Parallel inverse-problem solver for time-domain optical tomography with perfect parallel scaling

E. L. Gaggioli[a,b], and O. P. Bruno[c]

[a] *Instituto de Astronomía y Física del Espacio (IAFE, CONICET–UBA), Casilla de Correo 67 - Suc. 28 (C1428ZAA), Buenos Aires, Argentina.*
[b] *Departamento de Física, FCEyN, Universidad de Buenos Aires.*
[c] *Computing and Mathematical Sciences, Caltech, Pasadena, CA 91125, USA.*

Abstract

This paper presents an efficient *parallel* radiative transfer-based inverse-problem solver for time-domain optical tomography. The radiative transfer equation provides a physically accurate model for the transport of photons in biological tissue, but the high computational cost associated with its solution has hindered its use in time-domain optical-tomography and other areas. In this paper this problem is tackled by means of a number of computational and modeling innovations, including 1) A spatial parallel-decomposition strategy with *perfect parallel scaling* for the forward and inverse problems of optical tomography on parallel computer systems; and, 2) A Multiple Staggered Source method (MSS) that solves the inverse transport problem at a computational cost that is *independent of the number of sources employed*, and which significantly accelerates the reconstruction of the optical parameters: a six-fold MSS acceleration factor is demonstrated in this paper. Finally, this contribution presents 3) An intuitive derivation of the adjoint-based formulation for evaluation of functional gradients, including the highly-relevant general Fresnel boundary conditions—thus, in particular, generalizing results previously available for vacuum boundary conditions. Solutions of large and realistic 2D inverse problems are presented in this paper, which were produced on a 256-core computer system. The combined parallel/MSS acceleration approach reduced the required computing times by several orders of magnitude, from months to a few hours.

1. Introduction

This paper presents an efficient parallel inverse-problem solver, based on the radiative transfer equation (RTE), for time-domain optical tomography. As is well known, the RTE provides a physically accurate model for the transport of photons in biological tissue [1, 2], but its applicability in the context of inverse problems has been hindered by the high computational cost required for its solution. In this paper this difficulty is addressed by means of a combination of three main strategies, namely, (i) Use of the spectral FC-DOM approach [3] for the solution of the RTE equations; (ii) An effective parallel implementation of the FC-DOM method based on a spatial domain-decomposition strategy; and (iii) A Multiple Staggered Source (MSS) setup, which utilizes certain combinations of sources operating simultaneously instead of the sequences of single sources used in previous approaches. When implemented on a 256-core computer cluster the overall parallel/MSS approach results in computing-time acceleration by several orders of magnitude—thus making it possible to solve large 2D inverse problems, such as the problem of imaging within a neck model considered in Section 7.

The proposed direct parallel solvers provide a number of advantages. In contrast to other algorithms, for example, the approach presented in this contribution is highly efficient independently of the number of sources employed (cf. [4]). Alternative parallel strategies based on GPU architectures [5] might be appropriate for optical tomography based on the diffusion approximation. But, unfortunately, the diffusion approximation is not accurate for many applications, and, in view of the high storage required by the RTE-based time-domain inverse problem, a GPU-based RTE solver may not be viable. In the review [6] of parallelization strategies for this problem, for strong scaling tests, all parallelization strategies show an efficiency scaling significantly below the ideal. Ref. [7] presents the only parallel strategy known to the authors which reports a perfect parallel efficiency, but the applicability of the method is restricted by a fundamental limitation to non-absorbing and non-scattering media. In contrast, the parallel algorithm presented in this paper enjoys perfect parallel efficiency for general problems, including arbitrarily prescribed scattering and absorption, and finds no restrictions regarding the transport regime, the number of sources, or the number of discrete ordinates employed. A parallel efficiency of 136.7% with up to 256 cores for the benchmark presented in this work has been obtained by means of the proposed parallelization for the FC-DOM algorithm (see sec. 4.3 Figure 4 and its caption). For comparison, reference [8, p. 153] reports a computing time of 44.3 hours for a single 2D forward time solution of the

Email address: egaggioli@iafe.uba.ar (E. L. Gaggioli)

RTE, which, running on 64 cores, the algorithms presented in this paper obtain in less than thirty minutes.

As mentioned above, an additional significant reduction in the computing time required for the solution of the inverse problem is achieved by exploiting the proposed MSS method—which, combining multiple sources in each RTE solution, reduces the number of time-domain direct and adjoint RTE solutions required. MSS acceleration by a factor of six is demonstrated in this paper, without any degradation in the accuracy of the inverse problem reconstruction, relative to the time required by the Transport Sweep method [9, 10] (TS) ubiquitously encountered in optical tomography.

The rest of this paper is organized as follows. Section 2 presents the classical optical-tomography forward problem. Section 3 then presents a derivation of the adjoint gradient formalism under the highly relevant general Fresnel boundary conditions (thus generalizing results previously available for vacuum boundary conditions) which, in particular, incorporates a class of *generalized sources* inherent in the MSS method in addition to the classical source types inherent in the TS method. A direct comparison, presented in Section 3.6, between gradients for the objective function obtained by finite differences and the adjoint method demonstrate the validity and accuracy of the proposed adjoint approach. Section 4, in turn, presents the numerical methods employed for the numerical solution of the RTE, as well as the spatial domain decomposition strategy utilized for their efficient parallelization. Section 5 demonstrates the high-order convergence of the overall RTE parallel solver for smooth solutions, and it presents a demonstration for a photon density of a type used in laboratory settings. Section 7 presents results of a number of reconstructions obtained by the proposed methods, including large 2D inverse problems, such as the problem of imaging within a head model and an MRI-based neck model.

2. Preliminaries

Defining the transport operator

$$\mathcal{T}[u] = \frac{1}{c}\frac{\partial u(\mathbf{x},\theta,t)}{\partial t} + \hat{\theta}\cdot\nabla u(\mathbf{x},\theta,t) + a(\mathbf{x})u(\mathbf{x},\theta,t) + b(\mathbf{x})\left[u(\mathbf{x},\theta,t) - \int_{S^1}\eta(\hat{\theta}\cdot\hat{\theta}')u(\mathbf{x},\theta',t)d\theta'\right], \quad (1)$$

we consider the initial and boundary-value radiative transfer problem

$$\begin{aligned} &\mathcal{T}\big[u\big] = 0,\ (\mathbf{x},\theta)\in\Omega\times[0,2\pi)\\ &u(\mathbf{x},\theta,t=0) = 0,\ (\mathbf{x},\theta)\in\Omega\times[0,2\pi)\\ &u(\mathbf{x},\theta,t) = f(\hat{\theta}\cdot\hat{\nu})u(\mathbf{x},\theta_r,t) + q(\mathbf{x},\theta,t),\ (\mathbf{x},\theta)\in\Gamma_- \end{aligned} \quad (2)$$

for the amount of energy $u = u(\mathbf{x},\theta,t)$ irradiated at a point $\mathbf{x}$, propagating with direction $\hat{\theta}\in S^1$, at a time t, over the two-dimensional spatial domain Ω and for $0 \le t \le T$ ($T > 0$), where $\hat{\theta} = (\cos(\theta),\sin(\theta))$ ($0\le\theta\le 2\pi$), $\hat{\theta}' = (\cos(\theta'),\sin(\theta'))$ ($0\le\theta'\le 2\pi$), and where S^1, c, $a(\mathbf{x})$ and $b(\mathbf{x})$ denote the unit circle, the average speed of light in the medium, and the absorption and scattering coefficients, respectively. Further, $\hat{\nu}$ denotes the outward unit vector normal to the boundary $\partial\Omega$ of the domain Ω, $\Gamma_\pm = \{(\mathbf{x},\theta)\in\partial\Omega\times[0,2\pi), \pm\hat{\nu}\cdot\hat{\theta} > 0\}$—with subindex $-$ (resp. $+$) denoting the set of incoming (resp. outgoing) directions—and $q(\mathbf{x},\theta,t)$ denotes a given source function, which can be used to model illuminating lasers that inject radiation over some portion of Γ_-. Additionally, the function $f(\hat{\theta}\cdot\hat{\nu}) = f(\cos(\theta-\theta_\nu))$ denotes the Fresnel coefficient, according to which radiation is reflected into Ω and transmitted away from Ω in the directions $\hat{\theta}_r = (\cos(\theta_r),\sin(\theta_r))$ and $\hat{\theta}_{\rm tr} = \hat{\alpha}(\hat{\theta})$ respectively, where, in accordance with Snells' law for the refractive indexes n_Ω and n_0 within and outside Ω, $\theta_r = R(\theta) = 2\theta_\nu - \theta + \pi \pmod{2\pi}$ and $\hat{\alpha}(\hat{\theta})$ denote a reflection function and the transmission angle, respectively. And, finally

$$\eta(\hat{\theta}\cdot\hat{\theta}') = \frac{1}{2\pi}\frac{1-g^2}{(1+g^2-2g\,\hat{\theta}\cdot\hat{\theta}')^{3/2}} \quad (3)$$

denotes the widely used Henyey–Greeinstein phase function [11] which models the angle dependent redistribution probability after a photon collision event. The index $g\in[-1,1]$ characterizes the scattering anisotropy: pure forward scattering (resp. pure backward scattering), wherein all photons emerge in the forward (resp. backward) direction after a collision event, corresponds to a value of $g = 1$ (resp. $g = -1$). Isotropic scattering, in turn, where a photon emerges with equal probability in any given direction after collision corresponds to the value $g = 0$. The total density of photons at a particular point $\mathbf{x}$ at a time t, which is considered in particular in the context of the photon diffusion approximation, is quantified by the scalar flux

$$\Phi(\mathbf{x},t) = \int_{S^1} u(\mathbf{x},\theta,t)d\theta. \quad (4)$$

To conclude this section we mention the window function [12]

$$w(v) = \begin{cases} 1 & \text{for } v = 0,\\ \exp\left(\frac{2e^{-1/|v|}}{|v|-1}\right) & \text{for } 0<|v|<1,\\ 0 & \text{for } |v|\ge 1. \end{cases} \quad (5)$$

of the real variable v, which vanishes for $|v|\ge 1$ and smoothly transitions to one in the interval $-1 < v < 1$. This function is used in multiple roles in what follows—including modeling of both the temporal profile and the collimated irradiation of the laser beam, as well as the spatial dependence of the sensitivity of the detectors.

3. Inverse problem and adjoint gradient formalism

3.1. Objective functions

This section introduces the general form of the *objective functions* whose minima yield the solutions of the TS

and MSS inverse transport problems we consider, as well as the *adjoint formalism* we use for the efficient evaluation of the corresponding objective-function gradients under general Fresnel boundary conditions—that is, using a possibly non-vanishing Fresnel coefficient f in the last line in equation (2). For $f(\hat{\theta}\cdot\hat{\nu})\equiv 0$ the gradient expression we obtain coincides with the well-known result for homogeneous boundary conditions that are relevant in neutron transport [13] (vacuum boundary conditions) and that are also often assumed in the context of optical tomography [9, 10, 14].

The TS and MSS objective functions are based on use of sources of two different types, both of which can expressed in the form

$$q = q_i(\mathbf{x},\theta,t) = \sum_{k=1}^{N_s} s_{k,i}(\mathbf{x},\theta,t), \quad i = 1,2,\ldots,N_q \tag{6}$$

for certain integer values of N_q and N_s. Here, using the window function (5), we define

$$s_{k,i}(\mathbf{x},\theta,t) = \exp(-|\mathbf{x}-\mathbf{x}_{k,i}|^2/2\sigma_{\mathbf{x}}^2)w(\delta_{k,i})T(t-\tau_{k,i}), \tag{7}$$

for given laser-beam positions $\mathbf{x}_{k,i}\in\partial\Omega$, beam diameter $\sigma_{\mathbf{x}}$, time delays $\tau_{k,i}\geq 0$, and angular departures $\delta_{k,i} = |\theta-\theta_{k,i}|/\sigma_\theta$, where $\theta_{k,i}$ $(0\leq\theta_{k,i}<2\pi)$ and σ_θ model the center and the angular spread of the nearly collimated irradiation from the (k,i)-th laser beam. Note that, for both the TS and MSS configurations a total of

$$N_b = N_q N_s \tag{8}$$

laser beams are utilized.

For the TS-type sources we set $N_s = 1$ and, generally, $N_q > 1$ (that is to say, a sequence of $N_q > 1$ single-laser sources is used, requiring the solution of N_q pairs of forward/adjoint problems per gradient-descent iteration), while in the proposed MSS-type sources we let $N_s > 1$ and $N_q = 1$ (so that $N_s > 1$ laser beams are combined in a single "generalized" source, thus requiring the solution of only one forward/adjoint problem pair per gradient-descent iteration). In each "sweep" of the TS method each one of the $N_q > 1$ sources is applied independently of all others, with delays $\tau_{1,i} = 0$ $(1\leq i\leq N_q)$, and readings are recorded at all of the detectors used [9, 10]. In the proposed MSS method, instead, the single generalized source $(N_q = 1)$ is used which incorporates time-staggered contributions from $N_s > 1$ laser beams along $\partial\Omega$, with time delays $\tau_{k,1}\geq 0$. In view of the time delays it utilizes, the single forward/adjoint solve used in the MSS method requires longer computing time than each one of the $N_q > 1$ forward/adjoint solves required by the TS method. In all, as illustrated in Section 7, the combined-source strategy inherent in the MSS approach leads to significant gains in the overall inversion process without detriment in reconstruction accuracy.

Both the TS and MSS approaches rely on use of a number $N_d \geq 1$ of detectors, where the j-th detector $(1\leq j\leq N_d)$, which is placed at the point $\mathbf{x}_j\in\partial\Omega$, is characterized by a measurement operator $G_j = G_j[u](t)$ defined by

$$\begin{aligned} G_j[u] = \oint_{\partial\Omega}\int_{\hat{\theta}\cdot\hat{\nu}>0} &[1-f(\hat{\theta}\cdot\hat{\nu})]\hat{\theta}\cdot\hat{\nu} \\ &\times w\left(\frac{|\mathbf{x}-\mathbf{x}_j|}{\sigma_d}\right)u(\mathbf{x},\theta,t)d\theta dS \end{aligned} \tag{9}$$

for any given function $u = u(\mathbf{x},\theta,t)$ defined for $(\mathbf{x},\theta,t)\in\Omega\times[0,2\pi)\times[0,T]$. Here, using equation (5) and letting $\sigma_d > 0$ denote the effective area of the detectors, the factor $w(|\mathbf{x}-\mathbf{x}_j|/\sigma_d)$ characterizes the spatial sensitivity of the j-th detector and dS denotes the element of area on $\partial\Omega$. Clearly, the operator G_j quantifies the flux of transmitted photons over the surface of the detector. For each generalized source q_i we have a set of N_d time resolved detector readings. The position and number of detectors remain fixed throughout the inversion process.

3.2. Inverse problem

The reconstruction of the absorption properties in tissue enables the identification of tumors [15, 16, 17], functional imaging of the brain [18, 19, 20], and characterization of different tissue constituents in medical imaging. In this work we focus in the reconstruction of the absorption coefficient only, although the proposed approach can be easily extended to other reconstruction problems, such as, e.g., the problem of determining the RTE sources, with application in the related discipline of fluorescence optical tomography and bioluminescence tomography [1, 14, 21]. Prior information on the scattering coefficient $b(\mathbf{x})$, which can be obtained from high resolution imaging modalities, is generally assumed for cancer diagnosis and treatment monitoring [22, 23]. Such prior knowledge additionally provides (limited) information on the parameter $a(\mathbf{x})$, namely, the known absorption coefficient of, say, bone and air, on one hand, as well as upper and lower bounds on the absorption coefficient of soft tissue, which can be used to constrain the space of functions where the minimizer is sought. In view of these considerations, in what follows we make explicit the dependence of the transport operator $\mathcal{T}$ and the solution u in eq. (1) on the absorption coefficient $a = a(\mathbf{x})$ by denoting

$$\mathcal{T}[u] = \mathcal{T}[u,a] = \mathcal{T}[u,a](\mathbf{x},\theta,t) \tag{10}$$

and

$$u = u[a] = u[a](\mathbf{x},\theta,t), \tag{11}$$

respectively.

We express our inversion problem for the optical parameter $a(\mathbf{x})$ in terms of the problem of minimization of the objective function

$$\Lambda[a] = \sum_{i=1}^{N_q} g_i[u_i], \tag{12}$$

where, for a given absorption coefficient a,

$$u_i = u_i[a] = u_i[a](\mathbf{x}, \theta, t) \tag{13}$$

denotes the solution $u = u_i$ of equation (2) with $q = q_i$ (increasingly added detail is included in eq. (12) from left to right concerning the dependence of u_i on a and the spatial, angular and temporal variables), and where, for a given number $N_q \times N_d$ of detector measurements $\widetilde{G}_{j,i}$ (N_d detector readings $\widetilde{G}_{j,i}$ for each of the N_q generalized sources q_i) and using eq. (9) g_i, denotes the functional

$$g_i[u] = \frac{1}{2} \sum_{j=1}^{N_d} \int_0^T (G_j[u] - \widetilde{G}_{j,i})^2 dt. \tag{14}$$

3.3. Functional derivatives

To minimize the objective function eq. (12) we rely on a gradient descent algorithm based on use of the functional derivative $\frac{d\Lambda}{da}[a; \delta a]$ with respect to the absorption coefficient function $a = a(\mathbf{x})$ in the direction δa. Here $\frac{d}{da}$ denotes Gateaux differentiation [24]: for a given function $a = a(\mathbf{x})$ and a given perturbation $\delta a = \delta a(\mathbf{x})$, the Gateaux derivative of a given functional $h = h[a]$ in the direction δa is defined by

$$\frac{dh}{da}[a; \delta a] = \lim_{\varepsilon \to 0} \frac{h[a + \varepsilon \delta a] - h[a]}{\varepsilon}. \tag{15}$$

A similar definition can be given for partial Gateaux derivatives for an operator $w = w[a](\mathbf{x}, \theta, t)$ (such as, e.g., the operator (1), the solution $u = u[a] = u[a](\mathbf{x}, \theta, t)$ of equation (2), etc.):

$$\frac{\partial w}{\partial a}[a; \delta a](\mathbf{x}, \theta, t) = \lim_{\varepsilon \to 0} \frac{w[a + \varepsilon \delta a](\mathbf{x}, \theta, t) - w[a](\mathbf{x}, \theta, t)}{\varepsilon}. \tag{16}$$

In what follows we utilize Gateaux derivatives of composition of functionals and operators, for which the chain rule is satisfied. For example, for the composition $h \circ w[a] = h\big[w[a]\big]$ we have the chain-rule identity

$$\frac{d(h \circ w)}{da}\big[a; \delta a\big] = \frac{dh}{dw}\left[w[a]; \frac{\partial w}{\partial a}\big[a; \delta a\big]\right]. \tag{17}$$

In our context we may illustrate this relationship as follows. As a variation equal to a real number ε times a function $\delta a = \delta a(\mathbf{x})$ is added to the function a, a perturbed function $(a + \varepsilon \delta a)$ is obtained and, thus, a perturbed operator value $w[a + \varepsilon \delta a]$. (In our case, the perturbed operator value could be e.g. the solution $u[a + \varepsilon \delta a]$ of equation (2) with absorption coefficient $(a + \varepsilon \delta a)$; cf. eq. (11).) In view of the Gateaux-derivative definition (16) we obtain

$$w[a + \varepsilon \delta a] = w[a] + \varepsilon \frac{\partial w}{\partial a}[a; \delta a] + o(\varepsilon)$$

where $\frac{o(\varepsilon)}{\varepsilon} \to 0$ as $\varepsilon \to 0$. In other words, the error in the approximation $w[a + \varepsilon \delta a] \approx w[a] + \varepsilon \frac{\partial w}{\partial a}[a; \delta a]$ is much smaller than ε. We may thus utilize the approximation

$$h\big[w[a + \varepsilon \delta a]\big] \approx h\left[w[a] + \varepsilon \frac{\partial w}{\partial a}[a; \delta a]\right]$$

in the quotient of increments, of the form (16), for the derivative of the composite function $h\big[w[a]\big]$, which yields

$$\lim_{\varepsilon \to 0} \frac{h\big[w[a + \varepsilon \delta a]\big] - h\big[w[a]\big]}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{h\big[w[a] + \varepsilon \frac{\partial w}{\partial a}[a; \delta a]\big] - h\big[w[a]\big]}{\varepsilon},$$

and, thus, clearly, the right-hand side of (17), as desired.

The needed functional derivative of the objective function (12) is given by

$$\frac{d\Lambda}{da} = \sum_{i=1}^{N_q} \frac{d(g_i \circ u_i)}{da}[a; \delta a]. \tag{18}$$

In order to obtain the derivatives in the sum on the right-hand side of this equation we apply the chain rule identity (17), which yields

$$\frac{d(g_i \circ u_i)}{da}[a; \delta a] = \frac{dg_i}{du}\left[u_i[a]; \frac{\partial u_i}{\partial a}\big[a; \delta a\big]\right], \tag{19}$$

or, using (9) and (14), $\frac{d(g_i \circ u_i)}{da}[a; \delta a] = \mathcal{G}[a; \delta a]$ where

$$\begin{aligned} \mathcal{G}[a; \delta a] \coloneqq \\ \int_0^T \oint_{\partial\Omega} \int_{\hat{\theta} \cdot \hat{\nu} > 0} \sum_{j=1}^{N_d} \Big(G_j\big[u_i[a]\big] - \widetilde{G}_{j,i}\Big) [1 - f(\hat{\theta} \cdot \hat{\nu})] \\ \times \hat{\theta} \cdot \hat{\nu} w\left(\frac{|\mathbf{x} - \mathbf{x}_j|}{\sigma_d}\right) \frac{\partial u_i}{\partial a}[a; \delta a](\mathbf{x}, \theta, t) d\theta dS dt. \end{aligned} \tag{20}$$

Clearly, in view of eq. (20), the gradients (18) required by the gradient descent strategy in a fully discrete context could be produced by evaluating and substituting in this equation the derivative $\frac{\partial u_i}{\partial a}[a; \delta a]$, for each (discretized) absorption coefficient a in the gradient descent process, and for all (discretized) directions δa. But, the evaluation of these partial derivatives, say, by means of a simple finite difference scheme, requires evaluation of one fully spatio-temporal solution of the transport eq. (2) for each direction δa, which clearly entails an extremely high, crippling, computational burden. To avoid this computational expense we rely on the adjoint-method strategy, which is described in what follows.

3.4. Fast gradient evaluation via the adjoint method

To evaluate the derivative displayed in eq. (20) we seek to eliminate the quantity $\frac{\partial u_i}{\partial a}[a; \delta a]$ from the right-hand side of this equation. As indicated in what follows, this can be achieved by considering the initial and boundary problem that is obtained by differentiation, for the given a and in the direction δa, of each one of the three equations in the initial and boundary problem (2). From the first line in (2), in particular, we obtain

$$\begin{aligned} 0 = \frac{d\mathcal{T}}{da}\big[u_i[a], a; \delta a\big] = &\frac{\partial \mathcal{T}}{\partial u}\left[u_i[a], a; \frac{\partial u_i}{\partial a}\big[a; \delta a\big]\right] \\ &+ \frac{\partial \mathcal{T}}{\partial a}\big[u_i[a], a; \delta a\big]. \end{aligned} \tag{21}$$

But, by linearity of $\mathcal{T}$ we have

$$\frac{\partial \mathcal{T}}{\partial u}\left[u_i[a], a; \frac{\partial u_i}{\partial a}[a;\delta a]\right] = \mathcal{T}\left[\frac{\partial u_i}{\partial a}[a;\delta a], a\right], \quad (22)$$

and, thus, in view of (2), the relation

$$\frac{\partial \mathcal{T}}{\partial a}[u_i[a], a;\delta a] + \mathcal{T}\left[\frac{\partial u_i}{\partial a}[a;\delta a], a\right] = 0 \quad (23)$$

results. This relation provides, for each relevant triple $(\mathbf{x}, \theta, t)$, one linear equation for the two unknowns $u_i[a]$ and $\frac{\partial u_i}{\partial a}[a;\delta a]$.

In order to eliminate $\frac{\partial u_i}{\partial a}[a;\delta a]$ from the right-hand side of (20) we subtract from from both sides of this identity a "linear combination with suitable coefficients" λ of the relation (23)—or, more precisely, an integral of the product of this relation times a suitable function $\lambda(\mathbf{x}, \theta, t)$ over $(\mathbf{x}, \theta, t) \in \Omega \times [0, 2\pi) \times [0, T]$. (Below we incorporate additional equations related to the initial and boundary conditions in (2) as well.) For notational compactness we express such integrals in terms of the scalar product notation

$$\langle v, w\rangle = \int_0^T \int_\Omega \int_{S^1} v(\mathbf{x}, \theta, t) w(\mathbf{x}, \theta, t) d\theta d\mathbf{x} dt \quad (24)$$

for any two functions v and w of the variables $(\mathbf{x}, \theta, t)$. Thus, for a given function $\lambda_i = \lambda_i[a](\mathbf{x}, \theta, t)$ we obtain from (23) the equation

$$\left\langle \lambda_i, \frac{\partial \mathcal{T}}{\partial a}[u_i, a;\delta a]\right\rangle + \left\langle \lambda_i, \mathcal{T}\left[\frac{\partial u_i}{\partial a}[a;\delta a], a\right]\right\rangle = 0, \quad (25)$$

which, for a suitably selected function λ_i we intend to subtract from (20) to achieve the cancellation of the challenging derivative term.

To select the function λ_i that attains such cancellation we rely on an integration-by-parts calculation to express the second summand in (25) as an integral of a product of two functions, one of which is precisely $\frac{\partial u_i}{\partial a}$. Integration by parts of that second summand leads to a sum of a "volumetric" integral $\mathcal{A}$ (namely, an integral over $\Omega \times [0, 2\pi) \times [0, T]$) plus a sum $\mathcal{B} + \mathcal{C}$ of integrals over various portions of the boundary of this domain:

$$\left\langle \lambda_i, \mathcal{T}\left[\frac{\partial u_i}{\partial a}[a;\delta a], a\right]\right\rangle = \mathcal{A} + \mathcal{B} + \mathcal{C} \quad (26)$$

where

$$\mathcal{A}[a;\delta a] \coloneqq \int_0^T \int_\Omega \int_{S^1} \frac{\partial u_i}{\partial a}[a;\delta a]\Bigg[-\frac{1}{c}\frac{\partial \lambda_i}{\partial t} \quad (27)$$
$$-\hat{\theta}\cdot\nabla\lambda_i + (a+b)\lambda_i - b\int_{S^1}\eta(\hat{\theta}\cdot\hat{\theta}')\lambda_i d\theta'\Bigg] d\theta d\mathbf{x} dt,$$

$$\mathcal{B}[a;\delta a] \coloneqq \int_\Omega \int_{S^1}\left[\frac{\partial u_i}{\partial a}[a;\delta a]\lambda_i\right]_0^T d\theta d\mathbf{x}, \quad (28)$$

and

$$\mathcal{C}[a;\delta a] \coloneqq \int_0^T \oint_{\partial\Omega}\int_{S^1} \hat{\theta}\cdot\hat{\nu}\lambda_i \frac{\partial u_i}{\partial a}[a;\delta a] d\theta dS dt. \quad (29)$$

Subtracting the linear combination (25) from (20) and using (26)-(29) we obtain

$$\begin{aligned}\frac{d(g_i\circ u_i)}{da}[a;\delta a] =& \mathcal{G} - \mathcal{A} - \mathcal{B} - \mathcal{C} \\ &- \left\langle \lambda_i, \frac{\partial \mathcal{T}}{\partial a}[u_i, a;\delta a]\right\rangle.\end{aligned} \quad (30)$$

Clearly, the quantity $\frac{\partial u_i}{\partial a}$ in (30) will be eliminated, as desired, if and only if

$$\mathcal{A} + \mathcal{B} + \mathcal{C} = \mathcal{G}, \quad (31)$$

since the last term on the right-hand side of (30) does no contain $\frac{\partial u_i}{\partial a}$. Once we select λ_i such that (31) is satisfied, and using the Gateaux derivative relation

$$\frac{\partial \mathcal{T}}{\partial a}[u_i, a;\delta a] = \delta a(\mathbf{x}) u_i[a](\mathbf{x}, \theta, t), \quad (32)$$

the expression

$$\frac{d(g_i\circ u_i)}{da}[a;\delta a] = -\langle \lambda_i[a], \delta a u_i[a]\rangle \quad (33)$$

for the functional derivative, which does not contain the challenging term $\frac{\partial u_i}{\partial a}$, results from (30).

To obtain the solution $\lambda_i = \lambda_i[a](\mathbf{x}, \theta, t)$ of eq. (31) we note that, in view of the spatial integration domains in eqs. (20) and (27)-(29), eq. (31) is satisfied if and only if (i) $\mathcal{A} = 0$, (ii) $\mathcal{B} = 0$ and (iii) $\mathcal{C} - \mathcal{G} = 0$. Equation (i) is satisfied provided the term in brackets in the integrand of (27), which will be denoted by $\mathcal{T}^*[\lambda_i[a], a]$ in what follows, equals zero. Relation (ii) is satisfied by imposing the appropriate "final"condition $\lambda_i(\mathbf{x}, \theta, t = T) = 0$, since, in view of (2), we have $\frac{\partial u_i}{\partial a} = 0$ for $t = 0$. In order to fulfill point (iii), finally, we decompose the integral (29) into two integrals, $\mathcal{C}_+$ and $\mathcal{C}_-$, where integration ranges in the θ variable are restricted to angular domains $\hat{\theta}\cdot\hat{\nu} > 0$ and $\hat{\theta}\cdot\hat{\nu} < 0$, respectively. The integrand in the difference $\mathcal{C}_+ - \mathcal{G}$, which is only integrated over the angular domain $\hat{\theta}\cdot\hat{\nu} > 0$, equals the product of the common factor $F = \frac{\partial u_i}{\partial a}\hat{\theta}\cdot\hat{\nu}$ and the difference $P = \lambda_i - \sum_{j=1}^{N_d}\left(G_j[u_i] - \widetilde{G}_{j,i}\right)\times[1 - f(\hat{\theta}\cdot\hat{\nu})]w\left(\frac{|\mathbf{x}-\mathbf{x}_j|}{\sigma_d}\right)$. To incorporate the summand $\mathcal{C}_-$ under the same $\hat{\theta}\cdot\hat{\nu} > 0$ integration range, in turn, we first utilize the Fresnel boundary condition $\frac{\partial u_i}{\partial a}(\mathbf{x}, \theta, t) = f(\hat{\theta}\cdot\hat{\nu})\frac{\partial u_i}{\partial a}(\mathbf{x}, \theta_r, t)$ $((\mathbf{x}, \theta) \in \Gamma_-)$ that results from differentiation of the boundary condition in eq. (2), and we thus obtain

$$\begin{aligned}\mathcal{C}_-[a;\delta a] \coloneqq& \int_0^T \oint_{\partial\Omega}\int_{\hat{\theta}\cdot\hat{\nu}<0} \hat{\theta}\cdot\hat{\nu}\lambda_i(\mathbf{x}, \theta, t) f(\hat{\theta}\cdot\hat{\nu}) \\ &\times \frac{\partial u_i}{\partial a}[a;\delta a](\mathbf{x}, \theta_r, t) d\theta dS dt.\end{aligned}$$

Then, incorporating the change of variables $\theta = 2\theta_\nu - \theta_r + \pi$, so that $\hat{\theta}\cdot\hat{\nu} = \cos(\theta-\theta_\nu) = -\cos(\theta_r-\theta_\nu) = -\hat{\theta}_r\cdot\hat{\nu}$ we obtain

$$\mathcal{C}_-[a;\delta a] \coloneqq -\int_0^T \oint_{\partial\Omega}\int_{\hat{\theta}_r\cdot\hat{\nu}>0} \hat{\theta}_r\cdot\hat{\nu}\lambda_i(\mathbf{x},2\theta_\nu-\theta_r+\pi,t) \times f(\hat{\theta}_r\cdot\hat{\nu})\frac{\partial u_i}{\partial a}[a;\delta a](\mathbf{x},\theta_r,t)d\theta_r dSdt.$$

Substituting the dummy variable θ_r by θ in this equation, the angular argument in λ_i becomes $2\theta_\nu - \theta + \pi$ which coincides with θ_r, and, thus, calling $Q(\mathbf{x},\theta_r,t) = f(\hat{\theta}\cdot\hat{\nu})\lambda_i(\mathbf{x},\theta_r,t)$ we obtain

$$\mathcal{C}_-[a;\delta a] \coloneqq -\int_0^T \oint_{\partial\Omega}\int_{\hat{\theta}\cdot\hat{\nu}>0} \hat{\theta}\cdot\hat{\nu}Q(\mathbf{x},\theta_r,t) \times \frac{\partial u_i}{\partial a}[a;\delta a](\mathbf{x},\theta,t)d\theta dSdt.$$

Combining this result with the term $\mathcal{C}_+ - \mathcal{G}$ we obtain

$$(\mathcal{C}-\mathcal{G})[a;\delta a] \coloneqq \int_0^T \oint_{\partial\Omega}\int_{\hat{\theta}\cdot\hat{\nu}>0} F\times[P(\mathbf{x},\theta,t) - Q(\mathbf{x},\theta_r,t)]d\theta dSdt,$$

and, thus, (iii) is satisfied provided $P - Q = 0$.

In summary, denoting $\mathcal{T}^*\big[\lambda_i[a],a\big] = -\frac{1}{c}\frac{\partial\lambda_i}{\partial t} - \hat{\theta}\cdot\nabla\lambda_i + (a+b)\lambda_i - b\int_{S^1}\eta(\hat{\theta}\cdot\hat{\theta}')\lambda_i d\theta'$, we have shown that the conditions (i), (ii) and (iii) are satisfied provided the corresponding "adjoint" problem

$$\begin{aligned}
&\mathcal{T}^*\big[\lambda_i[a],a\big] = 0,\ (\mathbf{x},\theta)\in\Omega\times[0,2\pi)\\
&\lambda_i(\mathbf{x},\theta,t=T) = 0,\ (\mathbf{x},\theta)\in\Omega\times[0,2\pi),\quad \text{and}\\
&\lambda_i(\mathbf{x},\theta,t) = f(\hat{\theta}\cdot\hat{\nu})\lambda_i(\mathbf{x},\theta_r,t) + \sum_{j=1}^{N_d}\Big(G_j[u_i] \qquad (34)\\
&-\widetilde{G}_{j,i}\Big)\times[1-f(\hat{\theta}\cdot\hat{\nu})]w\left(\frac{|\mathbf{x}-\mathbf{x}_j|}{\sigma_d}\right),(\mathbf{x},\theta)\in\Gamma_+
\end{aligned}$$

hold. Thus, the function $\lambda_i(\mathbf{x},\theta,t)$ needed in eq. (33) can be obtained by solving the *adjoint back transport* problem (34) in the time interval $T \geq t \geq 0$ with homogeneous final data at time $t = T > 0$. Once the function λ_i has been obtained the component of the functional gradient (33) in the direction δa can inexpensively be obtained by integration, which, in view of (24), may be expressed in the form

$$\frac{d(g\circ u)}{da}[a;\delta a] = -\int_0^T\int_\Omega\int_{S^1}\lambda u\delta a d\theta d\mathbf{x}dt. \qquad (35)$$

The correctness and accuracy of the proposed approach for gradient evaluation are demonstrated in the following section via comparisons with direct finite-difference gradient computations.

3.5. Verification and accuracy assessment of the functional-derivative expression (33)

In this section we present numerical verifications and accuracy assessments for the functional derivative expression (35), with λ_i given as the solution of the adjoint problem (34). To do this we consider a problem of the type (2) with Fresnel boundary conditions, described in what follows—so as to illustrate, in particular, the ability of the functional-derivative expression to produce correct gradients in this case, for which corresponding adjoint treatments were not previously available. As a basis for comparison we obtain numerical gradients produced by direct use of the finite-difference approximation

$$\frac{d(g\circ u)}{da}[a;\delta a]^{FD} \sim \frac{g\big[u[a+\varepsilon\delta a]\big] - g\big[u[a]\big]}{\varepsilon} \qquad (36)$$

for a given direction $\delta a(\mathbf{x})$ and a suitable small value of ε. The index of refraction $n_\Omega = 1.4$ is assumed in the spatial domain $\Omega = [x_{\min},x_{\max}]\times[y_{\min},y_{\max}] = [0,3]\times[0,3]$ and with $n_0 = 1$ outside Ω. For simplicity we use $\delta a = 1$, with given spatially constant values $a(\mathbf{x}) = a$ and $b(\mathbf{x}) = b$ of the absorption and scattering coefficients, and, without loss of generality, we consider a single generalized source $q_1 = q$ for both a TS source case (a single laser incident beam located at $\mathbf{x}_s = (1.5, 0.0)$) and an MSS source case (assumed to consist of the combination of four laser incident beams, one located at the center of each one of the sides of the square domain Ω). (Further details on the modeling of sources can be found at the end of sec. 5, and eq. (48).) For these tests we employ a single detector placed at $\mathbf{x}_d = (0.0, 0.75)$. The time delays required by the MSS method are selected for these examples by enforcing a 50ps time-shift between successive laser start times. A total duration of 60ps was used for each single pulse, and the system was evolved for both TS and MSS cases up to a final time of 600ps. A mesh with $N_x = N_y = 200$, $M = 32$ discrete directions and $T = 60000$ time steps was used for the solution of the RTE and its adjoint. The relative error

$$e = \frac{\left|\frac{d(g\circ u)}{da}[a;\delta a]^{\mathrm{Adj}} - \frac{d(g\circ u)}{da}[a;\delta a]^{\mathrm{FD}}\right|}{|\frac{d(g\circ u)}{da}[a;\delta a]^{\mathrm{Adj}}|}, \qquad (37)$$

was used to quantify the quality provided by the proposed adjoint gradient expression, where $\frac{d(g\circ u)}{da}[a;\delta a]^{\mathrm{Adj}}$ and $\frac{d(g\circ u)}{da}[a;\delta a]^{\mathrm{FD}}$ denote the adjoint and finite difference derivatives, respectively; the value $\varepsilon = 0.0001$ was used to produce the finite difference approximation (36).

Table (1) demonstrates the agreement observed between the values of the functional derivative produced by the finite-difference and adjoint methods under various transport regimes, including several values of the scattering coefficient b and anisotropy coefficient g, and under both the TS and the MSS configurations; errors of similar magnitudes were obtained for a wide range of values of the parameters a, b and g. The excellent agreement observed in all cases suggests that the very large

Table 1: Functional derivative differences

$a[1/cm]$	$b[1/cm]$	g	e_{TS}	e_{MSS}
0.35	80	0.9	0.00008	0.00009
0.35	20	0.0	0.00008	0.00097
0.35	8.0	0.0	0.00027	0.00016
0.35	0.1	0.0	0.00009	0.00026

improvements in computational speed provided by the adjoint method, which would amount to a factor of the order of $(N_x+1)\times(N_y+1)\simeq 40,000$ for the evaluation of the full gradient in the present example, do not impact upon the accuracy in the gradient determination.

3.6. Numerical functional gradient calculation

All of the numerical gradients utilized in this paper were obtained by solving forward and the adjoint problems, followed by use of a discrete version of eq. (35) for a number of perturbation functions δa—each one of which is selected to provide a variation of the absorption coefficient at and around one of the discretized spatial coordinate points $\mathbf{x}_{\ell_1,\ell_2}\in\Omega$ in the discretization $\mathbf{x}_{\ell_1,\ell_2}=(x_{\min}+[\ell_1-1]\Delta x)\hat{x}+(y_{\min}+[\ell_2-1]\Delta y)\hat{y}$, $\ell_1=1,\ldots,N_x+1$, $\ell_2=1,\ldots,N_y+1$ of the domain Ω. The perturbation δa is selected as a pyramid-shaped function which equals one at a point $\mathbf{x}_{\ell_1,\ell_2}\in\Omega$, and becomes zero at and beyond the first neighbors in the discrete grid. At the discrete level, each pyramid-shaped function is approximated by a product of Kronecker delta functions $\delta a_{\ell_1,\ell_2}=\delta_{r,\ell_1}\times\delta_{s,\ell_2}$. Thus, denoting by $\nabla_a g(\mathbf{x}_{\ell_1,\ell_2})$ the value of the functional gradient in the direction $\delta a_{\ell_1,\ell_2}$, the discrete version of eq. (35) is given by

$$\nabla_a g(\mathbf{x}_{\ell_1,\ell_2})\sim-\sum_{m,j}\lambda_{\ell_1,\ell_2,m,j}u_{\ell_1,\ell_2,m,j}\Delta\theta\Delta x\Delta y\Delta t,\quad(38)$$

where $\lambda_{\ell_1,\ell_2,m,j}\sim\lambda(\mathbf{x}_{\ell_1,\ell_2},\theta_m,t_j)$ and where $u_{\ell_1,\ell_2,m,j}\sim u(\mathbf{x}_{\ell_1,\ell_2},\theta_m,t_j)$. Note that eq. (38) represents the functional derivative for a single direction $\delta a_{\ell_1,\ell_2}$, corresponding to the (ℓ_1,ℓ_2) component of the functional gradient. As a result of the adjoint method, the evaluation of eq. (38) for all (ℓ_1,ℓ_2) requires only one forward and one adjoint transport simulation for each generalized source $q=q_i$—a calculation which, if performed by direct use of eq. (36) requires a much larger number $(N_x+1)\times(N_y+1)$ of forward transport simulations and associated overwhelming computational cost. It must be noted, however, that the adjoint method requires storage in memory of full forward and adjoint transport solutions. The solver algorithm proposed in Section 4 is well suited for parallel distributed systems, as it simultaneously provides computational speed and distributed memory availability.

As an illustration, Figure 1 displays the full spatial gradient $\nabla_a g(\mathbf{x}_{\ell_1,\ell_2})$ for $1\le\ell_1\le(N_x+1)$ and $1\le\ell_2\le(N_y+1)$, for certain assumed values $\widetilde{G}_{j,i}$, with a single source and a single detector placed at $\mathbf{x}_s=(1.5,0)$ and $\mathbf{x}_d=(1.0,0)$ respectively.

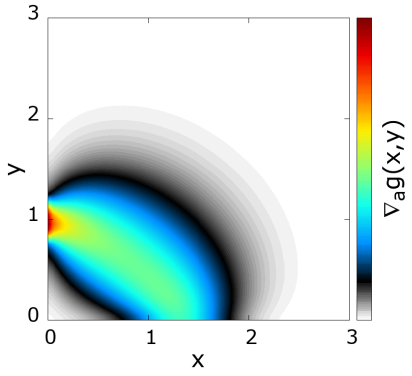

Figure 1: (Color online). Full spatial gradient eq. (38) (in arbitrary units) for a single source-detector pair, located at $\mathbf{x}_s=(1.5,0)$ and $\mathbf{x}_d=(0,1.0)$ respectively, for selected data values $\widetilde{G}_{1,1}$ (one source and one detector).

4. Parallel FC–DOM numerical implementation for the Radiative Transfer Equation

The numerical treatment of the time dependent RTE requires a discretization of all variables in phase space, namely, the spatial, directional and temporal variables. In this section we present a parallel algorithm for the numerical solution of the RTE on the basis of such a discrete grid in phase space. As evidenced by the approach used, and demonstrated by means of numerical experiments in Sections 4.4 and 5, the proposed algorithm enjoys high order accuracy for smooth solutions as well as high parallel efficiency.

4.1. Velocity-domain discretization

We discretize the RTE with respect to the velocity variable $\mathbf{v}=c\hat{\theta}$ by means of the discrete ordinates method. To do this a set of M discrete directions $\hat{\theta}(\theta_m)=\hat{\theta}_m=(\xi_m,\eta_m)$ $(1\le m\le M)$ is utilized where the direction cosines are given by $\xi_m=\hat{x}\cdot\hat{\theta}_m=\cos(\theta_m)$ and $\eta_m=\hat{y}\cdot\hat{\theta}_m=\sin(\theta_m)$ in terms of the Cartesian unit vectors $\hat{x}$ and $\hat{y}$, where

$$\theta_m=\frac{2\pi(m-1)}{M},\quad m=1,2,\ldots,M.\qquad(39)$$

The necessary angular integrations are produced by means of the trapezoidal rule using the associated quadrature weights $w_m=2\pi/M$. Given that the specific intensity u is 2π-periodic in the angular variable θ, and provided the transport solution is sufficiently smooth, the use of the trapezoidal rule gives spectral accuracy for integration with respect to this variable, as illustrated in Figure 5.

Letting $u_m=u(\mathbf{x},\theta_m,t)$, the semidiscrete version of the differential equation in (2) translates into the system of equations

$$\frac{1}{c}\frac{\partial u_m}{\partial t}+\hat{\theta}_m\cdot\nabla u_m+(a+b)u_m$$
$$-b\sum_{m'=1}^{M}w_{m'}p_{m,m'}\,u_{m'}=0,\quad m=1,2,\ldots,M.$$

To ensure energy conservation, and thus enable the stability of the numerical scheme, the phase function eq. (3) is normalized in such a way that its *numerical* integral, as produced, in our case, by means of the trapezoidal rule, equals one [25, 26].

4.2. Time propagation

To evolve the RTE solution in time with high order accuracy, we employ the fourth order Adams–Bashforth time-stepping method [27], which, in the present context, offers a reasonable compromise between accuracy and stability. Calling

$$u_m^k = u_m^k(x,y) = u(\mathbf{x}, \theta_m, t^k), \quad t^k = (k-1)\Delta t, \tag{40}$$

the time dependent RTE equation is evolved from the initial condition via the relations

$$u_m^{k+1} = u_m^k + c\Delta t \sum_{\ell=0}^{3} \chi_\ell \mathcal{L}(u_m^{k-\ell}) \tag{41}$$

where

$$\mathcal{L}(u_m^k) = b \sum_{m'=1}^{M} w_{m'} p_{m,m'} u_{m'}^k - (b+a) u_m^k - \hat{\theta}_m \cdot \nabla u_m^k,$$

with fourth order Adams–Bashforth coefficients given by $\chi_0 = 55/44$, $\chi_1 = -59/24$, $\chi_2 = 37/24$ and $\chi_4 = -3/8$. Throughout this work the four initial time-steps were set to vanish, as befits incident fields that smoothly ramp-up from zero. This arrangement corresponds to the actual experimental setup in time-dependent optical tomography. But we note that arbitrary initial conditions can be treated within our context, by proceeding e.g. as recommended in Section 5 of ref. [28] in a related context.

4.3. Discretization over the spatial domain: the FC–DOM method

For simplicity, in this work we consider a square spatial domain

$$\Omega = \{\mathbf{x} = (x,y) \in [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]\}, \tag{42}$$

although general curvilinear domains can be treated similarly; cf. e.g. [28, 29, 30]. We discretize the spatial domain by means of a uniform grid with grid-sizes Δx and Δy along the x and y directions, respectively; throughout this paper we have used $\Delta x = \Delta y$. We denote by $\mathcal{D}$ denote the discretized version of the spatial domain Ω: $\mathcal{D} = \{(x_i, y_j) \mid i = 1, \ldots, N_x+1 \text{ and } j = 1, \ldots, N_y+1\}$. In what follows we describe the method used for evaluation of spatial derivatives in the x direction, for which we use the grid

$$x_i = x_{\min} + (i-1)\Delta x, \quad \text{with}$$
$$\Delta x = \frac{x_{\max} - x_{\min}}{N_x} = \frac{x_{N_x+1} - x_1}{N_x};$$

the y derivatives are, of course, handled similarly.

We produce the necessary spatial derivatives by means of the Fourier Continuation method (FC) [31], that provides low-dispersion approximations with high-order accuracy on the basis of Fourier expansions for general (non periodic) functions. For a function defined on an interval in the real line, the FC method utilizes a smooth and periodic extension of the given function into an extended interval, for which a regular Fourier series is then obtained, which accurately approximates the extended periodic function, and, thus, in particular, the given function in the original interval. The FC method has extensively been studied and utilized; see e.g. [32, 29, 30, 3, 33] and references therein. In what follows we briefly review this method in the context of the RTE considered in the present contribution.

In order to produce the Fourier Continuation of a given function, such as e.g. $g(x) = u_m^k(x, y_j)$ for a given y_j in the y-discretization mesh, we consider the discrete vector $\mathbf{g} = [g_1, \ldots, g(x_i), \ldots, g_{N_x+1}]^T$ of values of the function g. Utilizing a few of the leftmost and rightmost entries of $\mathbf{g}$, called the "matching values" in what follows, the algorithm produces a continuation vector $\mathbf{g}_c$ that corresponds to the discrete function values of the desired continuation function g_c—which, once the vector $\mathbf{g}_c$ is available, can be obtained by an application of the Fast Fourier Transform (FFT) algorithm. To obtain the vector $\mathbf{g}_c$ the algorithm at first smoothly extends the aforementioned matching values to zero, towards the left and right, respectively, as illustrated in Figure 2. To achieve the desired extensions to zero, certain "projections" are used (which project the first d_l function values $(g_1, g_2, \ldots, g_{d_l})$ and the last d_r function values $(g_{N_x+2-d_r}, \ldots, g_{N_x+1})$ onto polynomial bases), and then, precomputed continuations to zero are utilized for each polynomial in the bases. Here d_l and d_r denote small integers; throughout this paper we have used $d_l = d_r = 5$.

In detail, the continuation procedure can be summarized in the following four steps [29]:

1. The d_l and d_r matching values $g_1, g_2, \ldots, g_{d_l}$ and $g_{N_x+2-d_r}, g_{N-d_r}, \ldots, g_{N_x+1}$ are projected on Gram polynomial bases.
2. Smooth continuations to zero $\widetilde{\mathbf{g}}_l$ and $\widetilde{\mathbf{g}}_r$ are produced on the basis of accurately precomputed extensions to zero [32, 29] of each one of the polynomials in the Gram basis.
3. Two new vectors $\mathbf{g}_l$ and $\mathbf{g}_r$ are generated which expand the dimension of $\widetilde{\mathbf{g}}_l$ and $\widetilde{\mathbf{g}}_r$ to $C+E$ by merely adding zero entries to the vector, where E is a number of extra zeroes used. This expansion is used to obtain vectors whose dimension can be factored into small primes, leading to efficient FFT evaluation.
4. The discrete continued vector $\mathbf{g}_c$ containing $N_p = N_x + 1 + C + E$ components is given by

$$\mathbf{g}_c = \begin{bmatrix} \mathbf{g} \\ \mathbf{g}_l + \mathbf{g}_r \end{bmatrix},$$

and the continued function g_c is obtained from $\mathbf{g}_c$ via the FFT.

Figure 2 illustrates the FC method and the resulting continuation function g_c for the non periodic function $g(x) = 12 + x^2 - e^{x/3}$ in the interval $[0, 6]$ using $N = 60$ points in the grid, $C = 25$ continuation points (dotted line), $d_l = d_r = 5$ matching points (shown with diamonds) and $E = 4$ extra points. The continuation vector $\mathbf{g}_c$ contains $N_p = N_x + 1 + C + E = 90$ entries, which can be factored into small prime numbers, $N_p = 2 \times 3 \times 3 \times 5$. As suggested above, the number E of extra points is selected so as to ensure that N_p equals the product of small prime factors; in practice we enforce that the prime factors in the factorization $N_p = \prod_{j=1}^{F} p_j$ satisfy $p_1 \leq p_2 \leq \ldots \leq p_F \leq 5$—leading to a efficient evaluation of Fourier coefficients vi

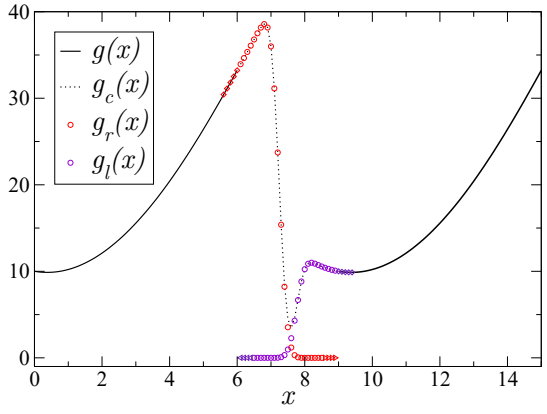

Figure 2: Fourier Continuation of the function $g(x) = 12 + x^2 - e^{x/3}$. For this example, the original function is defined for $x \in [0, 6]$, and a copy of it, displaced to the interval $[9, 15]$, is also included. The continuations to zero, $g_l(x)$ and $g_r(x)$, are shown in circles. The dotted line is the periodic continuation obtained. The diamonds are the matching points, with $d_l = d_r = 5$, and the triangles are the $E = 4$ extra points used in this example.

Using the Fourier coefficients obtained by means of the FFT, as indicated above, the continuation g_c is given by

$$g_c(x) = \sum_{k=-N_p/2}^{N_p/2} a_k \, \exp\left(2\pi i k \frac{(x - x_{\min})}{b}\right) \qquad (43)$$

(where i denotes the imaginary unit, and where $b = [N_p - 1]\Delta x$ denotes the period of the continued function). We additionally employ an exponential filter which, without deterioration in the accuracy [32], ensures the stability and robustness of the method. The filtered coefficients are given by

$$\hat{a}_k = \exp\left(-\alpha \left|\frac{2k}{N_p}\right|^{2\beta}\right) a_k \qquad (44)$$

for adequately selected values of α and β; following [32] throughout this paper we use the values $\alpha = 2$ and $\beta = 55$ in conjunction with the fourth order Adams–Bashforth method. The necessary derivatives of the function g are obtained by direct differentiation of the filtered version of the continuation function (which, for notational simplicity, will also be called g_c):

$$\frac{dg(x)}{dx} \sim \frac{dg_c(x)}{dx} = \sum_{k=-N_p/2}^{N_p/2} \frac{2\pi i k}{b} \hat{a}_k \, e^{2\pi i k \frac{(x - x_{\min})}{b}}. \qquad (45)$$

4.4. Domain decomposition and parallel implementation

In order to solve the RTE on parallel systems we decompose the discrete domain $\mathcal{D}$ introduced in Section (4.3) as a union $\mathcal{D} = \cup_{s=1}^{N_c} \mathcal{D}_s$ of a number N_c of overlapping subdomains $\mathcal{D}_s$ $(1 \leq s \leq N_c)$, with corresponding interior (non-overlapping) regions $\widetilde{\mathcal{D}}_s$, as illustrated in panels (a), (b) and (c) in Figure 3. The overlapping domains $\mathcal{D}_s$ equal the union of corresponding sets $\widetilde{\mathcal{D}}_s$ (shown as sets of black points partitioned along red dashed lines in panel (a)), and a set of "fringe points" (shown in gray in panels (b) and (c)).

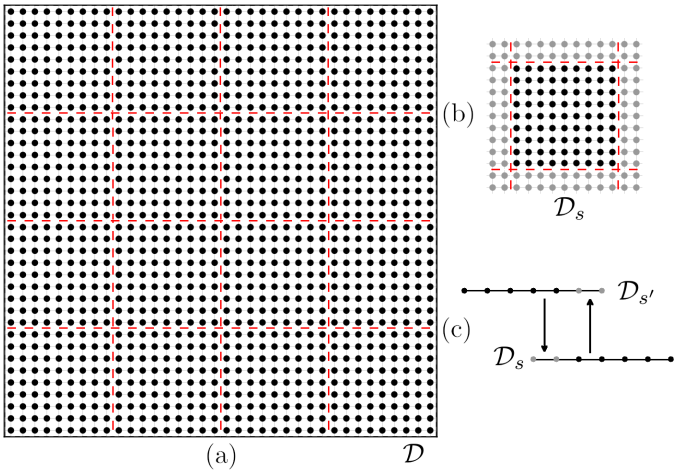

Figure 3: (Color online). (a) The domain $\mathcal{D}$ is decomposed as a union non-overlapping subdomains $\widetilde{\mathcal{D}}_s$. (b) Each subdomain $\widetilde{\mathcal{D}}_s$ extended by two "fringe points" in the x and y directions (gray circles), resulting in the overlapping subdomains $\mathcal{D}_s$. (Subdomains adjacent to the physical boundary Γ are only extended toward the interior of $\mathcal{D}$.) (c) The fringe points are used to enable exchange of information between neighboring subdomains $\mathcal{D}_s$ and $\mathcal{D}_{s'}$: in the one-dimensional cartoon, process s sends to its neighbor s' the u value at the third and fourth points and it receives from s' the u value at the first two points on the s grid.

Each one of the subdomains $\mathcal{D}_s$, which contains $N_{x,s}$ and $N_{y,s}$ points along the x and y directions, respectively, is assigned to one processing core—so that, in particular, N_c equals the number of processing cores used. Note from Figure 3 (b) (and the one-dimensional cartoon Figure 3 (c)) that neighboring subdomains $\mathcal{D}_s$ overlap by four grid points along the x and y directions. In spite of the depiction in Figure 3, which only includes square subdomains, in general, rectangular subdomains need to be used for certain values of N_c.

The subdomain $\mathcal{D}_s = [x_{\min,s}, x_{\max,s}] \times [y_{\min,s}, y_{\max,s}]$ contains grid points of x- and y-coordinates $x_i = x_{\min,s} + (i-1)\Delta x$ and $y_j = y_s(j) = y_{\min,s} + (j-1)\Delta y$, respectively.

The algorithm proceeds by time stepping the quantities $u_m^k(x_i, y_j)$ (cf. (40)) in parallel on each subdomain $\mathcal{D}_s$. At the end of each time step, a boundary exchange of four points is performed between neighboring subdomains as indicated in the caption of Figure 3. The proposed algorithm for the evolution of the time dependent RTE in parallel is summarized in Algorithm 1.

Algorithm 1 FC–DOM Parallel Algorithm

```
Generate the domain decomposition.
Assign one subdomain D_s per process.
for each D_s do in parallel
  Assign Adams–Bashforth initial values.
  for each time–step k do
   for each direction θ̂_m do
    for each y_j do along x
      Apply Fourier Continuation to u^k_m(x_i, y_j).
      Evaluate ∂u^k_m(x_i, y_j)/∂x using (45).
    end for
    for each x_i do along y
      Apply Fourier Continuation to u^k_{m,i}(x_i, y_j).
      Evaluate ∂u^k_m(x_i, y)/∂y using (45).
    end for
    Evaluate the right hand side of equation (41).
    Impose boundary conditions.
    Exchange non-physical boundaries with neighboring
  subdomains.
   end for
  end for
end for
```

5. Numerical Results I: Direct RTE problem

A Fortran 90 parallel implementation of the proposed algorithm for the RTE problem (2), parallelized with MPI and compiled with the intel Fortran compiler, was used to produce all of the numerical results presented in this paper. In all cases the parallel code was run on various numbers of nodes on a 16-node cluster, wherein each node contains an Intel Xeon E5-2630 v3 at 2.40GHz CPU with 24 physical cores and 128Gb of RAM per node. For geometrical simplicity, only sixteen cores per node were used, to match a square geometry containing multiples of 4×4 subdomains.

Three main examples are presented in this section, concerning parallel scaling, accuracy and a demonstration of a laboratory-type simulation.

Our first test concerns computational cost. Figure 4 presents the computational times required by the proposed algorithm for a fixed model problem run on various number of processing cores—a type of test known in the literature as a "strong scaling test". Constant values of the absorption and scattering coefficients were used for this test (although, of course, such selections do not affect the computing time), and the algorithm was run for a discretization with $N_x = N_y = 2000$ and of $M = 16$ directions, for a total of $T = 1000$ time steps. Computing times for other final times or discretization sizes, or even other numbers of computing cores, can easily be estimated from the results presented in Figure 4 in view of the linear scaling of the method. To avoid irregularities in the parallel acceleration caused by the "Intel turbo" technology, which operates as a few cores are used per node, but which is "incrementally" turned off as processors are incorporated in a run, we present scaling data up to sixteen nodes, using a single node as reference, and we report the computing times observed as the number of nodes increases. As indicated by the figure, the proposed algorithm enjoys *perfect parallel scaling* (and, even somewhat better than perfect, on account of the logarithmic cost factors associated with the FFT algorithm [32]) at least up to the 256 processing cores used—and, we conjecture, for arbitrarily large number of processing cores, as long as the computational domain can reasonably be decomposed in a corresponding num

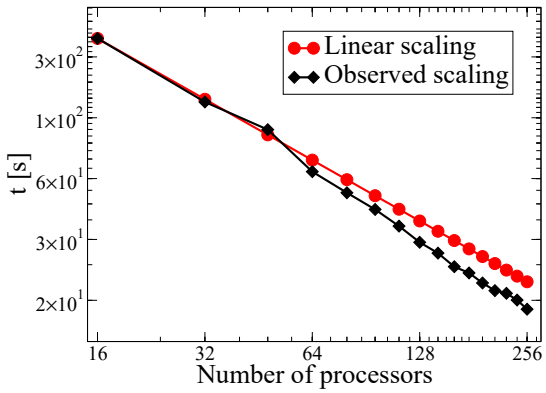

Figure 4: (Color online). Observed scalability of the proposed algorithm for the model problem detailed in the text. The computational time on 256 processors, t_{256} for perfect scalability (red circles) produces a speedup $t_{256}/t_{16} = 16$ with respect to the computational time for the simulation in 16 processors, t_{16}. The measured scalability with the proposed FC–DOM parallel algorithm (black diamonds) produces a speedup $t_{256}/t_{16} = 21.872$, providing an efficiency of 136.7%.

The second example in this section concerns the numerical convergence properties of the algorithm. To study solution errors we consider the "manufactured-solution" RTE problem

$$\begin{aligned} &\mathcal{T}\big[u[a], a\big] = q, \ (\mathbf{x}, \theta) \in \Omega \times [0, 2\pi) \\ &u(\mathbf{x}, \theta, t = 0) = u_0, \ (\mathbf{x}, \theta) \in \Omega \times [0, 2\pi), \\ &u(\mathbf{x}, \theta, t) = u_b, \ (\mathbf{x}, \theta) \in \Gamma_-. \end{aligned} \tag{46}$$

for which the exact solution

$$u^{\text{an}}(\mathbf{x}, \theta, t) = e^{-(x-t)^2-(y-t)^2-\cos(\theta)^2}. \tag{47}$$

is prescribed and accounted for by selecting the right-hand side q and initial and boundary conditions u_0 and

u_b that result as the proposed solution (47) is substituted in (46). For this test we use the square domain $\Omega = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$, with $x_{\min} = y_{\min} = 0$ cm and $x_{\max} = y_{\max} = 3$ cm, and we evolve the solution in the time interval $0 \leq t \leq T = 3$ ps. We use a domain decomposition with eight subdomains. The scattering medium considered is isotropic and homogeneous, with $g = 0$, $a = 0.35/\text{cm}$ and $b = 20/\text{cm}$. We study the convergence properties on all the variables involved, evaluating maximum errors, in each case, by comparison with numerical scalar fluxes (4) against the analytic scalar flux

$$\Phi^{\text{an}}(\mathbf{x}, t) = \int_{2\pi} u^{an}(\mathbf{x}, \theta, t) d\theta = k_\Phi \times e^{-(x-t)^2-(y-t)^2},$$
$$k_\Phi = \int_{2\pi} e^{-\cos(\theta)^2} d\theta \simeq 4.052876133898710,$$

where k_Φ was obtained, with 16-digit accuracy, using the software Wolfram Mathematica. The solution was evolved up to a fixed final time T, and the maximum error was then evaluated by means of the expression

$$\varepsilon = \max_{\mathbf{x}\in\Omega} |\Phi^{\text{num}}(\mathbf{x}, T) - \Phi^{\text{an}}(\mathbf{x}, T)|.$$

The high order convergence of the FC–DOM parallel approach against the proposed manufactured solution is demonstrated in figure 5, which displays convergence curves as Δt, $\Delta x = \Delta y$ and $\Delta\theta$ are refined. (When considering refinements in one of the variables, the mesh sizes in the other variables were kept fixed at sufficiently fine levels, so as to avoid error cross-contamination.) Clearly, excellent convergence is observed in all three cases.

In actual optical tomography contexts, for which an analytic solution is, of course, not known, typically collimated irradiation from a pulsed laser beam is incident at the boundary of the domain; the third and last example in this section, which is presented in what follows, concerns precisely such a scenario. In our test case the collimated irradiation from laser pulses is modeled by means of a peaked function in the directional variable, θ, wherein the function peak coincides with the laser beam direction. We solve equation (2) with Fresnel boundary conditions and using the well-known approach based on use of a Gaussian spatial variation for a collimated laser-beam source:

$$q(\mathbf{x}, \theta, t) = \exp(-|\mathbf{x} - \mathbf{x}_s|^2/2\sigma_{\mathbf{x}}^2) \times w\left(\frac{|\theta - \theta_s|}{\sigma_\theta}\right) \times w\left(\frac{|t - t_s|}{\sigma_t}\right); \quad (48)$$

the beam dependence on the temporal and angular variables, in turn, are modeled on the basis of the window function (5), but any other models of these variations could be used without difficulty. For our example the laser source is located at $\mathbf{x}_s = (1.5, 0)$ with spatial spread $\sigma_{\mathbf{x}} = 0.3$cm, it points in direction $\theta_s = \pi/2$ with angular spread $\sigma_\theta = \pi/4$, and is temporally centered at $t_s = 0$, with temporal spread $\sigma_t = 30$ ps. By choosing a smooth source, we avoid the appearance of singularities in the source function that would be inherited by the RTE solution and deteriorate the convergence properties of the method. We solve the RTE (2) in the 2D spatial domain $\Omega = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}] = [0, 3] \times [0, 3]$ with values $a(\mathbf{x}) = 0.35/\text{cm}$, $b(\mathbf{x}) = 20/\text{cm}$, $g = 0.8$, $n_\Omega = 1.4$ and $n_0 = 1$. Figure (6) displays the scalar flux eq. (4) obtained for this problem; clearly a smooth spatio-temporal distribution is obtained suggesting a solution of high quality in accordance with the accuracy studies presented above in this section.

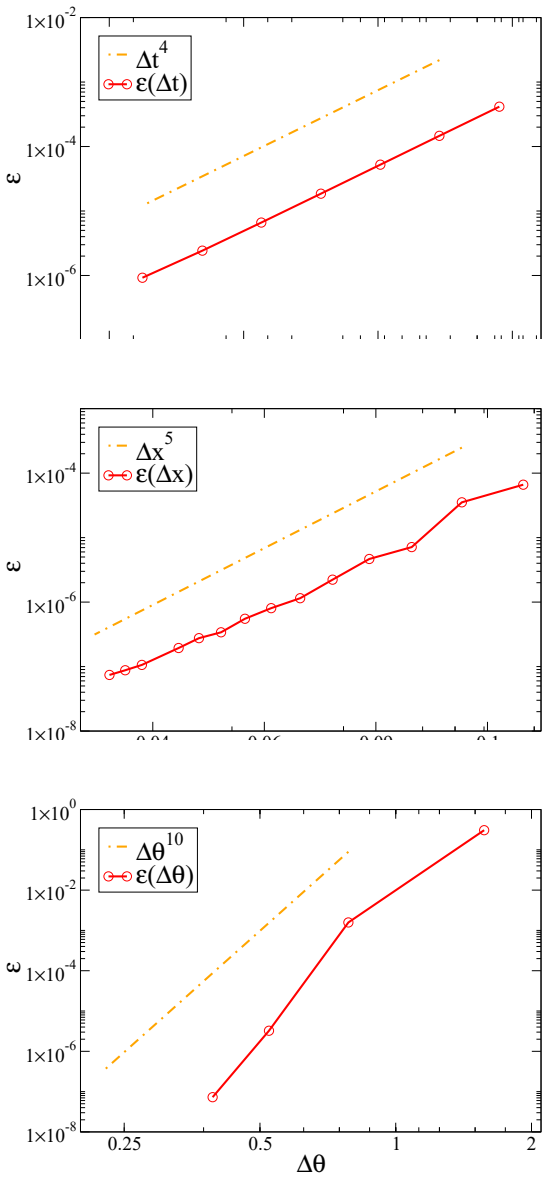

Figure 5: (Color online). Convergence for the FC–DOM parallel solver over each variable. The plot is shown in log-log scale. The dash dotted lines shows the slope for the given convergence order. The solver shows spectral convergence over the angular variable θ.

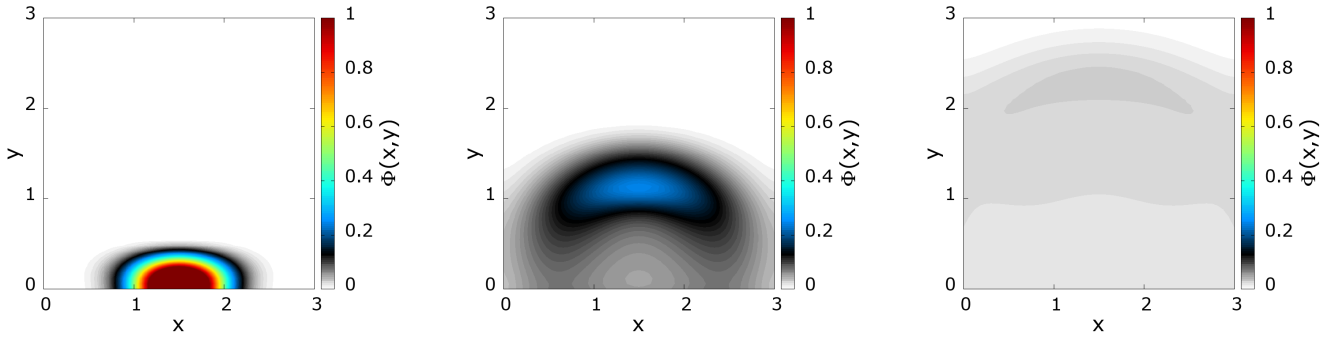

Figure 6: (Color online). Photon transport simulation produced by the parallel FC-DOM algorithm presented above. The scalar flux eq. (4) is shown at three different times from left to right, $t = 30$ps, $t = 100$ps and t

6. Inverse problem solver

As indicated in Section 1, the inverse problem solver proposed in this paper incorporates, in particular, the novel MSS strategy based on use of multiple staggered sources instead of the transport sweep TS strategy that underlies the previous related literature. In the previous TS source method, a separate forward-adjoint pair of simulations is used for each one of the laser-source illuminations, and the results are then combined to achieve the inversion via gradient descent. The proposed MSS approach, in turn, constructs the objective function on the basis of time-staggered laser sources in conjunction with *a single pair of forward and adjoint simulations*, and thus, as demonstrated in Section 7 (quantitatively in Figures 9 and 11, and qualitatively in Figures 10 and 12), it significantly reduces the computing time required for the solution of the inverse problem (e.g. by a factor of six in Figure 11) without any deterioration in image quality.

Detector readings corresponding to a TS source and an MSS source are displayed in the upper and lower panels of Figure 7, respectively; both noiseless and noisy data with 10% random noise added, are presented in the figure. The MSS time delays introduce a degree of decoupling in the portions of the detector signals originating from laser beams applied at different locations, thus enabling effective inversions without requiring independent forward and backward solutions for each one of the beams separately. As indicated in Section 3.1, depending on parameter choices, the expression (6) yields sources corresponding to the TS or MSS illumination approach, and Section 7 presents results obtained from both the TS and MSS methods.

We solve both the TS and MSS inverse problems on the basis of the iterative lm-BFGS [34] quasi-Newton gradient descent method. In detail, the inverse solver seeks the absorption coefficient function that minimizes eq. (12) subject to the constraints $a^\ell \leq a(\mathbf{x}) \leq a^u$, i.e. we seek the absorption function $a = a(\mathbf{x})$ given by

$$a = \text{argmin}_{a^\ell \leq \widetilde{a}(\mathbf{x}) \leq a^u} \Lambda[\widetilde{a}]. \tag{49}$$

The constraints on the values of admissible functions $\widetilde{a}(\mathbf{x})$

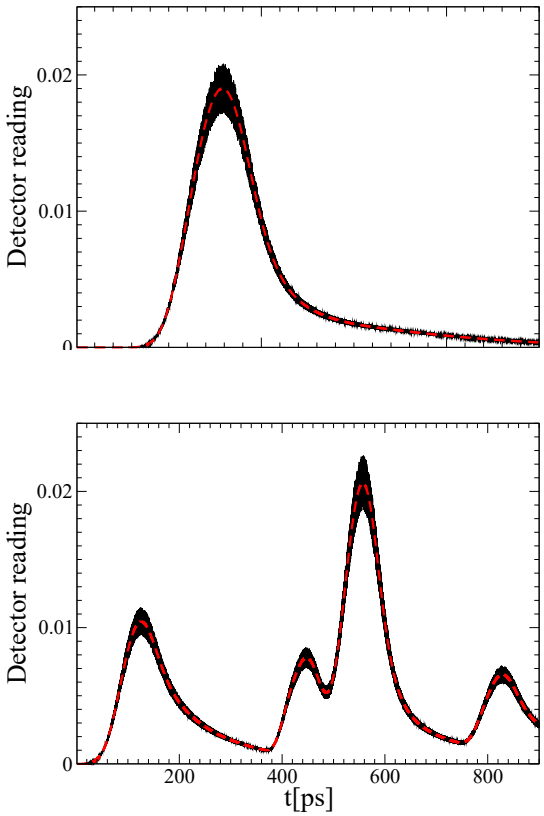

Figure 7: Detector readings ($G_1[u_1]$ in eq. (9)) for the TS and MSS methods, with $N_s = 1$ source for TS method, and $N_s = 4$ sources for the MSS method, and with and without addition of 10% of noise to the signal. The detector was placed at $\mathbf{x}_d = (3.0, 2.25)$. Each single MSS laser beam is located at the center of a face of the square domain, and the activation times $\tau_{1,1} = 0$ps, $\tau_{2,1} = 200$ps, $\tau_{3,1} = 400$ps, and $\tau_{4,1} = 600$ps were used. Clearly, the MSS detector readings combine signals arising from all sources.

are known as prior knowledge on the general absorption properties of the tissue under consideration.

The inverse solver, which is summarized in Algorithm 2, proceeds as follows. Starting from an initial guess $a^0(\mathbf{x})$ for the absorption coefficient and using the experimental detector readings $\widetilde{G}_{j,i}$, with $j = 1, \ldots, N_d$ and $i = 1, \ldots, N_q$, the forward and adjoint problems (2) and (34) are solved,

and the gradient (38) is computed. The functional gradient is then passed to the lm-BFGS algorithm, which returns an updated absorption coefficient $a^1(\mathbf{x})$ that reduces the mismatches between the experimental and simulated detector readings. The procedure is iterated with lm-BFGS convergence to a minimum of the objective function (12).

Algorithm 2 Parallel Inverse Problem Solver

1: **for** iteration $i = 1, \ldots, i_{\max}$ **do**
2: **for** each generalized source $q_j, j = 1, \ldots, N_q$ **do**
3: solve the forward problem by means of algorithm 1
4: solve the adjoint problem by means of algorithm 1
5: **end for**
6: Construct the gradient eq. (38), call lm-BFGS and update the coefficient $a^i(\mathbf{x})$.
7: **end for**

7. Numerical Results II: Inverse problem

This section demonstrates the character of the proposed inverse problem solver for two main model problems, namely, 1) Imaging of cancerous tissue within a human neck section; and, 2) Hemodynamic response in a human head model. In both cases, the optical tomography inverse problems under consideration concerns configurations in which inclusions characterized by absorption higher than that of the surrounding tissue are to be imaged; the higher absorption values arise from the excess of oxygenated hemoglobin under the presence of a tumor in tissue (due to the tumoral angiogenesis) in the neck-tumor problem [15, 16, 17, 22, 23], and from radiation absorption excess originated by the presence of oxygenated hemoglobin triggered by hemodynamic response due to the activation of a brain region [18, 19, 20, 35], in the brain-imaging problem. Accordingly, in what follows the lower bound $a^\ell(\mathbf{x})$ for $a(\mathbf{x})$ in (49) is set to equal the absorption value of the background tissue (which is assumed to be known a priori), and the corresponding upper bound is set to $a^u(\mathbf{x}) = 1/\text{cm}$, which provides a reasonable upper constraint for the absorption values of the types of tissue under consideration [36].

We solve these inverse problems on the basis of synthetic data obtained by running the forward problem for a given "target" absorption coefficient $a^t(\mathbf{x})$. In order to account for experimental noise, we add a 10% of random noise to the resulting detector readings prior to the inversion process, as illustrated in Figure 7. In particular, we study the convergence of Algorithm 2 for a varying number of sources and detectors for the given configuration. In order to evaluate the convergence of the reconstruction process we use the L^2-error norm

$$E(i) = \sqrt{\frac{\int_\Omega (a^t(\mathbf{x}) - a^i(\mathbf{x}))^2 d\mathbf{x}}{\int_\Omega (a^t(\mathbf{x}))^2 d\mathbf{x}}} \tag{50}$$

where $E(i)$ corresponds to the L^2-error in the absorption value $a^i(\mathbf{x})$ obtained at the i-th iteration of the inverse solver.

7.1. Neck tumor imaging

Our first test case concerns the application of optical tomography for diagnoses in patients for which a background Magnetic Resonance Imaging of a neck section is available (Figure 8); such situations arise as e.g. imaging of either evolving tumors or new metastatic tumors is sought within a body part (the neck, in this case) for which an existing MRI image was acquired months or years in advance. The portability and low cost of optical tomographic systems make optical tomographic devices much more accessible than MRI systems for such periodic monitoring and diagnostic applications. Studies for the forward modeling of light propagation in the human neck are reported in [17, 37]. In the present test case we consider the inverse problem for such a configuration (which, of course, requires the repeated solution of forward and adjoint problems, as described in Algorithm 2), focusing on the reconstruction itself and illustrating the convergence character and computational time required by the proposed algorithm. For the present test case the absorption coefficients for the spine, the spinal chord, and the trachea will remain fixed in the reconstruction process (in accordance with the values provided in [37]), since tumoral angiogenesis is only expected to exist in the soft tissue. Additionally, we restrict the presence of tumor inclusions to regions slightly away from the neck boundary, so as to avoid the error amplification associated with the existence of exponential boundary layers [39] (a full treatment of such near-boundary imaging configurations is left for future work). Accordingly, the value of the absorption coefficient in the proximity of the boundary is set to the background tissue value $a(\mathbf{x}) = a_b(\mathbf{x})$ at all points closer than 0.5cm from the boundary.

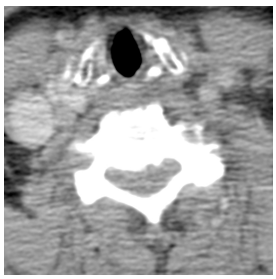

Figure 8: Magnetic Resonance Image [38] for the neck model employed. The absorption and scattering coefficients were set in accordance with reference [37].

We tackle the present neck-imaging problem by means of both the TS and MSS methods described in Section 3.1, in a configuration containing N_b laser beams (see eq. (8)), and we consider examples with $N_b = 4$, 8 and 16. Per the description in that section, a single generalized source containing $N_b = N_s$ beams is utilized in the MSS approach,

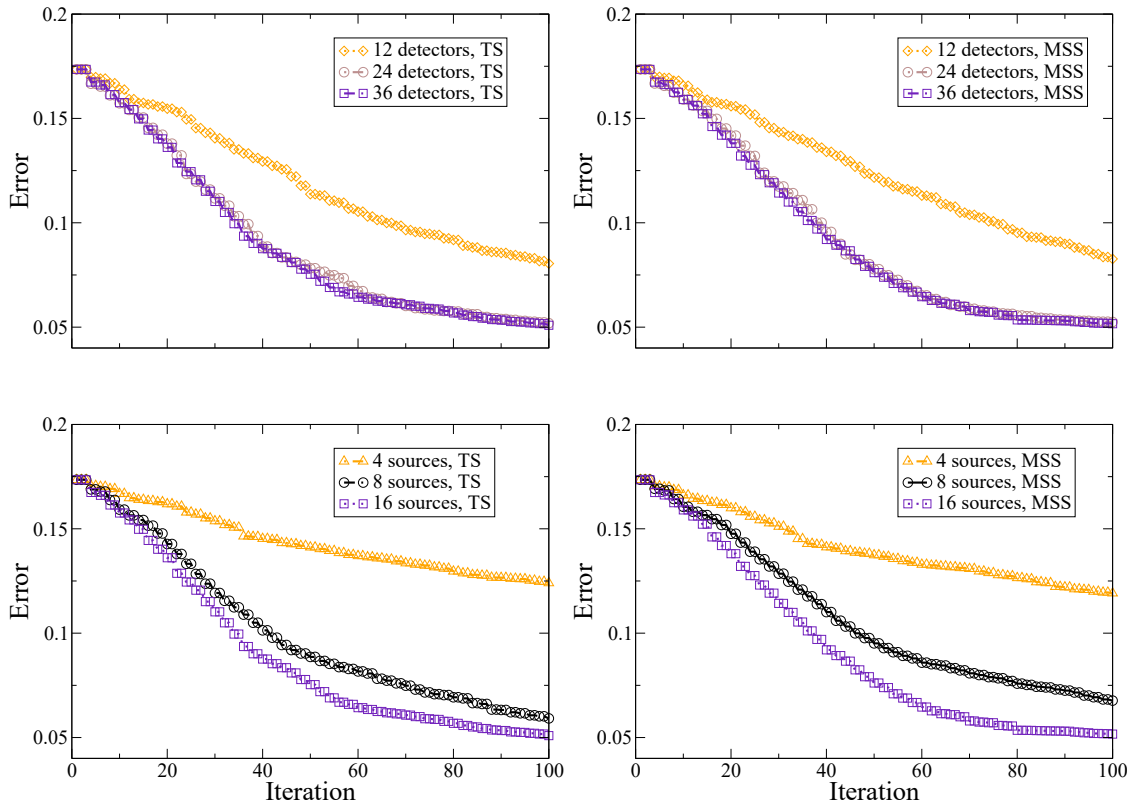

Figure 9: L^2-error norm eq. (50) convergence for the absorption coefficient reconstruction with respect to the number of iterations, for different number of detectors (top) and sources (bottom), for the TS and MSS methods. Left panel: convergence eq. (50) up to 100 iterations for the TS method. Right panel: convergence eq. (50) up to 100 iterations for the MSS method. For the simulations in the top panel 16 laser sources were used. For the simulations in the bottom panel, 36 detectors were used.

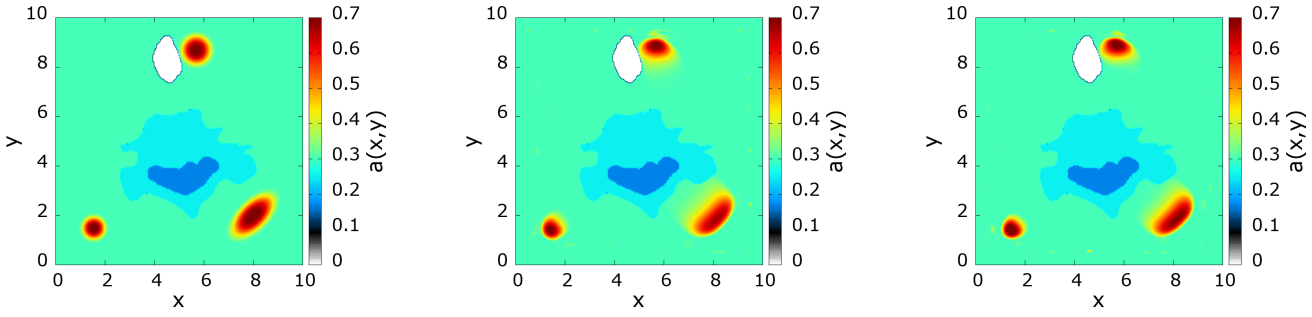

Figure 10: (Color online). From left to right: true absorption coefficient, reconstructed absorption coefficient after 100 iterations using the TS method, and reconstructed absorption coefficient after 100 iterations using the MSS method. For these reconstructions 36 detectors and 16 sources were used.

with $N_q = 1$, and groups of four sources, one per face in the square domain depicted in Figure 8, are simultaneously activated, with time delay of 300ps between groups of four sources. For example, in the case $N_b = 16$ and letting $\mathbf{x}_1^i$ denote the i-th beam in the group of four beams that is activated first we have $\tau_{1,1} = 0$ps for the simultaneous sources placed at $\mathbf{x}_1^1 = (2.0, 0.0)$, $\mathbf{x}_1^2 = (10.0, 2.0)$, $\mathbf{x}_1^3 = (8.0, 10.0)$, $\mathbf{x}_1^4 = (0.0, 8.0)$. The remaining sources, located at point $\mathbf{x}_k^i$, activated at time $\tau_{k,1}$ $(k = 2, 3, 4)$ are arranged as follows: $\mathbf{x}_2^5 = (4.0, 0.0)$, $\mathbf{x}_2^6 = (10.0, 4.0)$, $\mathbf{x}_2^7 = (6.0, 10.0)$, $\mathbf{x}_2^8 = (0.0, 6.0)$ activated at $\tau_{2,1} = 300$ps; $\mathbf{x}_3^9 = (6.0, 0.0)$, $\mathbf{x}_3^{10} = (10.0, 6.0)$, $\mathbf{x}_3^{11} = (4.0, 10.0)$, $\mathbf{x}_3^{12} = (0.0, 4.0)$ activated at $\tau_{3,1} = 600$ps; and $\mathbf{x}_4^{13} = (4.0, 0.0)$, $\mathbf{x}_4^{14} = (10.0, 4.0)$, $\mathbf{x}_4^{15} = (6.0, 10.0)$, $\mathbf{x}_4^{16} = (0.0, 6.0)$ activated at $\tau_{4,1} = 900$ps. This MSS arrangement was obtained by seeking to optimize the required simulation time, which, considering the exponential decay of the photon density wave, was achieved by using simultaneous laser sources that are as far away from each other as possible—and thus facilitate the discrimination of signals received at any given detector. In order to provide a sufficiently long relaxation time for the photon density wave produced by

the last sources to be activated, each forward simulation for the MSS method was carried on up to a final time $t_{\max} = 1400$ps. Similar arrangements, whose details are not provided explicitly for the sake of brevity, where used for the MSS cases $N_b = 4$ and $N_b = 8$. For the TS method, in turn, each independent forward simulation was carried up to the final time $t_{\max} = 600$ps, with detectors and sources placed at the same position as for the single sources of the MSS method with the corresponding value of N_b and using the same numerical grids, and number of processors for each case.

Figure 9 shows that the TS and the MSS methods enjoy similar convergence properties for a varying number N_b of sources and N_d of detectors (cf. equation (14)). However, for the $N_b = 16$ benchmarks, the MSS method required $12,689$ seconds to reach one-hundred iterations of the lm-BFGS algorithm while the TS method, required $88,093$ seconds—so that the MSS method produces an acceleration by a factor of almost seven to achieve one-hundred lm-BFGS iterations. The figure also shows that increasing the number of sources and detectors produces a significant improvement on the convergence of the inverse solver for a fixed number of iterations. The number of sources has a more significant impact on the reconstructions than the number of detectors employed. This can be understood as follows: the role of the sources is to produce the photon density waves that are used to sense the medium. Although the detector reading mismatches $(G_j[u_i] - \widetilde{G}_{j,i})$ are related to the sources for the adjoint problem (34), the intensity of these adjoint sources depends, in turn, on the amount of photons reaching a given detector. It can be argued that for the same reason the number of sources utilized has a more significant impact than whether the sources are run on independent forward simulations (as in the TS method), or simultaneously (as in the MSS method), which makes MSS strategy a reliable and efficient approach.

Figure 10 displays the true absorption coefficient as well as the reconstructed absorption coefficients obtained by both the TS and the MSS methods, with 16 sources and 36 detectors, after 100 iterations of the lm-BFGS algorithm. Clearly, the reconstructions obtained by both methods are of comparable quality, but the MSS reconstruction resulted in a reduction in computing time by a factor of 7.

7.2. Brain imaging based on hemodynamic activation or cancerous tissue

In this final test case we consider a "head model" similar to the one utilized in references [9, 40]. This head model mimics the typical situation where optical tomography is used to study hemodynamic activity in the brain, and captures one of its salient features, namely the clear layer that surrounds the brain. This is a region between the scalp and the brain filled by cerebrospinal fluid with negligible absorption and scattering coefficients—which, in the context of optical tomography, is not suitable for modeling under the diffusion approximation [18] and thus requires use of the full RTE.

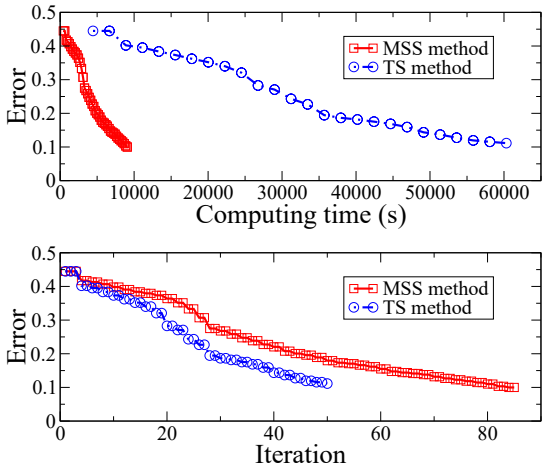

Figure 11: L^2-error norm eq. (50) evolution per iteration for the absorption coefficient reconstruction in the head model for the TS and the MSS method.

In the present context we employ the same number of sources as in reference [9], where a total of 16 sources, with four sources per face, are utilized for a similar head model. For our reconstructions we again look for inclusions over a known background, where the background value is used as the initial guess. We employ 32 detectors, with 8 detectors equally distributed per face. For this benchmark all sources that are placed at the same face of the domain are activated simultaneously. By employing the notation previously described in Section 7.1, the activation configuration for the MSS method for this benchmark is as follows: $\mathbf{x}_1^1 = (1.0, 0.0)$, $\mathbf{x}_1^2 = (2.0, 0.0)$, $\mathbf{x}_1^3 = (3.0, 0.0)$, $\mathbf{x}_1^4 = (4.0, 0.0)$, activated at $\tau_{1,1} = 0$ps; $\mathbf{x}_2^5 = (1.0, 5.0)$, $\mathbf{x}_2^6 = (2.0, 5.0)$, $\mathbf{x}_2^7 = (3.0, 5.0)$, $\mathbf{x}_2^8 = (4.0, 5.0)$ activated at $\tau_{2,1} = 100$ps; $\mathbf{x}_3^9 = (5.0, 1.0)$, $\mathbf{x}_3^{10} = (5.0, 2.0)$, $\mathbf{x}_3^{11} = (5.0, 3.0)$, $\mathbf{x}_3^{12} = (5.0, 4.0)$ activated at $\tau_{3,1} = 400$ps; and $\mathbf{x}_4^{13} = (0.0, 1.0)$, $\mathbf{x}_4^{14} = (0.0, 2.0)$, $\mathbf{x}_4^{15} = (0.0, 3.0)$, $\mathbf{x}_4^{16} = (0.0, 4.0)$ activated at $\tau_{4,1} = 500$ps. Figure 11 displays the evolution of the error eq. (50) for the TS and MSS methods, and Figure 12 presents the true absorption coefficient and the final reconstructed absorption coefficients for the head model.

A reconstruction using the TS method for a total of 50 iterations is presented in the middle panel of Figure 12. The right panel in this figure, in turn, presents the results produced by MSS using 79 iterations, which result in a similar error in the L^2 norm (50) as the one obtained from the 50 TS iterations. The absorption in the clear layer and the region exterior to the clear layer are assumed to be known and remain fixed during the reconstruction process. The search for the inclusions is performed in the region which is surrounded by the clear layer, given that this would be the region where brain activity should be looked for in a real optical tomography experiment. In terms of computational time, the fifty iterations of the TS method

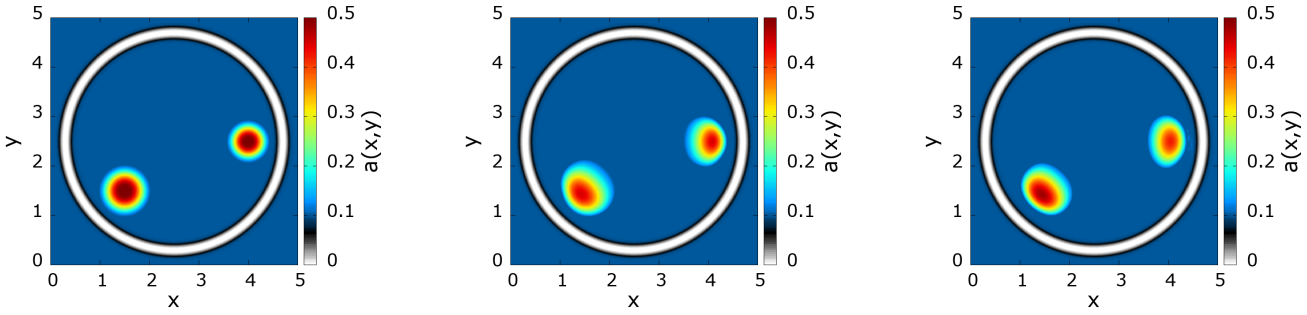

Figure 12: (Color online). From left to right: true absorption coefficient, reconstructed absorption coefficient after 50 iterations of the TS method, and reconstructed absorption coefficient after 79 iterations of the MSS method. For these reconstructions 32 detectors and 16 sources were used.

required more than six times longer than the 79 iterations required by the MSS method to achieve the same error.

8. Conclusions

In this work we have considered the inverse RTE problem in optical tomography by means of a non linear iterative optimization scheme based on the lm-BFGS gradient descent method. We have demonstrated that the proposed domain decomposition parallel strategy presents perfect parallel scaling, making it suitable to tackle the computational demands in optical tomography. Our newly obtained adjoint-gradient expressions under Fresnel boundary conditions enables correct accounting of refractive index mismatches at interfaces. In conjunction with the MSS strategy, the algorithm provided a reduction in computational times of several orders of magnitude over previous approaches. This allowed us to produce reconstructions of the absorption parameter in only a few hours, which in a single processor and by means of the typically used TS method would have required months of computing time to complete.

9. Acknowledgements

This work was supported by NSF, DARPA and AFOSR through contracts DMS-2109831 and HR00111720035 and FA9550-21-1-0373, and by the NSSEFF Vannevar Bush Fellowship under contract number N00014-16-1-2808. ELG acknowledge financial support from CONICET.

References

[1] A. D. Klose. Radiative transfer of luminescence light in biological tissue. In *Light Scattering Reviews 4*, pages 293–345. Springer Praxis Books, 2009.

[2] S. R. Arridge and J. C. Schotland. Optical tomography: Forward and inverse problems. *Inverse Problems*, 25(12), 2009.

[3] E. L. Gaggioli, O. P. Bruno, and D. M. Mitnik. Light transport with the equation of radiative transfer: The Fourier Continuation Discrete Ordinates (FC-DOM) Method. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 236, 2019.

[4] A. H. Hielscher and S. Bartel. Parallel programming of gradient-based iterative image reconstruction schemes for optical tomography. *Computer Methods and Programs in Biomedicine*, 73(2):101–113, 2004.

[5] M. Doulgerakis, A. Eggebrecht, S. Wojtkiewicz, J. Culver, and H. Dehghani. Toward real-time diffuse optical tomography: accelerating light propagation modeling employing parallel computing on GPU and CPU. *Journal of Biomedical Optics*, 22(12):1, 2017.

[6] Pedro J. Coelho. Advances in the discrete ordinates and finite volume methods for the solution of radiative heat transfer problems in participating media. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 145:121–146, 2014.

[7] G. Colomer, R. Borrell, F. X. Trias, and I. Rodríguez. Parallel algorithms for Sn transport sweeps on unstructured meshes. *Journal of Computational Physics*, 232(1):118–135, 2013.

[8] H. Fujii, S. Okawa, Y. Yamada, and Y. Hoshi. Hybrid model of light propagation in random media based on the time-dependent radiative transfer and diffusion equations. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 147:145–154, 2014.

[9] K. Prieto and O. Dorn. Sparsity and level set regularization for diffuse optical tomography using a transport model in 2D. *Inverse Problems*, 33(1), 2017.

[10] O. Dorn. A transport – backtransport method for optical tomography. *Inverse Problems*, 14:1107–1130, 1998.

[11] L. G. Henyey and J. L. Greenstein, Diffuse radiation in the galaxy, *The Astrophysical Journal*, 93:70–83, 1941.

[12] O. P. Bruno and B. Delourme. Rapidly convergent two-dimensional quasi-periodic Green function throughout the spectrum-including Wood anomalies. *Journal of Computational Physics*, 262:262–290, 2014.

[13] J. J. Duderstadt and W. R. Martin, Transport Theory. *John Wiley & Sons*, New York, USA, first edition, 1979.

[14] K. Ren. Recent developments in numerical techniques for transport-based medical imagingmethods. *Communications in Computational Physics*, 8(1):1–50, 2010.

[15] Q. Zhu, S. H. Kurtzman, P. Hegde, S. Tannenbaum, M. Kane, M. Huang, N. Chen, B. Jagjivan, and K. Zarfos. Utilizing optical tomography with ultrasound localization to image heterogeneous hemoglobin distribution in large breast cancers. *Neoplasia*, 7(3):263–270, 2005.

[16] Q. Zhu, P. U. Hegde, A. Ricci, M. Kane, E. B. Cronin, Y. Ardeshirpour, C. Xu, A. Aguirre, S. H. Kurtzman, P. J. Deckers, and S. H. Tannenbaum. Early-stage invasive breast cancers: Potential role of optical tomography with US localization in assisting diagnosis. *Radiology*, 256(2):367–378, 2010.

[17] H. Fujii, Y. Yamada, K. Kobayashi, M. Watanabe, and Y. Hoshi. Modeling of light propagation in the human neck for diagnoses of thyroid cancers by diffuse optical tomography. *International Journal for Numerical Methods in Biomedical Engineering*, 33(5):e2826, 2016.

[18] 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 Processing Magazine*, 18(6):57–75, 2001.

[19] A Bluestone, G Abdoulaev, C Schmitz, R Barbour, and A Hielscher. Three-dimensional optical tomography of hemodynamics in the human head. *Optics express*, 9(6):272–286, 2001.

[20] S. R. Arridge. Optical Tomography in medical imaging. *Inverse Problems*, 15:R41–R93, 1999.

[21] A. D. Klose, V. Ntziachristos, and A. H. Hielscher. The inverse source problem based on the radiative transfer equation in optical molecular imaging. *Journal of Computational Physics*, 202(1):323–345, 2005.

[22] M. Althobaiti, H. Vavadi, and Q. Zhu. Diffuse optical tomography reconstruction method using ultrasound images as prior for regularization matrix. *Journal of Biomedical Optics*, 22(2):026002, 2017.

[23] M. Guven, B. Yazici, X. Intes, and B. Chance. Diffuse optical tomography with a priori anatomical information. *Optical Tomography and Spectroscopy of Tissue V*, 4955:634, 2003.

[24] E. Hille, R.S. Phillips *Functional analysis and semi-groups*. Providence, Rhode Island, USA: American Mathematical Society, first edition, 1957.

[25] T. K. Kim and H. Lee. Effect of anisotropic scattering on radiative heat transfer in two-dimensional rectangular enclosures. *International Journal of Heat and Mass Transfer*, 31(8):1711–1721, 1988.

[26] L. H. Liu, L. M. Ruan and H. P. Tan. On the discrete ordinates method for radiative heat transfer in anisotropically scattering media. *International Journal of Heat and Mass Transfer*, 45(15):3259–3262, 2002.

[27] J. Stoer, R. Bulirsch *Introduction to numerical analysis*. New York, USA: Springer-Verlag, third edition, 2002

[28] O. P. Bruno and M. Cubillos. Higher-order in time "quasi-unconditionally stable" ADI solvers for the compressible Navier–Stokes equations in 2D and 3D curvilinear domains. *Journal of Computational Physics*, 307:476–495, 2016.

[29] F. Amlani and O. P. Bruno. An FC-based spectral solver for elastodynamic problems in general three-dimensional domains. *Journal of Computational Physics*, 307:333–354, 2016.

[30] O. P. Bruno, M. Cubillos and E. Jimenez. Higher-order implicit-explicit multi-domain compressible Navier-Stokes solvers. *Journal of Computational Physics*, 391:322–46, 2019.

[31] O. P. Bruno and M. Lyon. High-order unconditionally stable FC-AD solvers for general smooth domains I. Basic elements. *Computer Physics Communications*, 229(6):2009–33, 2010.

[32] N. Albin and O. P. Bruno. A spectral FC solver for the compressible Navier-Stokes equations in general domains I: Explicit time-stepping. *Journal of Computational Physics*, 230(16):6248–6270, 2011.

[33] M. Fontana, O. P. Bruno, P. D. Mininni and P. Dmitruk. Fourier continuation method for incompressible fluids with boundaries. *Journal of Computational Physics*, 391:322–46, 2020.

[34] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[35] E. Hernandez-Martin and J. L. Gonzalez-Mora Diffuse optical tomography in the human brain: A briefly review from the neurophysiology to its applications. *Brain Science Advances*, 6(4):289–305, 2020.

[36] M. Dehaes, L. Gagnon, F. Lesage, M. Pélégrini-Issac, A. Vignaud, R. Valabrègue, R. Grebe, F. Wallois, and H. Benali, "Quantitative investigation of the effect of the extra-cerebral vasculature in diffuse optical imaging: a simulation study," *Biomedical Optics Express*, vol. 2, no. 3, p. 680, 2011.

[37] H. Fujii, K. Nadamoto, S. Okawa, Y. Yamada, M. Watanabe, Y. Hoshi, and E. Okada. Numerical Modeling of Photon Migration in Human Neck Based on the Radiative Transport Equation. *Journal of Applied Nonlinear Dynamics*, 5(1):117–125, 2016.

[38] Case courtesy of RMH Core Conditions, Radiopaedia.org, rID: 26271.

[39] E. L. Gaggioli, D. M. Mitnik, and O. P. Bruno. Skin effect in neutron transport theory. *Physical Review E*, 104(3):L032801-1–L032801-6, 2021.

[40] A. D Klose, U. Netz, J. Beuthan, and A. H. Hielscher. Optical tomography using the time-independent equation of radiative transfer — Part 1: forward model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 72(5):691–713, 2002.

[41] U. Netz, J. Beuthan, and H. J. Cappius. Imaging of Rheumatoid Arthritis in Finger Joints *Medical Laser Application*, 16:306–310, 2001.