min}}$. Setting this relative error equal to ϵ_M, and using the top-layer diffusion constant $\kappa=\kappa_0$, gives the estimate Eq. (39).

References

[1] D.A. Boas, D.H. Brooks, E.L. Miller, C.A. DiMarzio, M. Kilmer, R.J. Gaudette, Q. Zhang, Imaging the body with diffuse optical tomography, IEEE Signal Processing Magazine 18 (6) (2001) 57–75.

[2] A. Villringer, B. Chance, Non-invasive optical spectroscopy and imaging of human brain function, Trends Neurosci. 20 (1997) 435–442.

[3] S.R. Hintz, W.-F. Cheong, J.P. van Houten, D.K. Stevenson, D.A. Benaron, Bedside imaging of intracranial hemorrhage in the neonate using light: comparison with ultrasound, computed tomography, and magnetic resonance imaging, Pediatr. Res. 45 (1999) 54–59.

[4] D. Grosenick, K.T. Moesta, H. Wabnitz, J. Mucke, C. Stroszczynski, R. Macdonald, P.M. Schlag, H. Rinneberg, Time-domain optical mammography: initial clinical results on detection and characterization of breast tumors, Appl. Opt. 42 (2003) 3170–3186.

[5] S. Fantini, M.A. Franceschini-Fantini, J.S. Maier, S.A. Walker, B. Barbieri, E. Gratton, Frequency-domain multichannel optical detector for noninvasive tissue spectroscopy and oximetry, Opt. Eng. 34 (1995) 32–42.

[6] A. Klose, A.H. Hielscher, K.M. Hanson, J. Beuthan, Three-dimensional optical tomography of a finger joint model for diagnostic of rheumatoid arthritis, in: D.A. Benaron, B. Chance, M. Ferrari, M. Kohl (Eds.), Photon Propagation in Tissue IV, Proc. SPIE 3566, 1998, pp. 151–160.

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

[8] F.E. Schmidt, M.E. Fry, E.M.C. Hillman, J.C. Hebden, D.T. Delpy, A 32-channel time-resolved instrument for medical optical tomography, Rev. Sci. Instruments 71 (2000) 256–265.

[9] S.R. Arridge, Optical tomography in medical imaging, Inverse Problems 15 (1999) R41–R93.

[10] K.M. Hanson, G.S. Cunningham, S.S. Saquib, Inversion based on computational simulations, in: G.J. Erickson, J.T. Rychert, C.R. Smith (Eds.), Maximum Entropy and Bayesian Methods, Kluwer Academic, Dordrecht, 1998, pp. 121–135.
[11] A. IshimaruWave Propagation and Scattering in Random Media, vol. 1, Academic, 1978.
[12] J.D. Moulton, Diffusion modelling of picosecond laser pulse propagation in turbid media, Masters thesis, McMaster University, 1990.
[13] S.S. Saquib, K.M. Hanson, G.S. Cunningham, Model-based image reconstruction from time-resolved diffusion data, in medical imaging: image processing, Proc. SPIE 3034 (1997) 369–380.
[14] A.H. Barnett, J.P. Culver, A.G. Sorensen, A. Dale, D.A. Boas, Robust inference of baseline optical properties of the human head with 3D segmentation from magnetic resonance imaging, Appl. Opt. 42 (2003) 3095–3108.
[15] S. Oh, A.B. Milstein, R.P. Millane, C.A. Bouman, K.J. Webb, Source–detector calibration in three-dimensional Bayesian optical diffusion tomography, J. Optical Soc. Am. A 19 (2002) 1983–1993.
[16] A.H. Hielscher, R.E. Alcouffe, R.L. Barbour, Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues, Phys. Med. Biol. 43 (1998) 1285–1302.
[17] A.H. Hielscher, S.L. Jacques, L. Wang, F.K. Tittel, The influence of boundary conditions on the accuracy of diffusion theory in time-resolved reflectance spectroscopy of biological tissues, Phys. Med. Biol. 40 (1995) 1957–1975.
[18] L. Wang, S.L. Jacques, L. Zheng, MCML-Monte Carlo modeling of light transport in multi-layered tissues, Comput. Methods Programs Biomed. 47 (1995) 131–146.
[19] A.H. Hielscher, H.L. Liu, B. Chance, F.K. Tittel, S.L. Jacques, Time-resolved photon emission from layered turbid media, Appl. Opt. 35 (1996) 719–728.
[20] A.H. Hielscher, A.D. Klose, K.M. Hanson, Gradient-based iterative image reconstruction scheme for time-resolved optical tomography, IEEE Trans. Medical Imaging 18 (1999) 262–271.
[21] M. Firbank, S.R. Arridge, M. Schweiger, D.T. Delpy, An investigation of light transport through scattering bodies with non-scattering regions, Phys. Med. Biol. 41 (1996) 767–783.
[22] E. Okada, D.T. Delpy, Near-infrared light propagation in an adult head model. I. Modeling of low-level scattering in the cerebrospinal fluid layer, Appl. Opt. 42 (2003) 2906–2914.
[23] V. Ntziachristos, B. Chance, Accuracy limits in the determination of absolute optical properties using time-resolved NIR spectroscopy, Med. Phys. 28 (2001) 1115–1124.
[24] V. Ntziachristos, A.H. Hielscher, A.G. Yodh, B. Chance, Diffuse optical tomography of highly heterogeneous media, IEEE Trans. Med. Imaging 20 (2001) 470–478.
[25] R.C. Haskell, L.O. Svaasand, T.-T. Tsay, T.-C. Feng, M.S. McAdams, B.J. Tromberg, Boundary conditions for the diffusion equation in radiative transfer, J. Opt. Soc. Am. A 11 (1994) 2727–2741.
[26] R.J. LeVeque, Finite Difference Methods for Differential Equations, notes available at http://www.amath.washington.edu/~rjl/booksnotes.html.
[27] R.D. Richtmeyer, K.W. Morton, Difference Methods for Initial-Value Problems, John Wiley & Sons, NY, 1967.
[28] S. Takatani, M.D. Graham, Theoretical analysis of diffuse reflectance from a two-layer tissue model, IEEE Trans. Biomed. Eng. 26 (1979) 656–664.
[29] J.M. Schmitt, G.X. Zhou, E.C. Walker, R.T. Wall, Multilayer model of photon diffusion in skin, J. Opt. Soc. Am. A 7 (1990) 2141–2153.
[30] I. Dayan, S. Havlin, G.H. Weiss, Photon migration in a two-layer turbid medium. A diffusion analysis, J. Mod. Opt. 39 (1992) 1567–1582.
[31] A. Kienle, M.S. Patterson, Improved solutions of the steady-state and the time-resolved diffusion equations for reflectance from a semi-infinite turbid medium, J. Opt. Soc. Am. A 14 (1997) 246–254.
[32] A. Kienle, M.S. Patterson, N. Dögnitz, R. Bays, G. Wagnières, H. van den Bergh, Noninvasive determination of the optical properties of two-layered media, Appl. Opt. 37 (1998) 779–791.
[33] J.-M. Tualle, J. Prat, E. Tinet, S. Avrillier, Real-space Green's function calculation for the solution of the diffusion equation in stratified turbid media, J. Opt. Soc. Am. A 17 (2000) 2046–2055.
[34] F. Martelli, A. Sassaroli, Y. Yamada, G. Zaccanti, Analytical approximate solutions of the time-domain diffusion equation in layered slabs, J. Opt. Soc. Am. A 19 (2002) 71–80.
[35] F. Martelli, A. Sassaroli, S. Del Bianco, Y. Yamada, G. Zaccanti, Solution of the time-dependent diffusion equation for layered diffusive media by the eigenfunction method, Phys. Rev. E 67 (2003) 056623.
[36] F. Martelli, S. Del Bianco, G. Zaccanti, Procedure for retrieving the optical properties of a two-layered medium from time-resolved reflectance measurements, Opt. Lett. 28 (2003) 1236–1238.
[37] F. Martelli, A. Sassaroli, G. Zaccanti, Y. Yamada, Properties of the light emerging from a diffusive medium: angular dependence and flux at the external boundary, Phys. Med. Biol. 44 (1999) 1257–1275.
[38] M.S. Patterson, B. Chance, B.C. Wilson, Time-resolved reflectance and transmittance for the non-invasive measurement of tissue optical properties, Appl. Opt. 28 (1989) 2331–2336.

[39] K.M. Yoo, F. Liu, R.R. Alfano, When does the diffusion approximation fail to describe photon transport in random media?, Phys. Rev. Lett. 64 (1990) 2647.
[40] G. Bal, Transport through diffusive and non-diffusive regions, embedded objects, and clear layers, SIAM J. Appl. Math. 62 (5) (2002) 1677–1697.
[41] G.J. Tearney, M.E. Brezinski, J.F. Southern, B.E. Bouma, M.R. Hee, J.G. Fujimoto, Determination of the refractive index of highly scattering human tissue by optical coherence tomography, Opt. Lett. 20 (1995) 2258–2260.
[42] F. Bevilacqua, D. Piguet, P. Marquet, J.D. Gross, B.J. Tromberg, C. Depeursinge, In vivo local determination of tissue optical properties: applications to human brain, Appl. Opt. 38 (1999) 4939–4950.
[43] In this work, we ignore complications arising from convolution by the instrument response function (IRF). We have found that even if IRF convolution is included in the forward model, this does not affect our choice of t_{min}.
[44] We exclude the in vitro white matter value of $\mu_s' = 9.1\ \mathrm{mm}^{-1}$ used by Okada [22] since it is a whole order of magnitude larger than any in vivo literature value. Disagreements remain about in vivo baseline properties of the scalp, skull, grey and white matter [16,42,14,22].
[45] This simple picture is complicated by the fact that time-correlation in the forward signals I(R,t) can effectively average statistically-independent measurement errors, forcing us to make the forward model more accurate. Ultimately we plan an error analysis of the full (linearized) inverse problem in order to determine what types and sizes of forward errors have negligible effect on answers to the inverse problem.
[46] J.D. Jackson, Classical Electrodynamics, Wiley, New York, 1975.
[47] L. Greengard, J.A. Strain, A fast algorithm for evaluating heat potentials, Comm. Pure Appl. Math XLIII (1990) 949–963.
[48] L. Greengard, P. Lin, Spectral approximation of the free-space heat kernel, Appl. Computat. Harmonic Anal. 9 (2000) 83–97.
[49] J.R. Kuttler, V.G. Sigillito, Eigenvalues of the laplacian in two dimensions, SIAM Rev. 26 (1984) 163–193.
[50] LAPACK, available at http://www.netlib.org/lapack.
[51] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C: The Art of Scientific Computing (Cambridge University Press, 1992), p. 848, available at http://lib-www.lanl.gov/numerical/bookcpdf.html.
[52] J.-M. Tualle, private communication.