Time-dependent radiative transfer and pulse evolution

J. Tessendorf*

Areté Associates, P.O. Box 6024, Sherman Oaks, California 91413

Received February 1, 1988; accepted August 3, 1988

The time-dependent radiative transfer equation in an absorbing and scattering medium is recast as an evolution equation that is similar to the global formulation of Preisendorfer. All the properties of radiative transfer are embodied in the evolution operator, which can be studied and manipulated in the Dirac operator notation. The reciprocity theorem is obtained in a simple way from the fundamental components of the operator, and in a homogeneous and globally isotropic medium this leads to the statement of the theorem in terms of the radiometric distributions of a laser and a spherical point source. A path integral expression is derived for the evolution operator by using the Dirac notation. By collapsing the Hilbert space of Euclidean and angular degrees of freedom into a finite-sized subspace, an explicit finite-difference numerical scheme is obtained. A feature of the scheme is causal interpolation, which adjusts the spatial interpolation on a step-by-step basis in order to guarantee that the speed of propagation measured from the calculation is the physical (causal) speed to within the spatial and temporal resolution of the calculation. This finite-difference scheme is unconditionally stable and with only weak conditions is consistent. Lax's theorem guarantees that the scheme converges to the continuous solution as the spatial, temporal, and angular grids become arbitrarily fine. Example results from a code written to execute this algorithm are presented. In this example grid and ray affects are illustrated. Finally, the scheme is modified to account for sources of radiant energy located within the medium, and an algorithm is presented to include reflections from surfaces within the medium.

1. INTRODUCTION

The evolution of the radiant distribution of light in a laser pulse as it propagates through a scattering and absorbing medium can in many circumstances be described by the time-dependent radiative transfer (TDRT) equation. Preisendorfer gave an excellent account of the conditions under which this equation is valid,[1] including an outline of the connection between it and the fundamental Maxwell equations. Under the appropriate conditions of validity, the pulse can be characterized by its distribution of energy density in various propagation directions at each point in the medium. The TDRT equation describes the temporal evolution of such a pulse, given some phenomenological optical information about the medium, in the form of the total attenuation coefficient $c(\mathbf{x})$ and the volume-scattering function $\beta(\mathbf{x}, \hat{n}, \hat{n}')$. The dependence of these quantities is on the spatial position $\mathbf{x}$ and the incoming and outgoing directions of propagation $\hat{n}', \hat{n}$. The pulse is characterized by the radiance $L(s, \mathbf{x}, \hat{n})$, which is the power per unit projected area per unit solid angle that would be incident upon a surface located at the position $\mathbf{x}$ from the direction $\hat{n}$. The TDRT equation for the radiance is

$$\left\{\frac{\partial}{\partial s} + \hat{n}\cdot\nabla + c(\mathbf{x})\right\} L(s, \mathbf{x}, \hat{n}) = \int \mathrm{d}^2 n' \beta(\mathbf{x}, \hat{n}, \hat{n}') L(s, \mathbf{x}, \hat{n}'), \quad (1)$$

where s is time measured in units of length with $s = vt$, t is time measured in units of seconds, and v is the speed of light in the medium. The TDRT problem is well posed in the outer product space $R^3 \otimes S^2$ of three-dimensional Euclidean space with the unit sphere when the initial distribution $L_0(\mathbf{x}, \hat{n})$ is specified throughout the space.

A formal solution of Eq. (1) can be written in terms of a space and angle convolution of an evolution operator with the initial distribution and is equivalent to the global formulation discussed by Preisendorfer. This global formulation is a practical starting point for numeric and analytic evaluation schemes because, without approximation or loss of generality, it recasts the TDRT equation as an explicit finite-difference problem. The global evolution formulation can be written compactly by employing the Dirac bra–ket notation widely used in quantum mechanics[2] and by using Hilbert space techniques; this notation is assumed below. The radiance distribution L is represented by an operator $\mathcal{L}$, which depends on time and satisfies the evolution equation

$$\mathcal{L}(s) = \mathcal{G}(s)\mathcal{L}(0), \quad (2)$$

where $\mathcal{G}(s)$ is the evolution operator. These operators act on a Hilbert space with ket vectors $|\mathbf{x}, \hat{n}\rangle$, constructed by using an outer product of the Hilbert spaces of Euclidean space R^3 and the unit sphere S^2:

$$|\mathbf{x}, \hat{n}\rangle = |\mathbf{x}\rangle \otimes |\hat{n}\rangle.$$

The vectors in the corresponding dual space are the bra vectors $\langle \mathbf{x}, \hat{n}|$. The operator $\mathcal{L}$ is related to the radiance distribution by the formula

$$L(s, \mathbf{x}, \hat{n}) = \int \mathrm{d}^3 x' \mathrm{d}^2 n' \langle \mathbf{x}, \hat{n}|\mathcal{L}(s)|\mathbf{x}', \hat{n}'\rangle, \quad (3)$$

so that the components $\langle \mathbf{x}, \hat{n}|\mathcal{L}(s)|\mathbf{x}', \hat{n}'\rangle$ are interpreted as the amount of radiance per unit volume per unit solid angle that originates at time $s = 0$ at position x' propagating in direction $\hat{n}'$ and is at position and direction $(\mathbf{x}, \hat{n})$ at time s. The initial distribution operator can be taken to have the form

$$\langle \mathbf{x}, \hat{n}|\mathcal{L}(0)|\mathbf{x}', \hat{n}'\rangle = L_0(\mathbf{x}, \hat{n})\delta(\mathbf{x} - \mathbf{x}')\delta(\hat{n} - \hat{n}'),$$

and the evolution Eq. (2) is equivalent to

0740-3232/89/020280-18$02.00

$$L(s, \mathbf{x}, \hat{n}) = \int d^3x' d^2n' G(s, \mathbf{x}, \hat{n}; \mathbf{x}', \hat{n}') L_0(\mathbf{x}', \hat{n}'), \quad (4)$$

where $G(s, \mathbf{x}, \hat{n}; \mathbf{x}', \hat{n}')$ is just the component $\langle \mathbf{x}, \hat{n} | \mathcal{G}(s) | \mathbf{x}', \hat{n}' \rangle$. The closure property of the Hilbert space has been used to obtain Eq. (4). The explicit connection between this global evolution formulation and the TDRT equation comes from choosing $\mathcal{G}$ to be the exponentiated operator

$$\mathcal{G}(s) = \exp(-s\mathcal{H}), \quad (5)$$

where the extinction operator $\mathcal{H}$ has components

$$\langle \mathbf{x}, \hat{n} | \mathcal{H} | \mathbf{x}', \hat{n}' \rangle = \{\delta(\hat{n} - \hat{n}')[\hat{n} \cdot \nabla_x + c(\mathbf{x})] - \beta(\mathbf{x}, \hat{n}, \hat{n}')\}\delta(\mathbf{x} - \mathbf{x}'). \quad (6)$$

In operator notation,

$$\mathcal{H} = \hat{\mathcal{N}} \cdot \nabla + \mathcal{C} - \mathcal{B},$$

where

$$\begin{aligned} \langle \mathbf{x}, \hat{n} | \hat{\mathcal{N}} | \mathbf{x}', \hat{n}' \rangle &= \hat{n}\delta(\hat{n} - \hat{n}')\delta(\mathbf{x} - \mathbf{x}'), \\ \langle \mathbf{x}, \hat{n} | \mathcal{C} | \mathbf{x}', \hat{n}' \rangle &= c(\mathbf{x})\delta(\hat{n} - \hat{n}')\delta(\mathbf{x} - \mathbf{x}'), \\ \langle \mathbf{x}, \hat{n} | \mathcal{B} | \mathbf{x}', \hat{n}' \rangle &= \beta(\mathbf{x}, \hat{n}, \hat{n}')\delta(\mathbf{x} - \mathbf{x}'), \\ \langle \mathbf{x}, \hat{n} | \nabla | \mathbf{x}', \hat{n}' \rangle &= \nabla_{\mathbf{x}}\delta(\mathbf{x} - \mathbf{x}')\delta(\hat{n} - \hat{n}'). \end{aligned} \quad (7)$$

Equation (5) follows from the initial condition that $\mathcal{G}(0) = 1$ and from the fact that

$$\left\{\frac{\partial}{\partial s} + \mathcal{H}\right\} \mathcal{G}(s) = 0$$

is equivalent to Eq. (1).

The compact bra–ket notation can be a useful tool. For example, Appendix A uses Eq. (5) as the starting point for a derivation of a path integral expression for the components of $\mathcal{G}$. The bra–ket notation is also used to construct the numerical algorithm in Section 2. As a demonstration, the reversibility property of radiative transfer, known as the reciprocity theorem, can quickly be derived from the fundamental structure of the TDRT equation. For this purpose, a Hermitian idempotent operator $\mathcal{R}$, called the reversal operator, can be defined by the operation

$$\mathcal{R}|\mathbf{x}, \hat{n}\rangle = |\mathbf{x}, -\hat{n}\rangle.$$

Notice from Eq. (6) that when the volume-scattering function satisfies

$$\beta(\mathbf{x}, \hat{n}, \hat{n}') = \beta(\mathbf{x}, -\hat{n}', -\hat{n})$$

the extinction operator has the reversibility property

$$\mathcal{H} = \mathcal{R}\mathcal{H}^\dagger\mathcal{R},$$

from which it follows that the evolution operator also satisfies

$$\mathcal{G} = \mathcal{R}\mathcal{G}^\dagger\mathcal{R}. \quad (8)$$

Equation (8) is the reciprocity theorem. In component form,

$$G(s, \mathbf{x}, \hat{n}; \mathbf{x}', \hat{n}') = G(s, \mathbf{x}', -\hat{n}'; \mathbf{x}, -\hat{n}).$$

The meaning of this theorem is clear: The radiance distribution at $(\mathbf{x}, \hat{n})$, which is to radiance initially at $(\mathbf{x}', \hat{n}')$, is the same as the radiance distribution at $(\mathbf{x}', -\hat{n}')$, which is due to radiance initially at $(\mathbf{x}, -\hat{n})$.

If the medium is homogeneous and globally isotropic, a common method of stating the theorem is to connect the radiance from a spherical-point initial distribution with the irradiance from a point-laser initial distribution. This form of reciprocity follows from the operator form of the theorem. The homogeneity and global isotropy imply that the optical properties c and β are independent of spatial position and that β depends on the directions $\hat{n}$ and $\hat{n}'$ only through the quantity $\hat{n} \cdot \hat{n}'$. It follows that the components $G_h(s, \mathbf{x}, \hat{n}; \mathbf{x}', \hat{n}')$ (the subscript h is a reminder that the medium is homogeneous and globally isotropic) depend on the ten space and direction dimensions only through the four quantities

$$|\mathbf{x} - \mathbf{x}'|,$$
$$(\mathbf{x} - \mathbf{x}') \cdot \hat{n},$$
$$(\mathbf{x} - \mathbf{x}') \cdot \hat{n}',$$
$$\hat{n} \cdot \hat{n}'.$$

With this restriction and the reciprocity theorem, the evolution operator satisfies the particular property

$$\langle \mathbf{x}, \hat{n} | \mathcal{G}_h | \mathbf{x}', \hat{n}' \rangle = \langle \mathbf{x}, \hat{n}' | \mathcal{G}_h | \mathbf{x}', \hat{n} \rangle. \quad (9)$$

The radiance operator for a laser pulse located initially at $(\mathbf{x}_p, \hat{n}_p)$ is

$$\mathcal{L}_p(s) = E_p{}^0 \mathcal{G}_h(s) |\mathbf{x}_p, \hat{n}_p\rangle \langle \mathbf{x}_p, \hat{n}_p|,$$

where $E_p{}^0/v$ is the energy in the initial pulse. This distribution has an irradiance of

$$E_p(s, \mathbf{x} - \mathbf{x}_p, \hat{n}_p) = E_p{}^0 \int d^2n \int d^3x' \, d^2n' \langle \mathbf{x}, \hat{n} | \mathcal{L}_p(s) | \mathbf{x}', \hat{n}' \rangle$$
$$= E_p{}^0 \int d^2n \langle \mathbf{x}, \hat{n} | \mathcal{G}_h(s) | \mathbf{x}_p, \hat{n}_p \rangle.$$

A spherical-point distribution initially located at $\mathbf{x}_s$ has the radiance operator

$$\mathcal{L}_S(s) = L_S{}^0 \int d^2n' \mathcal{G}_h(s) |\mathbf{x}_S, \hat{n}'\rangle \langle \mathbf{x}_S, \hat{n}'|,$$

with total initial energy $4\pi L_S{}^0/v$. The corresponding radiance distribution is

$$L_S(s, \mathbf{x} - \mathbf{x}_S, \hat{n}) = L_S{}^0 \int d^2n' \langle \mathbf{x}, \hat{n} | \mathcal{G}_h(s) | \mathbf{x}_S, \hat{n}' \rangle,$$

and by using Eq. (9), the relation

$$\frac{E_p(s, \mathbf{x}, \hat{n})}{E_p{}^0} = \frac{L_S(s, \mathbf{x}, \hat{n})}{L_S{}^0}$$

emerges. This equation is the explicit statement that the irradiance distribution from a laser point pulse and the radiance distribution from a spherical point pulse are proportional.

As mentioned above, the global evolution operator formulation of TDRT is a convenient starting point for numerical schemes of solving the equation. In fact, Eq. (2) is equivalent to the finite-difference scheme

$$\mathcal{L}(s + \Delta s) = \mathcal{G}(\Delta s)\mathcal{L}(s), \quad (10)$$

which follows from the time-translation invariance of the

TDRT equation. In particular, time-translation invariance implies that

$$\mathcal{G}(s + \Delta s) = \mathcal{G}(\Delta s)\mathcal{G}(s),$$

and this property gives Eq. (10) when applied to Eq. (2). The construction of a numerical scheme can then focus on developing approximate and/or discretized versions of $\mathcal{G}(\Delta s)$, $\mathcal{L}(s)$, $\mathcal{L}(s + \Delta s)$, and the product of operators. In fact, this is the approach taken in the sections below in order to construct a real-space finite-difference algorithm. The result has several advantages over other commonly used schemes, such as Monte Carlo simulation,[3] finite-element techniques,[4] and traditional finite-difference approximation of the TDRT equation.[5] Some of these advantages include unconditional stability and great flexibility in choosing a spatial interpolation algorithm that maintains consistency. As with other finite-difference (and finite-element) methods, most interpolation schemes permit unphysical (acausal) propagation of the light distribution; that is, the apparent speed of light is faster than the physical speed. A general framework has been found, however, in the form of criteria for choosing the spatial points to be used in the interpolation, which guarantees causal propagation at the physical speed of light.

The construction of this finite-difference scheme is the subject of Section 2. The development is restricted to the homogeneous and globally isotropic problem in which c and β are constant throughout the medium and β has the form

$$\beta(\hat{n}, \hat{n}') = bP(\hat{n} \cdot \hat{n}').$$

The constant b is the scattering coefficient, and P is the unit-normalized phase function that satisfies

$$\int d^2nP(\hat{n} \cdot \hat{n}') = 1. \tag{11}$$

The limitations of the real-space algorithm in terms of spatial and temporal resolution are discussed. The special interpolation algorithm for ensuring the correct physical propagation speed is called causal interpolation. The set of rules for causal interpolation is also presented in Section 2, and the reasoning behind the rules and justification for them are reserved for Section 4. Section 3 demonstrates that the algorithm is unconditionally stable and conditionally consistent, and, according to Lax's theorem, the algorithm is convergent when the appropriate consistency requirements are satisfied.[6] Section 5 presents an example calculation of a cylindrically symmetric pulse propagating through a homogeneous medium with an isotropic phase function. In this example ray and grid affects are demonstrated.

If a source of radiant energy is present in the medium, the TDRT equation must be modified in an appropriate way. Section 6 develops the formalism and numerical algorithm, including a source. Also in Section 6, an algorithm is presented for reflection of a propagating pulse from a surface characterized by its bidirectional reflectivity distribution function.

2. FINITE-DIFFERENCE EVOLUTION ALGORITHM

The development of a finite-difference algorithm that is based on Eq. (10) and is suitable for numerical implementation requires a series of steps for discretizing the spatial and angular dimensions to a finite set of degrees of freedom. This is accomplished by following the steps below in order:

1. Choose a partition of the unit sphere S^2 that reduces the set of directions of propagation to a finite set $\{\hat{n}_k\}$, $k = 1, \ldots, N$, and a corresponding set of solid angles $\{\Delta\Omega_k\}$. By using this partition, the components of the evolution operator can be explicitly constructed in Fourier-space conjugate to the real-space R^3.
2. Restrict the size of the time-step Δs to be smaller than one scattering length, i.e., $\Delta s \ll 1/b$. This allows for an approximation of the expression for the evolution operator, which is necessary for the construction of a real-space algorithm, and places an upper bound on the achievable spatial resolution.
3. Choose a set of spatial points (which need not be uniform) and implement the causal interpolation algorithm.

Along with the execution of these steps, the properties of the algorithm that provide for energy conservation are illustrated.

A partition of the unit sphere consists of a set of N bins on the surface of the sphere, which are labeled by the index k and do not overlap, although neighboring bins may share a common boundary. The points on the boundary constitute a set of measure zero and thus do not affect the outcome of the discretization. Each bin is characterized by a unit vector $\hat{n}_k$ pointing in the direction of the centroid of the bin and a solid angle $\Delta\Omega_k$. Because the bins do not overlap,

$$\sum_{k=1}^{N} \Delta\Omega_k = 4\pi.$$

To find an explicit expression for the components of $\mathcal{G}(\Delta s)$ on this partitioned set of propagation directions, it is necessary to collapse the infinite-dimensional Hilbert space of propagation directions with basis kets $|\hat{n}\rangle$ to a finite-dimensional Hilbert subspace whose basis kets are denoted $|k\rangle$. By defining $|k\rangle$ in terms of its action on the basis members of the full space, it is demonstrated that this set constitutes a basis of the collapsed Hilbert space. The ket $|k\rangle$ acting on $|\hat{n}\rangle$ produces the result

$$\langle k|\hat{n}\rangle = \begin{cases} 0 & \text{if } \hat{n} \text{ is not in bin } k \\ 1/\Delta\Omega_k & \text{if } \hat{n} \text{ is in bin } k \end{cases}. \tag{12}$$

This definition is summarized in the expansion

$$|k\rangle = \frac{1}{\Delta\Omega_k} \int_k d^2n|\hat{n}\rangle,$$

where $\int_k d^2n$ means angular integration within the solid angle $\Delta\Omega_k$ of bin k. From this definition it follows that the collapsed basis is orthogonal. When the closure of the full Hilbert space is used,

$$\langle k|k'\rangle = \int d^2n\langle k|\hat{n}\rangle\langle\hat{n}|k'\rangle$$

$$= \left(\frac{1}{\Delta\Omega_k}\right)\delta_{kk'}.$$

Because the basis kets are not normalized to a magnitude of unity, the identity operator in this subspace has the closure expansion

$$\sum_{k=1}^{N} \Delta\Omega_k |k\rangle\langle k| = 1. \tag{13}$$

The behavior of operators in the collapsed space can be obtained from their behavior in the full space coupled with Eqs. (12) and (13). The radiance operator at time $s = 0$ is

$$\begin{aligned} L_{kk'}(0, \mathbf{x}; \mathbf{x}') &= \langle \mathbf{x}, k|\mathcal{L}(0)|\mathbf{x}', k'\rangle \\ &= \int \mathrm{d}^2 n \int \mathrm{d}^2 n' \langle k|\hat{n}\rangle \langle \mathbf{x}, \hat{n}|\mathcal{L}(0)|\mathbf{x}', \hat{n}'\rangle \langle \hat{n}'|k'\rangle \\ &= \frac{1}{\Delta\Omega_k} \frac{1}{\Delta\Omega_{k'}} \int_k \mathrm{d}^2 n \int_{k'} \mathrm{d}^2 n' \delta(\hat{n} - \hat{n}') \delta(\mathbf{x} - \mathbf{x}') \\ &\quad \times L(0, \mathbf{x}, \hat{n}) \\ &= \frac{1}{\Delta\Omega_k} L_k(0, \mathbf{x}) \delta_{kk'} \delta(\mathbf{x} - \mathbf{x}'), \end{aligned}$$

where the radiance in bin k is the average over that bin

$$L_k(s, \mathbf{x}) = \frac{1}{\Delta\Omega_k} \int_k \mathrm{d}^2 n L(s, \mathbf{x}, \hat{n}).$$

This form for the radiance in a bin follows from the collapse of Eq. (3) onto the discrete subspace

$$L_k(s, \mathbf{x}) = \sum_{k'=1}^{N} \Delta\Omega_{k'} \int \mathrm{d}^3 x' \langle \mathbf{x}, k|\mathcal{L}(s)|\mathbf{x}', k'\rangle$$

and from the use of Eq. (12) and the closure relation in Eq. (13). When Eq. (10) is applied to the right-hand side, the discrete evolution equation is

$$\begin{aligned} L_k(s + \Delta s, \mathbf{x}) &= \sum_{k'=1}^{N} \Delta\Omega_{k'} \int \mathrm{d}^3 x' \langle \mathbf{x}, k|\mathcal{G}(\Delta s)\mathcal{L}(s)|\mathbf{x}', k'\rangle \\ &= \sum_{k'=1}^{N} \Delta\Omega_{k'} \int \mathrm{d}^3 x' \sum_{k''=1}^{N} \Delta\Omega_{k''} \int \mathrm{d}^3 x'' \\ &\quad \times \langle \mathbf{x}, k|\mathcal{G}(\Delta s)|\mathbf{x}'', k''\rangle \langle \mathbf{x}'', k''|\mathcal{L}(s)|\mathbf{x}', k'\rangle \\ &= \sum_{k'=1}^{N} \Delta\Omega_{k'} \int \mathrm{d}^3 x' \langle \mathbf{x}, k|\mathcal{G}(\Delta s)|\mathbf{x}', k'\rangle L_{k'}(s, \mathbf{x}'). \end{aligned} \tag{14}$$

Because the medium is homogeneous by assumption, the convolution in the $\mathbf{x}'$ variable in Eq. (14) can be converted to a simple product by Fourier transforming the equation. In the operator formalism this is accomplished by introducing the conjugate basis with kets $|\mathbf{q}\rangle$, which has the closure relation

$$\int \frac{\mathrm{d}^3 q}{(2\pi)^3} |\mathbf{q}\rangle\langle\mathbf{q}| = 1$$

and satisfies the conjugation property

$$\langle \mathbf{x}|\mathbf{q}\rangle = \exp(i\mathbf{q} \cdot \mathbf{x}).$$

The expression for the radiance in this Fourier-transform basis is identical to Eq. (14) when $\mathbf{x}$ is replaced everywhere by $\mathbf{q}$. Because of homogeneity, however, the convolution can be evaluated, and the result is an evolution equation in terms of a matrix of equations. In particular, homogeneity implies that

$$\langle \mathbf{q}, k|\mathcal{G}(\Delta s)|\mathbf{q}', k'\rangle \Delta\Omega_{k'} = \delta(\mathbf{q} - \mathbf{q}') G_{kk'}(\Delta s, \mathbf{q}).$$

The solid angle $\Delta\Omega_k$ has been incorporated into this definition in order to remove it from the matrix expressions below.

The quantities $G_{kk'}(\Delta s, \mathbf{q})$ are the matrix elements of the exponentiated matrix

$$\mathbf{G}(\Delta s, \mathbf{q}) = \exp[-\Delta s \mathbf{H}(\mathbf{q})]$$

and

$$\mathbf{H}(\mathbf{q}) = i\hat{\mathbf{N}} \cdot \mathbf{q} + c - b\mathbf{P}.$$

The extinction matrix $\mathbf{H}$ is obtained from the extinction operator $\mathcal{H}$ by using the relation

$$\langle \mathbf{q}, k|\mathcal{H}|\mathbf{q}', k'\rangle \Delta\Omega_{k'} = \delta(\mathbf{q} - \mathbf{q}') H_{kk'}(\mathbf{q}).$$

From Eqs. (7), the terms $\hat{\mathbf{N}}$ and $\mathbf{P}$ have matrix elements

$$\hat{N}_{kk'} = \delta_{kk'} \frac{1}{\Delta\Omega_k} \int_k \mathrm{d}^2 n \, \hat{n}, \tag{15}$$

$$P_{kk'} = \frac{1}{\Delta\Omega_k} \int_k \mathrm{d}^2 n \int_{k'} \mathrm{d}^2 n' P(\hat{n} \cdot \hat{n}'). \tag{16}$$

$\mathbf{P}$ is called the phase matrix, and its elements are dimensionless. When the number of bins N is large, the average of the unit vectors in a bin can be approximated by just the unit vector at the centroid of the bin. In the implementation of the numerical algorithm this has been assumed, and there is a replacement of

$$\hat{N}_{kk'} = \delta_{kk'} \hat{n}_k.$$

The full TDRT equation in the partition basis is

$$L_k(s + \Delta s, \mathbf{q}) = \sum_{k'=1}^{N} G_{kk'}(\Delta s, \mathbf{q}) L_{k'}(s, \mathbf{q}). \tag{17}$$

The matrix quantities $\mathbf{G}$ and $\mathbf{H}$ satisfy a conservation property and a symmetry property, which can be useful. In particular, the normalization of the phase function in Eq. (11) implies that the phase matrix satisfies the condition

$$\sum_{k'=1}^{N} P_{kk'} = 1. \tag{18}$$

When this property is used, the extinction matrix satisfies the condition

$$\sum_{k'=1}^{N} H_{kk'} = i\hat{n}_k \cdot \mathbf{q} + a.$$

The structures of matrices $\hat{\mathbf{N}}$ and $\mathbf{P}$ in Eqs. (15) and (16) imply the symmetry properties

$$\Delta\Omega_k \hat{N}_{kk'} = \Delta\Omega_{k'} \hat{N}_{k'k},$$

$$\Delta\Omega_k P_{kk'} = \Delta\Omega_{k'} P_{k'k},$$

which consequently apply to the full extinction matrix and to the evolution matrix:

$$\Delta\Omega_k G_{kk'}(\Delta s, \mathbf{q}) = \Delta\Omega_{k'} G_{k'k}(\Delta s, \mathbf{q}). \tag{19}$$

These conservation and symmetry properties are applied below to the real-space algorithm in order to obtain equations for energy and irradiance flux.

A real-space algorithm cannot be obtained directly from Eq. (17) because a Fourier transformation of that equation cannot be done analytically. An analytic transformation would be possible if the evolution matrix could be written in the form

$$\mathbf{G}(\Delta s, \mathbf{q}) \sim \mathbf{G}_1(\Delta s)\exp[-i\mathbf{q}\cdot\mathbf{M}(\Delta s)],$$

where $\mathbf{G}_1$ and $\mathbf{M}$ are some matrices. However, this decomposition is not possible because $\mathbf{P}$ and $\hat{\mathbf{N}}$ do not commute. The alternative is to restrict the size of the time step Δs and to place a lower bound on the smallest resolvable scale $\Delta x \equiv 2\pi/\max\{|q|\}$ included in a calculation. The result, which will be shown below, is that the evolution matrix can be approximated by

$$\mathbf{G}(\Delta s, \mathbf{q}) \approx \exp(-c\Delta s)\mathbf{T}(\Delta s)\exp(-i\Delta s\hat{\mathbf{N}}\cdot\mathbf{q}), \tag{20}$$

where

$$\mathbf{T}(\Delta s) = \exp(b\Delta s\mathbf{P})$$

is the transition matrix. Because $\hat{\mathbf{N}}$ is a diagonal matrix, the Fourier transformation can be evaluated to produce the real-space algorithm

$$L_k(s+\Delta s, \mathbf{x}) = \exp(-c\Delta s)\sum_{k'=1}^{N} T_{kk'}(\Delta s)L_{k'}(s, \mathbf{x}-\hat{n}_{k'}\Delta s). \tag{21}$$

A complete numerical algorithm follows from Eq. (21) when a set of spatial grid points $\{\mathbf{x}_i\}$ and an interpolation scheme for estimating the radiance at the points $\mathbf{y}_{ik} \equiv \mathbf{x}_i - \hat{n}_k\Delta s$ are chosen. As is demonstrated in Section 3, there is considerable freedom in choosing the spatial grid relative to the partition and time step, but Section 4 shows that the interpolation algorithm must be chosen carefully in order for us to ensure the correct speed of propagation.

An interpretation of Eq. (21) is that the radiance propagating in the direction $\hat{n}_k$ at the point $\mathbf{x}$ arrived there by propagating along the line of sight from the surrounding points $\mathbf{x} - \hat{n}_{k'}\Delta s$ to $\mathbf{x}$, then scattering from the arriving direction to the direction $\hat{n}_k$. In each time step one scattering event occurs, so that a distribution that includes multiple-scattering events is achieved after several time steps.

The approximation of $\mathbf{G}$ by expression (20) and the restrictions on time-step and spatial resolution that make the approximation valid follow when we consider the exponentiation of the sum of two matrices $\mathbf{A}$ and $\mathbf{B}$, which do not commute. In the context of radiative transfer, $\mathbf{A} = -b\mathbf{P}$ and $\mathbf{B} = i\hat{\mathbf{N}}\cdot\mathbf{q}$. In general, the exponentiated sum can be decomposed as

$$\exp[-s(\mathbf{A}+\mathbf{B})] = \exp(-s\mathbf{A})\exp(-s\mathbf{B})\exp[\mathbf{C}(s)],$$

where $\mathbf{C}(s)$ is some matrix constructed from $\mathbf{A}$, $\mathbf{B}$, and s. In fact, $\mathbf{C}$ satisfies the initial value problem

$$\mathbf{C}(0) = 0,$$

$$\frac{\partial\mathbf{C}(s)}{\partial s} = \exp(s\mathbf{B})\exp(s\mathbf{A})[\exp(-s\mathbf{A}), \mathbf{B}]\exp(-s\mathbf{B}).$$

By expanding each of the exponents in a Taylor expansion and by collecting powers of s, the differential equation can be integrated to give

$$\mathbf{C}(s) = \sum_{j=1}^{\infty}\frac{s^{j+1}}{(j+1)!}\mathbf{W}_j,$$

where the coefficient matrices are

$$\mathbf{W}_j = \sum_{k_1,k_2,k_3=0}^{j}\frac{j!}{k_1!k_2!k_3!(j-k_1-k_2-k_3)!}(-1)^{k_2+k_3} \times \mathbf{B}^{j-k_1-k_2-k_3}\mathbf{A}^{k_1}[\mathbf{A}^{k_2}, \mathbf{B}]\mathbf{B}^{k_3}.$$

The first term of $\mathbf{C}$ in this case is

$$\mathbf{C}(\Delta s) \approx -b\frac{(\Delta s)^2}{2}i(\hat{\mathbf{N}}\cdot\mathbf{q}, \mathbf{P}).$$

The requirement necessary in order for expression (20) to be valid is that this term be small in an asymptotic sense compared with the terms that were kept. When $\mathbf{C}$ is compared with $i\Delta s\hat{\mathbf{N}}\cdot\mathbf{q}$, the first condition is that

$$b\Delta s \ll 1, \tag{22}$$

i.e., that the time step be small compared with the scattering length. This condition ensures that there is a negligible amount of scattering during a single time step. The second condition comes from requiring that $\mathbf{C}$ be small compared with $b\Delta s\mathbf{P}$,

$$\Delta s|q| \ll 1.$$

For a particular numerical calculation, the chosen spatial grid has a minimum resolvable length scale denoted Δx. The second condition requires that

$$\Delta x \gg \frac{\Delta s}{2\pi}, \tag{23}$$

which ensures that the scattering process occurs at the grid points and does not have a contribution from multiple scattering along a path of length Δs. If the spatial resolution were finer than $\Delta s/2\pi$, the scattering that physically occurs at the finer scales would not be accounted for accurately by Eq. (21). Using a spatial grid that includes finer scales is not harmful, however, because finer scales do not degrade the stability and consistency of the algorithm. It should be kept in mind, however, that the behavior of a numerical solution on spatial scales smaller than $\Delta s/2\pi$ is not necessarily a good reproduction of the correct physical behavior.

The two conditions also ensure that higher-order terms in $\mathbf{C}$ are small in the asymptotic sense as well.

It is instructive to compare this form of the evolution equation with Eq. (1) and in particular to examine the mechanisms that each has for the absorption, transport, and conservation of energy and irradiance flux. The purpose of this comparison is simply to show that the equations for irradiance and energy transport from the partitioned equation can be matched term by term with corresponding quantities in the continuous problem.

In the continuous problem the scalar and vector irradiance are defined, respectively, as

$$E(s, \mathbf{x}) = \int \mathrm{d}^2n L(s, \mathbf{x}, \hat{n}),$$

$$\mathbf{E}(s, \mathbf{x}) = \int \mathrm{d}^2n\hat{n}L(s, \mathbf{x}, \hat{n})$$

and from Eq. (1) satisfy ($a = c - b$ is the absorption coefficient)

$$\left(\frac{\partial}{\partial s} + a\right) E(s, \mathbf{x}) = -\nabla \cdot \mathbf{E}(s, \mathbf{x}). \tag{24}$$

Equation (24) can be converted to the finite-difference form

$$E(s + \Delta s, \mathbf{x}) = \exp(-a\Delta s) \times \left\{E(s, \mathbf{x}) - \int_s^{s+\Delta s} ds' \exp[-a(s - s')]\nabla \cdot \mathbf{E}(s', \mathbf{x})\right\}.$$

Similarly, the energy contained in a volume V within the medium is

$$e_V(s) = \frac{1}{v}\int_V d^3x\, E(s, \mathbf{x}),$$

and it satisfies

$$\left(\frac{\partial}{\partial s} + a\right) e_V(s) = -e_{\partial V}(s).$$

The energy flux $e_{\partial V}(s)$ is the rate of energy transfer out of the volume V through its boundary ∂V:

$$e_{\partial V}(s) = \frac{1}{v}\int_{\partial V} dA\, \hat{n}_V(\mathbf{x}) \cdot \mathbf{E}(s, \mathbf{x}).$$

In this expression dA is the area measure on the boundary of the volume and $\hat{n}_V(\mathbf{x})$ is the outward normal of the boundary. This differential equation can similarly be converted to the finite-difference form

$$e_V(s + \Delta s) = \exp(-a\Delta s) \times \left\{e_V(s) - \int_s^{s+\Delta s} ds' \exp[-a(s - s')]e_{\partial V}(s')\right\}.$$

To construct the corresponding equations in the partitioned basis for the finite-difference algorithm of Eq. (21), the normalization and symmetry properties of Eqs. (18) and (19) can be used to obtain the relationships for $\mathbf{T}$:

$$\sum_{k'=1}^{N} T_{kk'}(\Delta s) = \exp(b\Delta s), \tag{25}$$

$$\Delta\Omega_k T_{kk'}(\Delta s) = \Delta\Omega_{k'} T_{k'k}(\Delta s). \tag{26}$$

The scalar and vector irradiances in the partition basis are, respectively,

$$E^p(s, \mathbf{x}) = \sum_{k=1}^{N} \Delta\Omega_k L_k(s, \mathbf{x}),$$

$$\mathbf{E}^p(s, \mathbf{x}) = \sum_{k=1}^{N} \Delta\Omega_k L_k(s, \mathbf{x})\hat{n}_k,$$

where the superscript p refers to the partitioned basis. From Eqs. (21), (25), and (26), the irradiance transport equation is

$$E^p(s + \Delta s, \mathbf{x}) = \exp(-c\Delta s) \sum_{k,k'=1}^{N} \Delta\Omega_k T_{kk'}(\Delta s) L_{k'}(s, \mathbf{x} - \hat{n}_{k'}\Delta s)$$

$$= \exp(-a\Delta s)[E^p(s, \mathbf{x}) - \delta E^p(s, \mathbf{x})],$$

where

$$\delta E^p(s, \mathbf{x}) = \Delta s \sum_{k=1}^{N} \Delta\Omega_k \left[\frac{L_k(s, \mathbf{x}) - L_k(s, \mathbf{x} - \hat{n}_k\Delta s)}{\Delta s}\right].$$

The quantity δE^p is the discrete form of the continuous flux out of the volume

$$\int_s^{s+\Delta s} ds' \exp[-a(s - s')]\nabla \cdot \mathbf{E}(s', \mathbf{x}).$$

Note that in the limit when $\Delta s \to 0$ this equation becomes Eq. (24) because

$$\lim_{\Delta s \to 0}\left[\frac{L_k(s, \mathbf{x}) - L_k(s, \mathbf{x} - \hat{n}_k\Delta s)}{\Delta s}\right] = \hat{n}_k \cdot \nabla L_k(s, \mathbf{x});$$

thus

$$\delta E^p(s, \mathbf{x}) \to \Delta s \nabla \cdot \mathbf{E}^p(s, \mathbf{x}).$$

If the equation for E^p is integrated over a volume V, the equation for the energy in the volume is

$$e_V{}^p(s + \Delta s) = \exp(-a\Delta s)[e_V{}^p(s) - \delta e_V{}^p(s)],$$

where

$$\delta e_V{}^p(s) = \frac{\Delta s}{v}\sum_{k=1}^{N} \Delta\Omega_k \int_V d^3x \left[\frac{L_k(s, \mathbf{x}) - L_k(s, \mathbf{x} - \hat{n}_k\Delta s)}{\Delta s}\right].$$

A partitioned boundary $\partial^p V$ can be defined as the set of points $\mathbf{x}$ in the volume V, for which $\mathbf{x} - \hat{n}_k\Delta s$ is not in the volume for some direction bin k. For points in V that are not on the boundary, the limits of integration may be translated, leaving

$$\delta e_V{}^p(s) = \frac{\Delta s}{v}\sum_{k=1}^{N} \Delta\Omega_k \int_{\partial^p V} d^3x \left[\frac{L_k(s, \mathbf{x}) - L_k(s, \mathbf{x} - \hat{n}_k\Delta s)}{\Delta s}\right],$$

which is the discrete version of the quantity

$$\int_s^{s+\Delta s} ds' \exp[-a(s - s')]e_{\partial V}(s').$$

As in the continuous problem, the quantity $\delta e_V{}^p(s)/\Delta s$ is the energy flux out of the volume V.

The final step necessary to execute Eq. (21) in a numerical algorithm is to choose a discrete set of points $X = \{\mathbf{x}_i, i = 1, \ldots, M\}$ on which to solve the finite-difference equation. For any finite set of points there is a set of one or more points $\mathbf{y}_{ik} \equiv \mathbf{x}_i - \hat{n}_k\Delta s$ that are not grid points; thus it is necessary to obtain the radiance $L_k(s, \mathbf{y}_{ik})$ by interpolation, using nearby grid points. Denoting an interpolation algorithm by $\Im$, the full numerical algorithm is

$$L_k(s + \Delta s, \mathbf{x}_i) = \exp(-c\Delta s) \sum_{k'=1}^{N} T_{kk'}(\Delta s)\Im[L_{k'}(s, \mathbf{x}_i - \hat{n}_{k'}\Delta s)]. \tag{27}$$

A particular interpolation scheme works by choosing a set of points from X that are in some sense near the point $\mathbf{y}_{ik}$ and by choosing a set of unit normalized weights. If the grid X is not a simple, regular lattice, the set of weights and nearby points can depend on the particular point $\mathbf{x}_i$ and the propagation direction $\hat{n}_k$ and are denoted w_{ikl} and $\mathbf{x}_{ikl}$, respective-

ly, where $l = 1, \ldots, M_{ik}$ labels the set of points near $\mathbf{y}_{ik}$. The interpolation scheme is

$$\Im[L_k(s, \mathbf{y}_{ik})] = \sum_l w_{ikl} L_k(s, \mathbf{x}_{ikl}),$$

and the normalization of the weights is

$$\sum_l w_{ikl} = 1.$$

In addition, if some of the weights are negative it is possible to produce a negative interpolated value for the radiance. Because this is not desirable, the restriction

$$w_{ikl} \geq 0$$

is imposed.

Most interpolation schemes, for example, nearest-neighbor and bilinear schemes, use the set of nearest points that completely surround the position $\mathbf{y}_{ik}$. However, $\mathbf{y}_{ik}$ is at a distance Δs from $\mathbf{x}_i$, which is the longest distance that light can physically travel in one time step. Some of the surrounding points used in these interpolation schemes lie at a distance from $\mathbf{x}_i$ that is greater than Δs, and the interpolated radiance includes radiant energy that cannot physically have arrived at $\mathbf{x}_i$ in the time Δs. This acausal behavior occurs in many numerical schemes for radiative transfer, including finite elements and more traditional finite-difference algorithms. In addition, the excess distance and excess apparent speed of light depend on the direction of propagation and position in the grid. This acausal behavior can be overcome by using the following set of steps, called causal interpolation:

(a) A position, direction, and time-dependent length scale $l_{ik}(s)$ is defined, which replaces the time step length in the interpolation formula. The modified algorithm is

$$L_k(s + \Delta s, \mathbf{x}_i) = \exp(-c\Delta s) \sum_{k'=1}^{N} T_{kk'}(\Delta s) \Im\{L_{k'}[s, \mathbf{x}_i - \hat{n}_{k'} l_{ik'}(s)]\}. \tag{28}$$

The initial value $l_{ik}(0)$ is Δs.

(b) Choose the interpolation points $\mathbf{x}_{ikl}(s)$ from the set of points within a distance $l_{ik}(s)$ of $\mathbf{x}_i$ and within the angular bin k. Generate the weights $w_{ikl}(s)$ and interpolate the radiance. Typically, the set of points and weights used varies from time step to time step.

(c) Find the distance

$$\delta_{ik}(s) \equiv \max_l [|\mathbf{x}_i - \mathbf{x}_{ikl}(s)|], \tag{29}$$

which is the greatest distance that light traveled in the direction $\hat{n}_k$ to reach $\mathbf{x}_i$ during this time step.

(d) Update l_{ik} to the next time step by using the procedure

$$l_{ik}(s + \Delta s) = l_{ik}(s) + \Delta s - \delta_{ik}(s). \tag{30}$$

(e) A time-step index $m_{ik}(s)$ is defined to account for situations in which no points except $\mathbf{x}_i$ are found for the interpolation. $m_{ik}(s)$ is equal to the number of time steps backward that the distribution must be sampled. For situations in which points in addition to $\mathbf{x}_i$ were found in the previous time, $m_{ik}(s) = 1$. If no points in addition to $\mathbf{x}_i$ were found, $m_{ik}(s) = m_{ik}(s - \Delta s) + 1$. Equation (28) is the algorithm used when $m_{ik'}(s) = 1$. For larger values, the algorithm is generalized to

$$L_k(s + \Delta s, \mathbf{x}_i) = \sum_{k'=1}^{N} \exp[-cm_{ik'}(s)\Delta s] T_{kk'}(m_{ik'}\Delta s) \times \Im\{L_{k'}[s - (m_{ik'} - 1)\Delta s, \mathbf{x}_i - \hat{n}_{k'} l_{ik'}(s)]\}.$$

The reason for each of these steps and their justification are contained in Section 4. The effect of the causal interpolation scheme is to ensure that at all times the apparent speed of light is equal to the physical speed to within the spatial and temporal resolution of the calculation. In Section 3 it is shown that this scheme does not spoil the consistency properties of the algorithm.

An important concept to introduce when the grid X is not a regular lattice is that of a local grid spacing σ_{ik}, which is the distance from the point $\mathbf{x}_i$ to its nearest neighboring point within the solid angle defined by bin k in the partition. This quantity is used in Sections 3 and 4 in the discussions on interpolation.

3. CONVERGENCE

The fundamental property desired of any finite-difference algorithm is convergence, i.e., the solution of the algorithm on a finite-sized grid should approach the corresponding continuous solution as the grid becomes finer. Lax's theorem[6] shows that the convergence is ensured if and only if the finite-difference algorithm is uniformly stable and consistent with the original continuous equation as the grid becomes finer.

It should be noted that the restrictions placed on the time and space resolution in expressions (22) and (23) are not related to the stability of the algorithm embodied by Eq. (27). These conditions specify the necessary temporal resolution and the minimum grid spacing in order to ensure that the real-space algorithm includes all the desired radiative-transfer processes. It is reasonable to permit the smallest spatial scale in a calculation to be smaller than $\Delta s/2\pi$. However, the solution obtained at the smaller scales will not be reliable because it does not include the small-scale multiple scattering that should physically be present. The solution on the larger scales will contain the correct multiple-scattering and attenuation features.

The stability of the algorithm follows from an examination of the temporal evolution of the maximum value K of the radiance, defined as

$$K(s) = \max_{i,k} [L_k(s, \mathbf{x}_i)].$$

The elements of the transition matrix $\mathbf{T}$ are all positive, and when Eq. (25) is used, the maximum K satisfies

$$K(s + \Delta s) = \exp(-c\Delta s) \max_{i,k} \left\{ \sum_{k'=1}^{N} T_{kk'}(\Delta s) \Im[L_{k'}(s, \mathbf{x}_i - \hat{n}_{k'}\Delta s)] \right\}$$

$$\leq \exp(-a\Delta s) \max_{i,k} \{\Im[L_k(s, \mathbf{x}_i - \hat{n}_k \Delta s)]\}$$

$$\leq \exp(-a\Delta s) K(s).$$

The last inequality follows from the fact that the interpola-

tion weights are nonnegative and normalized to 1. When this result is iterated,

$$K(s) \leq K(0),$$

and the uniform stability of the algorithm is proved.

The consistency of the algorithm depends on how the algorithm reproduces the TDRT equation as the three grids—spatial, angular, and temporal—become arbitrarily dense. For the demonstrations of consistency in this section the causal interpolation scheme is not considered explicitly. Constraints added to ensure consistency under causal interpolation are discussed in Section 4.

The approach to demonstrating consistency is to reproduce the original TDRT equation from the finite-difference scheme. When we begin with Eq. (27), the surviving terms as $\Delta s \to 0$ are

$$\frac{1}{\Delta s}[L_k(s+\Delta s, \mathbf{x}_i) - L_k(s, \mathbf{x}_i)] + \Im[\hat{n}_k \cdot \nabla L_k(s, \mathbf{x}_i)] + cL_k(s, \mathbf{x}_i) = b\sum_{k'} P_{kk'} L_{k'}(s, \mathbf{x}_i),$$

where

$$\Im[\hat{n}_k \cdot \nabla L_k(s, \mathbf{x}_i)] = \frac{1}{\Delta s}\{L_k(s, \mathbf{x}_i) - \Im[L_k(s, \mathbf{x}_i - \hat{n}_k \Delta s)]\}.$$

Consistency follows if, as the grids become dense, the relations

$$\Im\{\hat{n}_k \cdot \nabla L_k(s, \mathbf{x}_i)\} \to \hat{n}_k \cdot \nabla L_k(s, \mathbf{x}_i),$$

$$\sum_{k'} P_{kk'} L_{k'}(s, \mathbf{x}_i) \to \int \mathrm{d}^2 n' P(\hat{n} \cdot \hat{n}') L(s, \mathbf{x}_i, \hat{n}')$$

hold. The second relation follows from the definition of the phase matrix $\mathbf{P}$ in Eq. (16) in the limit when $N \to \infty$ while $\sum_{k=1}^{N} \Delta\Omega_k = 4\pi$. The validity of the first relation depends on the specific spatial interpolation scheme used. In particular, two schemes—linear interpolation and $1/r^2$ weighting—can be demonstrated to satisfy consistency.

In the case of linear interpolation the interpolated radiance is

$$\Im[L_k(s, \mathbf{x}_i - \hat{n}_k \Delta s)] = \left(1 - \frac{\Delta s}{\sigma_{ik}}\right) L_k(s, \mathbf{x}_i) + \frac{\Delta s}{\sigma_{ik}} L_k(s, \mathbf{z}_{ik}),$$

where σ_{ik} is the local grid spacing and $\mathbf{z}_{ik}$ is the grid point at the distance σ_{ik} from $\mathbf{x}_i$ in the partition bin containing the direction $-\hat{n}_k$. It follows that

$$\Im[\hat{n}_k \cdot \nabla L_k(s, \mathbf{x}_i)] = \frac{L_k(s, \mathbf{x}_i) - L_k(s, \mathbf{z}_{ik})}{\sigma_{ik}},$$

and in the limit when the spatial grid becomes dense the first relation is established.

If the time step is larger than the local grid spacing, then some of the weights in the linear interpolation scheme are negative. Because this can induce an artificially negative value for the radiance, linear interpolation may not be the best scheme for a particular application. An alternative scheme with positive weights is a $1/r^2$ method. When we choose a set of points $\mathbf{x}_{ikl}$ in the neighborhood of $\mathbf{x}_i - \hat{n}_k \Delta s$ for the interpolation, they are weighted according to

$$w_{ikl} \propto \frac{1}{{\Delta_{ikl}}^2},$$

where

$${\Delta_{ikl}}^2 = |\mathbf{x}_{ikl} - \mathbf{x}_i + \hat{n}_k \Delta s|^2.$$

Assume that the points $\mathbf{x}_{ikl}$ have been ordered so that $\Delta_{ik1} < \Delta_{ik2} < \Delta_{ik3} < \ldots$. The unit normalization of the weights implies that

$$w_{ikl} = \frac{{\Delta_{ik1}}^2}{{\Delta_{ikl}}^2}\left(\sum_{l'} \frac{{\Delta_{ik1}}^2}{{\Delta_{ikl'}}^2}\right)^{-1}.$$

As the spatial grid becomes dense, the weights w_{ikl} are dominated by those points $\mathbf{x}_{ikl}$ that approach $\mathbf{x}_i - \hat{n}_k \Delta s$, so that the interpolated radiance

$$\Im[L_k(s, \mathbf{x}_i - \hat{n}_k \Delta s)] = \sum_l w_{ikl} L_k(s, \mathbf{x}_{ikl})$$

approaches the desired value of $L_k(s, \mathbf{x}_i - \hat{n}_k \Delta s)$. In this case

$$\Im[\hat{n}_k \cdot \nabla L_k(s, \mathbf{x}_i)] \to \frac{L_k(s, \mathbf{x}_i) - L_k(s, \mathbf{x}_i - \hat{n}_k \Delta s)}{\Delta s},$$

which approaches the desired value as $\Delta s \to 0$, and the first relation is satisfied.

From the two interpolation schemes illustrated, it is clear that essentially any interpolation scheme will give a consistent algorithm if the interpolation becomes more accurate as the spatial grid becomes dense. There are many finite-difference algorithms for TDRT that differ only in their essentials by the method of approximating the spatial derivative along the direction of propagation.[5] However, although many algorithms enjoy consistency, the stability restrictions differ substantially for the different algorithms. The advantage of the present algorithm is that the grids and the interpolation algorithm can be designed to suit particular problems. Because uniform stability is guaranteed regardless of the interpolation algorithm and because consistency follows from reasonably constructed interpolation algorithms, it is relatively easy to construct a convergent algorithm that is tailored to a specific problem.

4. CAUSAL INTERPOLATION

Nearly any reasonable spatial interpolation scheme can produce a convergent TDRT algorithm. However, most schemes also permit the radiance to propagate at a speed faster than the physical speed of light. The reason for this is simple: Interpolation using surrounding points to the point $\mathbf{x}_i - \hat{n}_k \Delta s$ usually includes points that are more than the distance Δs from $\mathbf{x}_i$. Consequently, in a single time step, radiance travels more than Δs, which produces acausal behavior.

To eliminate this problem, the interpolation algorithm could be modified to one in which only points within the distance Δs from $\mathbf{x}_i$ are used. This scheme suffers, however, from a propagation speed that is generally slower than the physical speed. A simple one-dimensional model can illustrate this behavior. Assume a regular one-dimensional array with spacing Δx, and choose $\Delta s = m\Delta x + \eta$, where m is an integer and $0 \leq \eta < \Delta x$. In a single time step the distance moved is $m\Delta x$, so the apparent speed of propagation is

$$v^{\text{app}} = \frac{m}{m + \eta/\Delta x} < 1,$$

which is less than the physical speed (in these units in which time s is measured in units of length, the physical speed of light is 1). In particular, if $\Delta s < \Delta x$, then $m = 0$, and there is no movement of the radiance. By making $\Delta s > \Delta x$ and η sufficiently small, the apparent speed v^{app} approaches the physical speed to within the desired accuracy. In a higher-dimensional calculation, however, η cannot easily be made small in all directions simultaneously, and in practical applications, when computing resources are limited, the discrepancy between apparent and physical speeds can be significant. Also, in higher-dimensional calculations the apparent speed would depend on the direction of propagation and the structure of the spatial grid.

The fundamental difficulty is that in all the interpolation schemes discussed, the distance traveled in each time step is always less than (or greater than) the physically required distance. The error in the distance propagated may be less than the local grid spacing and so is not measureable. The error accumulates, however, and after some number of time steps exceeds the threshold of resolution.

The solution of this problem is to adjust the distance actually propagated in each time step in a way that ensures that the accumulated difference between the actual distance and the physically required distance does not exceed the local grid spacing, i.e., is not measureable. Steps (a)–(e) in Section 2 succeed in accomplishing this goal and maintaining consistency of the algorithm, with a weak restriction that the spatial, angular, and temporal grids must be sufficiently dense to implement step (e) when necessary. The term causal interpolation arises from the intent of the algorithm to preserve the physical (causal) propagation speed.

Two properties of causal interpolation demonstrate that it is an algorithm that meets the above objectives:

1. Consistency follows from the property that

$$\Delta s \le l_{ik}(s) < \Delta s + \sigma_{ik},$$

where σ_{ik} is the local grid spacing, because $l_{ik} \to \Delta s$ as $\sigma \to 0$, and the original finite-difference algorithm is restored.

2. After m time steps, the apparent speed is

$$v_{ik}{}^{\text{app}}(s) = 1 + O\left(\frac{\sigma_{ik}}{s}\right),$$

where $s = m\Delta s$.

Both of these properties are proved below.

Property 2 follows, assuming that property 1 is valid. From the definition of δ_{ik} in Eq. (29), the apparent speed at position $\mathbf{x}_i$ in the direction $\hat{n}_k$ after m time steps is

$$v_{ik}{}^{\text{app}} = \frac{1}{s}\sum_{m'=0}^{m} \delta_{ik}(m'\Delta s),$$

where $s = m\Delta s$. From Eq. (30),

$$v_{ik}{}^{\text{app}}(s) = 1 + \sum_{m'=0}^{m}\left\{\frac{l_{ik}[(m'+1)\Delta s] - l_{ik}(m'\Delta s)}{s}\right\}$$

$$= 1 + \left[\frac{l_{ik}(s+\Delta s) - l_{ik}(0)}{s}\right].$$

From property 1, $|l_{ik}(s + \Delta s) - l_{ik}(0)| < \sigma_{ik}$, i.e., it is bounded, and property 2 follows. As a result of this property, the apparent speed approaches the physical speed as more time steps are calculated, and the error at any individual time step is below the spatial and temporal resolution.

An inductive proof of property 1 comes in two parts. In part 1 we assume that $\Delta s > \sigma_{ik}$, and write

$$\sigma_{ik} = \alpha_{ik}\Delta s,$$

where $\alpha_{ik} < 1$. Also, we assume that $l_{ik}(s) = \Delta s + \epsilon_{ik}$, with $\epsilon_{ik} < \sigma_{ik}$. The largest distance traveled is

$$\delta_{ik}(s) = n_{ik}\sigma_{ik}$$

$$= n_{ik}\alpha_{ik}\Delta s,$$

where

$$n_{ik} = \text{int}\left(\frac{1 + \epsilon_{ik}/\Delta s}{\alpha_{ik}}\right).$$

For the next time step,

$$l_{ik}(s + \Delta s) = \Delta s + \Delta s\left(1 + \frac{\epsilon_{ik}}{\Delta s} - \alpha_{ik}n_{ik}\right).$$

The quantity in the parentheses on the right-hand side satisfies the inequality

$$1 + \frac{\epsilon_{ik}}{\Delta s} - \alpha_{ik}n_{ik} < \alpha_{ik},$$

and this produces the desired result:

$$l_{ik}(s + \Delta s) < \Delta s + \alpha_{ik}\Delta s$$

$$< \Delta s + \sigma_{ik}.$$

In part 2, we assume that $\Delta s < \sigma_{ik}$. In this case radiance does not propagate for $m_{ik} - 1$ time steps, where

$$m_{ik}\Delta s + \epsilon_{ik} \ge \sigma_{ik},$$
$$(m_{ik} - 1)\Delta s + \epsilon_{ik} < \sigma_{ik}.$$

This situation is identical to that of part 1, with a time step of $m_{ik}\Delta s$, and the same result as in part 1 holds. The finite-difference problem to be solved has a time step of $m_{ik}\Delta s$, and so the algorithm in step (e) of Section 2 is used. In this situation there are $m_{ik} - 1$ time steps in which radiance does not propagate in the $\hat{n}_k$ direction from the point $\mathbf{x}_i$. In some geometries this has the appearance of a high-frequency oscillation. However, it should be emphasized that the oscillation is controlled by the uniform stability of this algorithm and can be reduced by performing calculations with finer grids.

5. EXAMPLE

The algorithms for TDRT, contained in steps (a)–(e) in Section 2, have been coded and executed in FORTRAN on a VAX 11/785 computer. The results of a calculation can be examined in several ways, some of which are provided here for a particular choice of initial conditions and optical properties.

The example under consideration is the propagation of a cylindrically symmetric pulse through 98 m of a medium with no absorption ($a = 0$), a scattering coefficient $b = 0.1$ m^{-1}, and an isotropic phase function

$$P_{\text{isotropic}}(\hat{n} \cdot \hat{n}') = \frac{1}{4\pi}.$$

Cylindrical symmetry is the property that the radiance distribution is invariant when the spatial and directional azimuthal angles are rotated rigidly together. If the propagation direction is given by the angle (θ, ϕ) in spherical coordinates and the position is given by (ρ, φ, z) in cylindrical coordinates, then the symmetry is

$$L(s, \rho, \varphi, z, \theta, \phi) = L(s, \rho, \varphi + \Delta, z, \theta, \phi + \Delta). \quad (31)$$

The FORTRAN code obtains the radiance in the plane $\varphi = 0$ at each time step. This requires knowledge of the radiance at nonzero spatial azimuths in each previous time step, which is obtained by using Eq. (31).

The phase matrix for the isotropic phase function is

$$P_{kk'} = \frac{\Delta\Omega_{k'}}{4\pi}.$$

The chosen partition has the 26 directions shown in Table 1, consisting of an equiangular division in θ and an equiangular division in ϕ for the bins away from $\theta = 0°$ and $\theta = 180°$. The spatial grid in the $\varphi = 0$ plane is rectangular, with the grid spacing $\Delta\rho = \Delta z = 2$ m, $N_\rho = 10$, $N_z = 50$, and the time step $\Delta s = 2$ m. The initial distribution of the pulse has energy in the forward direction ($k = 1$) bin only, located at three spatial points on the axis of propagation and the three points adjacent to those axis points. The evolving distribution was obtained for a total of 50 time steps, which required approximately 4 h of CPU time and 3/4 Mbyte memory on the VAX 11/785.

Figures 1 and 2 show the spatial distribution at several times in the forward ($k = 1$) and backward ($k = 26$) directions, respectively. The numbers next to the contours indicate the order of magnitude of the adjacent contour below the initial pulse strength (e.g., -1 is 10^{-1} below the original strength). The most prominent feature in these plots is the ray effect characterized by the lobes corresponding to radiance moving at 45 and 135° from the initial direction of propagation. This type of ray effect is present in any scheme in which the angular dimensions are replaced by a finite set of directions. The forward edge of the pulse at time $s = 20$ m has a ripple structure, owing to the course grid, which smooths in time as the distribution occupies more volume. The isotropic character of the phase function gives a high weight to multiple-scattering events, and the distribution behind the pulse quickly evolves into a smoothly stratified one in both of the figures.

If causal interpolation had not been used in the calculation, the edge of the distribution would outline a square with sides moving at the physical speed of light in the perpendicular directions, and the corners of the square would move faster than the physical speed.

Figures 3 and 4 show multiple time series in the forward and backward directions, respectively. Each time-series trace is the value of the radiance propagating in the appropriate direction at a designated point along the pulse axis.

One clear outcome of this calculation is that the magnitude of the forward-direction peak diminishes exponentially in time as

$$L(s) \sim \exp(-K_p s),$$

and in this calculation $K_p \sim 0.096\ \text{m}^{-1}$. Because there is no absorption and the pulse dimensions are small compared with the scattering length, K_p should have the approximate form

$$K_p = bg(\mathbf{P}),$$

where g is a dimensionless quantity that depends on the phase matrix. The resulting value $g \sim 0.96$ is approximately the same as $1 - P_{00}$. This allows for the interpretation that the peak of the distribution attenuates in time at the rate given by the scattering coefficient b and is partially restored by single scattering of the forward-directed pulse into the forward direction. Multiple-scattering events in the forward direction would contribute additional powers of $\mathbf{P}$ through

$$g(\mathbf{P}) = 1 - P_{00} + \sum_{n=2}^{\infty} g_n (\mathbf{P}^n)_{00} (bs)^{n-1},$$

and these appear to be small compared with single scattering, even over time scales equivalent to many scattering lengths.

In contrast, the backward-direction radiance shows attenuation that is due to single and multiple scattering. The peak backward radiance at each point attenuates at the same rate as the forward peak, indicating that the backward direction obtains energy directly from the forward direction. Over time, however, the attenuation rate at a given point diminishes, indicating that multiple-scattering effects contribute substantial amounts of energy to the backward direction. In fact, multiple scattering is important in all directions (including the forward direction) in regions away from the pulse.

The rough picture that develops from this simple analysis is that energy is removed from the propagating pulse by

Table 1. Angular Partition for the Example

Bin k	θ, ϕ	Maximum θ, ϕ	Minimum θ, ϕ
1	0, 0	22.5, 360	0, 0
2	45, 0	67.5, 22.5	22.5, −22.5
3	45, 45	67.5, 67.5	22.5, 22.5
4	45, 90	67.5, 112.5	22.5, 67.5
5	45, 135	67.5, 157.5	22.5, 112.5
6	45, 180	67.5, 202.5	22.5, 157.5
7	45, 225	67.5, 247.5	22.5, 202.5
8	45, 270	67.5, 292.5	22.5, 247.5
9	45, 315	67.5, 337.5	22.5, 292.5
10	90, 0	67.5, 22.5	67.5, −22.5
11	90, 45	112.5, 67.5	67.5, 22.5
12	90, 90	112.5, 112.5	67.5, 67.5
13	90, 135	112.5, 157.5	67.5, 112.5
14	90, 180	112.5, 202.5	67.5, 157.5
15	90, 225	112.5, 247.5	67.5, 202.5
16	90, 270	112.5, 292.5	67.5, 247.5
17	90, 315	112.5, 337.5	67.5, 292.5
18	135, 0	157.5, 22.5	112.5, −22.5
19	135, 45	157.5, 67.5	112.5, 22.5
20	135, 90	157.5, 112.5	112.5, 67.5
21	135, 135	157.5, 157.5	112.5, 112.5
22	135, 180	157.5, 202.5	112.5, 157.5
23	135, 225	157.5, 247.5	112.5, 202.5
24	135, 270	157.5, 292.5	112.5, 247.5
25	135, 315	157.5, 337.5	112.5, 292.5
26	180, 0	180, 360	157.5, 0

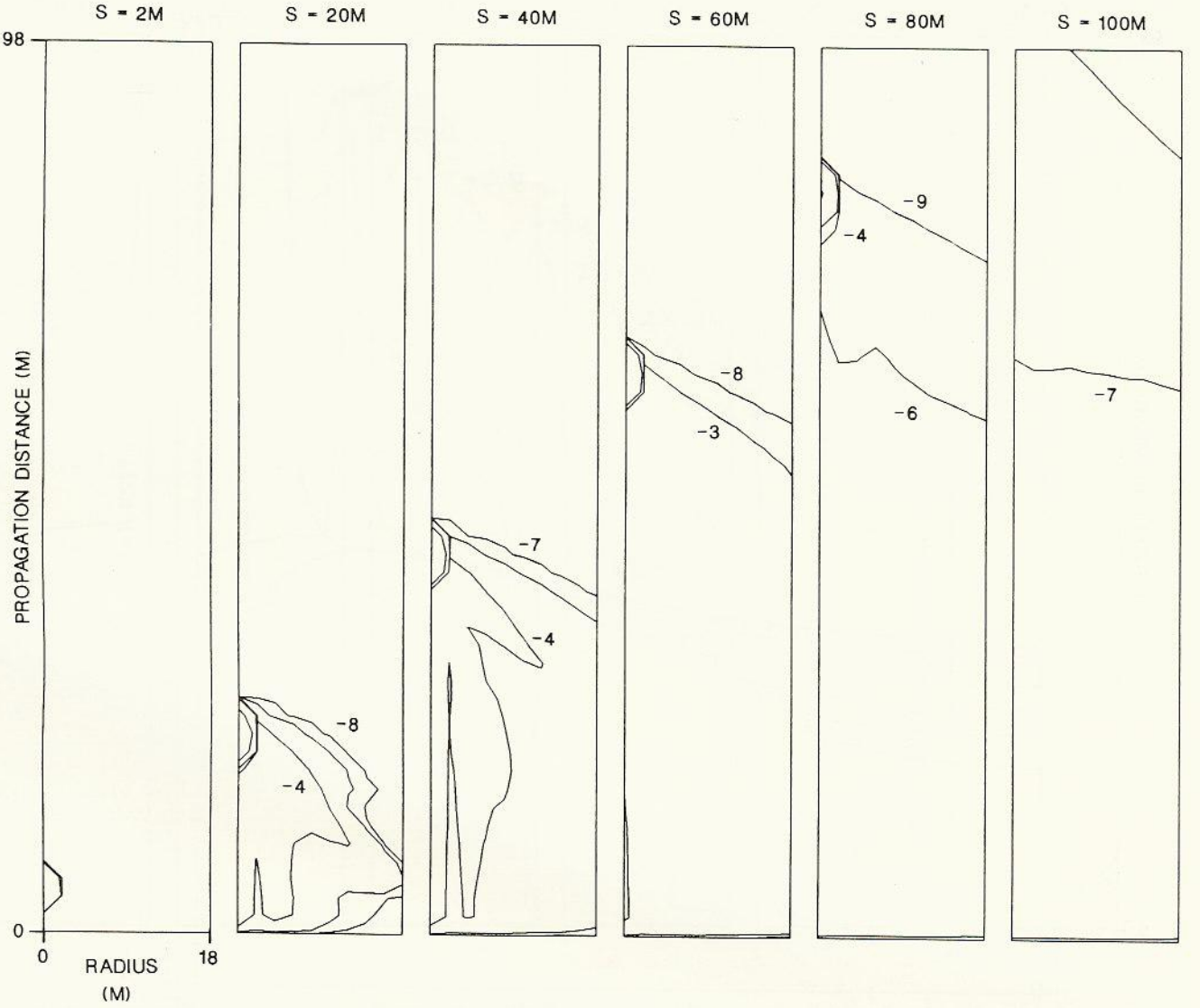

Fig. 1. Contour plots of the spatial distribution of the pulse in the forward ($k = 1$) direction at times $s = 2, 20, 40, 60, 80, 100$ m.

single-scattering events, with insignificant replenishment of the pulse by multiple scattering, and the removed energy is distributed in space and directions by multiple scattering. This description probably breaks down when the pulse has propagated for an amount of time that is sufficient to increase the relative magnitude of the multiple-scattered energy to a level comparable with the attenuated pulse. A rough estimate of the number of scattering lengths N_b required for multiple scattering to be seen in the pulse is

$$N_b \sim \frac{1 - P_{00}}{{P_{00}}^2},$$

i.e., N_b is the number of scattering events required to make double scattering comparable with single scattering in the attenuation factor g. For the example, $N_b \sim 25$, which is more than twice the number of events obtained in the calculation.

The example presented here is a calculation on a coarse set of angular and spatial grids, with an isotropic phase function. Realistic calculations in hydrologic or atmospheric optics would use phase functions that have a sharp forward peak. To reproduce this peak in the phase matrix it would be necessary to choose an angular partition that is correspondingly fine, potentially requiring prohibitive amounts of computational resources. However, experience with the existing code with coarse calculations indicates that such resources are available in modern work stations.

The CPU time and memory requirements for a given calculation are determined principally by the product of the number of space points times the number of partitioned directions. The total memory requirement (in bytes) is approximately

$$M = N_b N_g N_d N_f,$$

where N_b is the number of bytes per real number, N_g is the number of spatial grid points, N_d is the number of partitioned directions, and N_f is a scaling factor for the storage of more than one copy of the distribution and other variables. In practice, N_f is approximately 4. For example, if the unit sphere is partitioned into 10-deg intervals, the number of

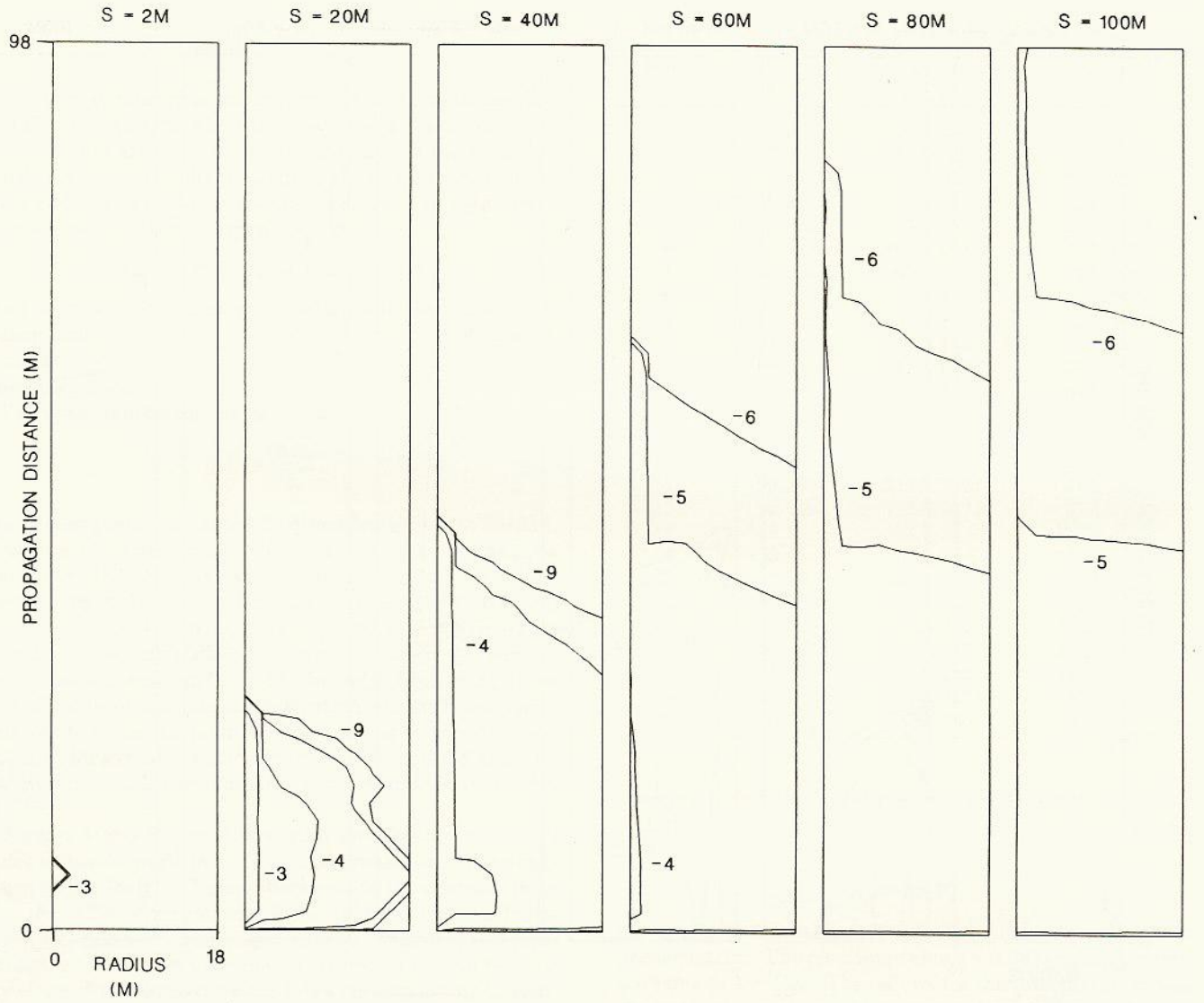

Fig. 2. Contour plots of the spatial distribution of the pulse in the backward ($k = 26$) direction at times $s = 2, 20, 40, 60, 80, 100$ m.

directions is 648. When 100 spatial grid points are used in a calculation and when we assume 16 bytes per number, the total memory requirement is approximately 4 Mbyte, which is readily available in work stations and supercomputers.

In principle, CPU time requirements also scale linearly with $N_g N_d$, but the method of estimating the CPU time needed for a single time step is complicated significantly by the particular characteristics of computer architecture and hardware. In almost all calculations that we have completed, the execution time exceeds the linear scaling relationship because of the virtual memory page swapping overhead on the VAX 11/785. A calculation in which $N_g = 110$ and $N_d = 26$ requires approximately 2–3 sec of VAX CPU time for each time step, whereas when $N_g = 1122$ and $N_d = 26$, 1 h per time step is required. There are several possible ways to improve the efficiency of the code beyond these times, however. The existing code spends 96% of the CPU time constructing the causal interpolation weights for each time step. This is because it treats the set of spatial grid points as an unordered set, so that the search for the interpolation points is conducted over the entire set of points. The execution time could be reduced substantially by ordering the grid points, at least partially, so that the search can be confined to a limited subset.

A second technique to reduce the CPU time is to use nonlinear, moving, or expanding spatial grids. If the radiance distribution is confined initially, the grid can also be confined initially and expanded periodically, either by adding more points or by spreading existing points farther apart so that CPU time is not used at locations for which the radiance is zero. The grid can also follow the motion of the core of a pulse so that the number of points remains fixed.

6. ADDITIONAL DEVELOPMENTS

Until now the discussion has focused on a finite-difference algorithm for the evolution of an initial radiance distribution in an unbounded medium. Generalizations can be made to include input of radiant energy over a period of time by a source located within the medium and the reflection of

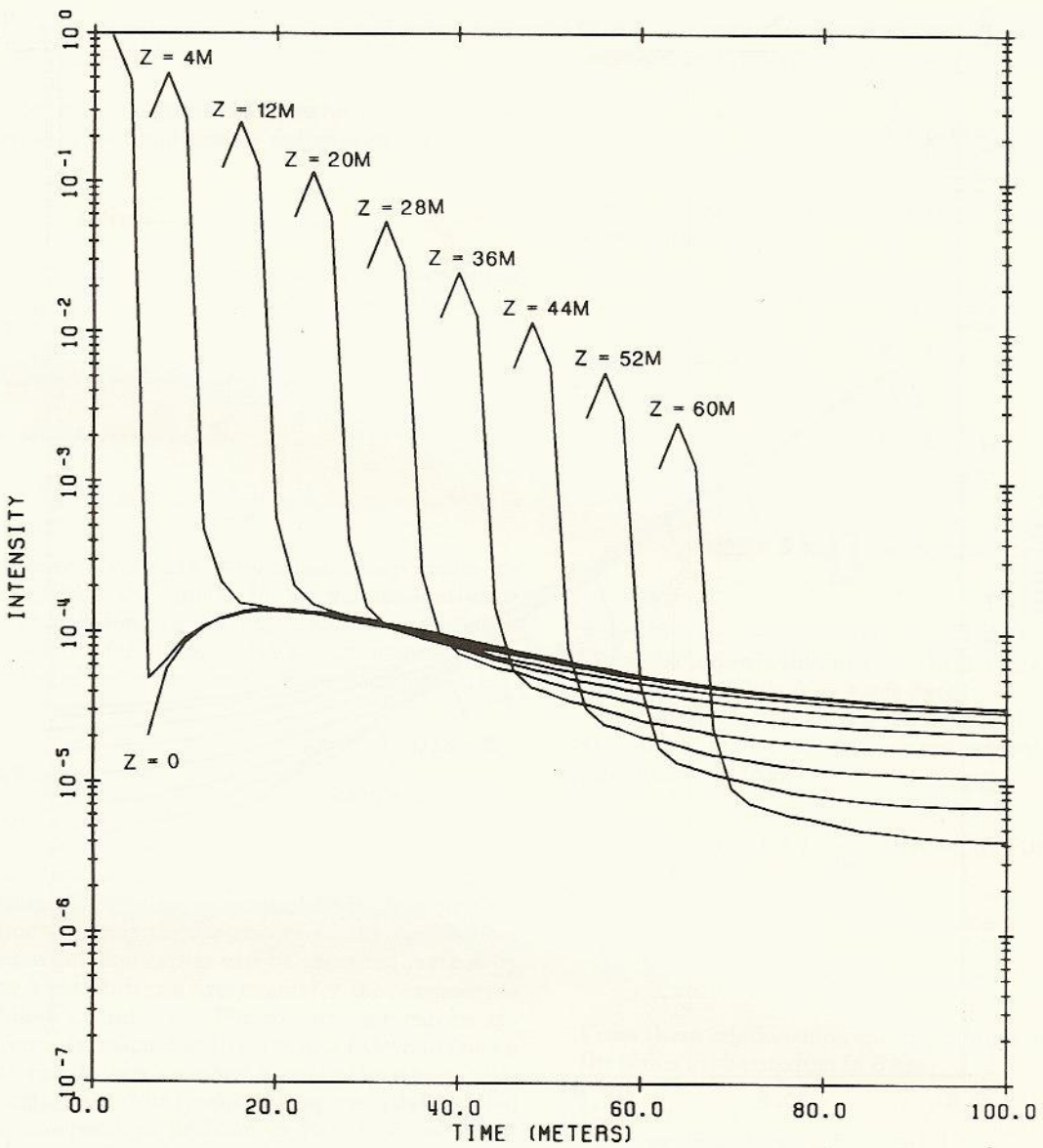

Fig. 3. Time series of the forward-direction radiance at several positions along the pulse axis.

radiance from an arbitrarily shaped surface. These two additions are discussed below.

A source of energy within the medium is described by the amount of radiant energy delivered into a volume ΔV in the solid angle $\Delta\Omega$ of directions during a time interval Δs. The source function is given by

$$\Delta(\text{energy}) = F(s, \mathbf{x}, \hat{n})\Delta s\Delta\Omega\Delta V/v$$

and appears in the TDRT equation as

$$\left[\frac{\partial}{\partial s} + \hat{n}\cdot\nabla + c(\mathbf{x})\right] L(s, \mathbf{x}, \hat{n}) = \int d^2n'\beta(\mathbf{x}, \hat{n}, \hat{n}')L(s, \mathbf{x}, \hat{n}') + F(s, \mathbf{x}, \hat{n}).$$

The corresponding operator equation is

$$\left(\frac{\partial}{\partial s} + \mathcal{H}\right)\mathcal{L}(s) = \mathcal{F}(s)$$

and has the solution

$$\mathcal{L}(s) = \mathcal{G}(s)\mathcal{L}(0) + \int_0^s ds'\mathcal{G}(s - s')\mathcal{F}(s').$$

To construct a finite-difference algorithm with a source term, the operator solution is written as

$$\mathcal{L}(s + \Delta s) = \mathcal{G}(\Delta s)\mathcal{L}(s) + \int_s^{s+\Delta s} ds'\mathcal{G}(s + \Delta s - s')\mathcal{F}(s').$$

In addition, the condition $b\Delta s \ll 1$ allows for the simplification of the evolution operator in the source term to

$$\mathcal{G}(s + \Delta s - s') \rightarrow \exp[-a(s + \Delta s - s')].$$

Assuming that the source does not vary rapidly during a single time step, we find the operator algorithm to be

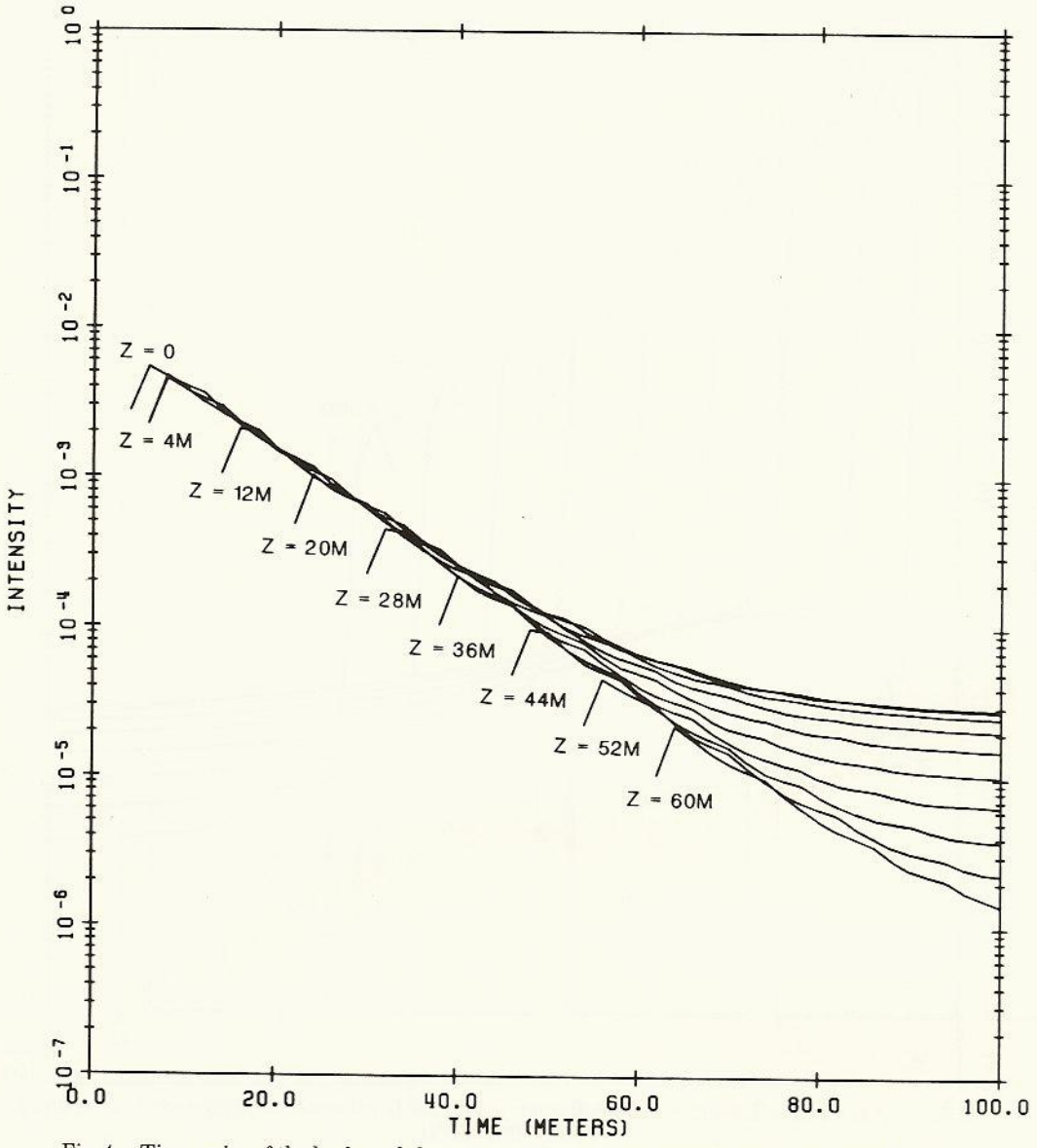

Fig. 4. Time series of the backward-direction radiance at several positions along the pulse axis.

$$\mathcal{L}(s + \Delta s) = \mathcal{G}(\Delta s)\mathcal{L}(s) + \frac{1}{a}[\exp(-a\Delta s/2) - \exp(-a\Delta s)]\mathcal{J}(s) + \frac{1}{a}[1 - \exp(-a\Delta s/2)]\mathcal{J}(s + \Delta s).$$

If we execute the partitioning and discretization as above, the complete algorithm is

$$L_k(s + \Delta s, \mathbf{x}) = \exp(-c\Delta s)\sum_{k'=1}^{N} T_{kk'}(\Delta s)L_{k'}(s, \mathbf{x} - \hat{n}_{k'}\Delta s) + \frac{1}{a}[\exp(-a\Delta s/2) - \exp(-a\Delta s)]F_k(s, \mathbf{x}) + \frac{1}{a}[1 - \exp(-a\Delta s/2)]F_k(s + \Delta s, \mathbf{x}).$$

An algorithm for time-dependent reflection of radiance from a surface can be constructed by appropriate interpretation of the quantities in the definition of the bidirectional reflectivity distribution function. Suppose that a surface S within the medium has a normal $\hat{n}_S$ at the point $\mathbf{x}_S$. Incident radiance $L^i(\mathbf{x}_S, \hat{n})$ is reflected at the surface in many directions to produce the reflected distribution $L^r(\mathbf{x}_S, \hat{n})$. The connection between the two can be written as

$$L^r(\mathbf{x}_S, \hat{n}) = \int \mathrm{d}^2n' R(\hat{n}, \hat{n}', \mathbf{x}_S)(-\hat{n}_S \cdot \hat{n}')L^i(\mathbf{x}_S, \hat{n}'),$$

where R is the bidirectional reflectivity distribution function.[7] In partitioned form this equation is

$$L_k{}^r(\mathbf{x}_S) = \sum_{k'} R_{kk'}(\mathbf{x}_S)L_{k'}{}^i(\mathbf{x}_S),$$

where

$$R_{kk'}(\mathbf{x}_S) = \frac{1}{\Delta\Omega_k}\int_k \mathrm{d}^2n \int_{k'} \mathrm{d}^2n' R(\hat{n}, \hat{n}', \mathbf{x}_S)(-\hat{n}_S \cdot \hat{n}').$$

The time-dependent algorithm follows by interpreting the reflected radiance as the observed radiance at the surface point,

$$L_k{}^r(\mathbf{x}_S) = L_k(s + \Delta s, \mathbf{x}_S),$$

and by interpreting the incident radiance as the distribution near the surface in the previous time step, propagated to the surface and reduced in magnitude by absorption,

$$L_k{}^i(\mathbf{x}_S) = \exp(-a\Delta s)L_k(s, \mathbf{x}_S - \hat{n}_k\Delta s).$$

In this case the full algorithm is

$$L_k(s + \Delta s, \mathbf{x}_S) = \exp(-a\Delta s)\sum_{k'} R_{kk'}L_{k'}(s, \mathbf{x}_S - \hat{n}_{k'}\Delta s).$$

Modifications and restrictions for causal interpolation apply to this algorithm the same as for the volume-scattering algorithm. In addition, a numerical execution of this algorithm must account for hiding by the surface and reflection of some portion of the causal interpolation propagation distance l_{ik}.

APPENDIX A

The development of TDRT in the Dirac notation in Sections 1 and 2 highlights the similarities in the formal mathematical structure between quantum mechanics and radiative transfer, although the extent of similarities in their physical interpretation depends on the nature of the application. The mathematical similarities can be extended further by constructing a path integral expression for the components of the evolution operator $\mathcal{G}$. The construction can be approached from a functional analysis point of view, or from a more traditional piecewise method by using the closure relation of Hilbert space. The functional approach was applied to the time-independent problem in Ref. 8, so the more traditional approach[9] is taken here to complement that derivation.

Although the path integral expression is a formal solution, it offers the hope of leading to significantly new and varied approximate solutions of the TDRT equation. As an example, a new small-angle approximation for the time-independent problem was extracted from its path integral solution and compared with data on the in-water solar distribution with some success.[10] Path integral methods have been powerful tools in the fields of high-energy physics,[11] solid-state and condensed-matter physics,[12] and even fluid dynamics.[13] General schemes have been established for approximating path integral expressions.[14] An outline is given at the end of this appendix as to how a nonperturbative approximate evaluation might be obtained for the TDRT path integral expression.

The construction of the path integral expression begins with the definition of the components of the evolution operator

$$G(s, \mathbf{x}, \hat{n}; \mathbf{x}', \hat{n}') = \langle \mathbf{x}, \hat{n}|\exp(-s\mathcal{H})|\mathbf{x}', \hat{n}'\rangle.$$

The evolution operator can be broken into a product of N short-time evolution operators, as in

$$\exp(-s\mathcal{H}) = \prod_{i=1}^{N} \exp(-\Delta s\mathcal{H}),$$

with $s = N\Delta s$. Between each term in the product, the closure relation

$$\int \mathrm{d}^3x\mathrm{d}^2n|\mathbf{x}, \hat{n}\rangle\langle\mathbf{x}, \hat{n}| = 1$$

can be inserted to give

$$G(s, \mathbf{x}, \hat{n}; \mathbf{x}', \hat{n}') = \int\left(\prod_{i=1}^{N-1} \mathrm{d}^3x_i\mathrm{d}^2n_i\right) \times \prod_{i=1}^{N} \langle \mathbf{x}_i, \hat{n}_i|\exp(-\Delta s\mathcal{H})|\mathbf{x}_{i-1}, \hat{n}_{i-1}\rangle,$$

where $|\mathbf{x}_N, \hat{n}_N\rangle = |\mathbf{x}, \hat{n}\rangle$ and $|\mathbf{x}_0, \hat{n}_0\rangle = |\mathbf{x}', \hat{n}'\rangle$. The path integral solution is obtained by considering the limiting form when $N \to \infty$, while $N\Delta s = s$ is fixed.

To proceed further, it is useful to introduce the variable $\mathbf{p}$ as a conjugate to $\hat{n}$, as there is a representation of the phase function in the form

$$P(\hat{n}, \hat{n}') = \int \frac{\mathrm{d}^3p}{(2\pi)^3}\Pi(\mathbf{p})\exp[i\mathbf{p}\cdot(\hat{n} - \hat{n}')]$$

and there is an inner-product relation

$$\langle\hat{n}|\mathbf{p}\rangle = \exp(i\mathbf{p}\cdot\hat{n}).$$

From these relationships and the closure of the $\{|\mathbf{p}\rangle\}$ basis, the terms in the product in G are

$$\langle \mathbf{x}_i, \hat{n}_i|\exp(-\Delta s\mathcal{H})|\mathbf{x}_{i-1}, \hat{n}_{i-1}\rangle = \int \mathrm{d}^3y \frac{\mathrm{d}^3p_i}{(2\pi)^3}\langle \mathbf{x}_i, \hat{n}_i|\mathbf{y}, \mathbf{p}_i\rangle \times \langle \mathbf{y}, \mathbf{p}_i|\exp(-\Delta s\mathcal{H})|\mathbf{x}_{i-1}, \hat{n}_{i-1}\rangle.$$

The small time-step argument used to obtain expression (20) can be used here to give

$$\begin{aligned}\langle \mathbf{y}, \mathbf{p}_i|\exp(-\Delta s\mathcal{H})|\mathbf{x}_{i-1}, \hat{n}_{i-1}\rangle &= \langle \mathbf{y}, \mathbf{p}_i|\exp(-c\Delta s)\exp(b\Delta s\mathcal{P}) \\ &\quad\times|\mathbf{x}_{i-1} + \hat{n}_{i-1}\Delta s, \hat{n}_{i-1}\rangle \\ &= \delta(\mathbf{y} - \mathbf{x}_{i-1} - \hat{n}_{i-1}\Delta s) \\ &\quad\times\exp(-i\mathbf{p}_i\cdot\hat{n}_{i-1})\exp(-c\Delta s) \\ &\quad\times\exp[b\Delta s\Pi(\mathbf{p}_i)],\end{aligned}$$

whereas the conjugation property gives

$$\langle \mathbf{x}_i, \hat{n}_i|\mathbf{y}, \mathbf{p}_i\rangle = \delta(\mathbf{x}_i - \mathbf{y})\exp(i\mathbf{p}_i\cdot\hat{n}_i).$$

When the terms of the product are assembled,

$$G(s,\mathbf{x},\hat{n};x',\hat{n}') = \exp(-cs)\int\left(\prod_{i=1}^{N-1} d^2n_i\frac{d^3p_i}{(2\pi)^3}\right)$$
$$\times\delta\left(\mathbf{x}-\mathbf{x}'-\sum_{i=0}^{N}\Delta s\hat{n}_i\right)$$
$$\times\exp\left[\sum_{i=1}^{N} b\Delta s\Pi(\mathbf{p}_i)\right]$$
$$\times\exp\left[i\sum_{i=1}^{N}\Delta s\mathbf{p}_i\cdot\left(\frac{\hat{n}_i-\hat{n}_{i-1}}{\Delta s}\right)\right].$$

In the $N\to\infty$ limit the sums over i in the exponents become integrals over time, the discrete set $\{\hat{n}_i\}$ becomes the continuous function $\hat{\beta}(s')$, the measure becomes

$$\prod_{i=1}^{N-1} d^2n_i\frac{d^3p_i}{(2\pi)^3}\to[D\beta]_{4\pi}[Dp]\delta[\hat{\beta}(0)-\hat{n}']\delta[\hat{\beta}(s)-\hat{n}],$$

where the 4π subscript means integration over the unit sphere, and

$$\frac{\hat{n}_i-\hat{n}_{i-1}}{\Delta s}\to\frac{\partial\hat{\beta}(s')}{\partial s'}.$$

The resulting path integral expression for the evolution operator is

$$G(s,\mathbf{x},\hat{n};\mathbf{x}',\hat{n}') = \exp(-cs)\int[D\beta]_{4\pi}[Dp]\delta[\hat{\beta}(0)-\hat{n}']$$
$$\times\delta[\hat{\beta}(s)-\hat{n}]\delta[\mathbf{x}-\mathbf{x}'-\int_0^s ds'\hat{\beta}(s')]$$
$$\times\exp\left\{b\int_0^s ds'\Pi[\mathbf{p}(s')]\right\}$$
$$\times\exp\left[i\int_0^s ds'\mathbf{p}(s')\cdot\frac{\partial\hat{\beta}(s')}{\partial s'}\right]. \qquad (A1)$$

This expression involves an integral $d^2\beta(s')$ at each time over the unit sphere and so is a path integral over a compact space. Alternatively, it can be viewed as an integral over all the three-dimensional space in which the unit sphere lies, with a constraint on the magnitude of $\hat{\beta}(s')$ at each time. Path integration of constrained systems is discussed in many texts, with some of the original work available in Ref. 15. The net result in this case is that the integration measure $[D\beta]_{4\pi}$ becomes

$$[D\beta]_{4\pi} = [D\beta]\prod_{s'=0}^{s}\delta[\beta^2(s')-1],$$

and the measure $[D\beta]$ is over the three-dimensional space. When we use the functional delta-function identity

$$\prod_{s'=0}^{s}\delta[\beta^2(s')-1] = \int[D\phi]\exp\left\{i\int_0^s ds'\phi(s')[1-\beta^2(s')]\right\},$$

the expression for the evolution operator becomes

$$G(s,\mathbf{x},\hat{n};\mathbf{x}',\hat{n}') = \exp(-cs)\int[D\beta][Dp][D\phi]\delta[\beta(0)-\hat{n}']$$
$$\times\delta[\beta(s)-\hat{n}]\delta\left[\mathbf{x}-\mathbf{x}'-\int_0^s ds'\beta(s')\right]$$
$$\times\exp\left\{i\int_0^s ds'\phi(s')[1-\beta^2(s')]\right\}$$
$$\times\exp\left\{b\int_0^s ds'\Pi[\mathbf{p}(s')]\right\}$$
$$\times\exp\left[i\int_0^s ds'\mathbf{p}(s')\cdot\frac{\partial\beta(s')}{\partial s'}\right], \qquad (A2)$$

and now all the integration measures are over unbounded spaces.

Equations (A1) and (A2) are equivalent path integral solutions of the TDRT equation. Equation (A2) is frequently more useful, however, because path integration is generally easier to evaluate over an unbounded space than over a bounded space. Although this particular expression cannot be evaluated in closed form exactly, it is possible to construct an approximate evaluation.

When the phase function is strongly forward peaked (as, for example, in ocean water) it might be possible to characterize it simply by the mean square of the scattering angle $\mu \ll 1$ and to approximate it by

$$\Pi(\mathbf{p})\approx 1-\frac{\mu}{2}p^2.$$

In this situation the integration over $\mathbf{p}(s')$ is Gaussian and can be evaluated to give

$$\int[Dp]\exp\left[-\frac{b\mu}{2}\int_0^s ds'p^2(s')\right]\exp\left\{i\int_0^s ds'\mathbf{p}(s')\cdot\frac{\partial\beta(s')}{\partial s'}\right\}$$
$$=\exp\left\{-\frac{1}{2b\mu}\int_0^s ds'\left[\frac{\partial\beta(s')}{\partial s'}\right]^2\right\}.$$

When the integration function $\phi(s')$ is redefined as $\lambda(s')/2b\mu$ and when the $\mathbf{x}-\mathbf{x}'$ variable is Fourier transformed, the expression for the evolution operator becomes

$$\tilde{G}(s,\mathbf{q},\hat{n},\hat{n}') = \exp(-as)\int[D\beta][D\lambda]\delta[\beta(0)-\hat{n}']\delta[\beta(s)-\hat{n}]$$
$$\times\exp\left[\frac{i}{2b\mu}\int_0^s ds'\lambda(s')\right]\exp\left[-i\int_0^s ds'\mathbf{q}\cdot\beta(s')\right]$$
$$\times\exp\left(-\frac{1}{2b\mu}\int_0^s ds'\left\{\left[\frac{\partial\beta(s')}{\partial s'}\right]^2+i\lambda(s')\beta^2(s')\right\}\right).$$

The Gaussian integral over β can be evaluated exactly by translating β to the integration variable η by

$$\beta(s') = \mathbf{W}(s')+\eta(s'),$$

where $\mathbf{W}(s')$ is a functional of λ and has the decomposition

$$\mathbf{W}(s') = \hat{n}f(s')+\hat{n}'f(s-s')-ib\mu\mathbf{q}g(s'),$$
$$g(s') = \int_0^s dtD(s',t).$$

The quantities D and f satisfy the equations

$$\left[-\left(\frac{\partial}{\partial s'}\right)^2 + i\lambda(s')\right] f(s') = 0,$$

$$\left[-\left(\frac{\partial}{\partial s'}\right)^2 + i\lambda(s')\right] D(s', s'') = \delta(s' - s''), \tag{A3}$$

with boundary conditions

$$f(0) = 0,$$
$$f(s) = 1,$$
$$D(0, s'') = 0,$$
$$D(s, s'') = 0.$$

Note that an exact solution for D in terms of f is

$$D(s', s'') = \frac{f(s_<)f(s - s_>)}{f'(s - s'')f(s'') + f'(s'')f(s - s'')}, \tag{A4}$$

where $s_> = \max\{s', s''\}$ and $s_< = \min\{s', s''\}$.

After translation, the path integral to be evaluated is

$$\int [D\eta]\delta[\eta(0)]\delta[\eta(s)] \times \exp\left(-\frac{1}{2b\mu}\int_0^s ds' \left\{\left[\frac{\partial\eta(s')}{\partial s'}\right]^2 + i\lambda(s')\eta^2(s')\right\}\right)$$

and has the solution

$$\exp\left[\frac{3}{2}\,\mathrm{Tr}\log(D)\right]. \tag{A5}$$

The path integral over $\mathbf{p}$, β, and λ has been reduced to a path integral over just λ and is

$$\tilde{G}(s, \mathbf{q}, \hat{n}, \hat{n}') = \int [D\lambda]\exp[\Gamma(s, \mathbf{q}, \hat{n}, \hat{n}', \lambda)],$$

where

$$\Gamma(s, \mathbf{q}, \hat{n}, \hat{n}', \lambda) = -as + \frac{3}{2}\,\mathrm{Tr}\log(D) + \frac{i}{2b\mu}\int_0^s ds'\lambda(s') - i\int_0^s ds'\mathbf{q}\cdot\mathbf{W}(s') - \frac{1}{2b\mu}\int_0^s ds' \times \left\{\left[\frac{\partial\mathbf{W}(s')}{\partial s'}\right]^2 + i\lambda(s')W^2(s')\right\}.$$

This exact path integral expression cannot be evaluated analytically. However, one of its potentially useful features is that there is no explicit small-angle restriction—and probably no implicit small-angle restriction. An approximate evaluation of the integral should produce an expression that is equally valid for all angles $\hat{n}$ and $\hat{n}'$.

The steps to an approximate evaluation begin[14] with the definition of a function $\lambda_c(s')$, which characterizes the rate at which $\mathbf{W}$ converts from the initial direction of propagation $\hat{n}'$ to the final direction $\hat{n}$. This function is defined by the condition

$$\frac{\delta\Gamma}{\delta\lambda(t)}(\lambda = \lambda_c) = 0.$$

When the definitions of Γ, $\mathbf{W}$, f, and D are used, this condition is the implicit functional equation

$$3b\mu D_c(t, t) = 1 - W_c^2(t), \tag{A6}$$

and the c subscript implies that $\lambda = \lambda_c$ in this expression. By expanding Γ in a functional Taylor series around the function λ_c,

$$\Gamma(\lambda_c + \sigma) \approx \Gamma(\lambda_c) + \frac{1}{2}\int_0^s dt \int_0^s dt'\sigma(t)\frac{\delta^2\Gamma(\lambda_c)}{\delta\lambda(t)\delta\lambda(t')}\sigma(t').$$

When only this quadratic expansion is kept, the leading-order approximation is

$$\tilde{G}(s, \mathbf{q}, \hat{n}, \hat{n}') = \exp(\Gamma_c)\exp\left(-\frac{1}{2}\,\mathrm{Tr}\log\left|-\frac{\delta^2\Gamma_c}{\delta\lambda\delta\lambda}\right|\right), \tag{A7}$$

and

$$\frac{\delta^2\Gamma_c}{\delta\lambda(t)\delta\lambda(t')} = -\frac{3}{2}D_c(t, t')D_c(t', t) - \frac{1}{b\mu}D_c(t, t')\mathbf{W}_c(t)\cdot\mathbf{W}_c(t').$$

What remains to be done is to solve Eqs. (A3) and (A6) by using Eq. (A4) and to substitute the solutions into Eq. (A7).

ACKNOWLEDGMENTS

I would like to thank Robert Kelly, Samuel Gasster, and William Helliwell for many interesting discussions on optics and radiative transfer. Special thanks go to Cynthia Piotrowski for coding and executing the algorithms described in Sections 2 and 6 and to William Helliwell for his kind reading and criticism of parts of the manuscript. The preparation of the manuscript and calculation of the example in Section 5 were supported by Areté Associates.

* Present address, Department of Meteorology, Texas A&M University, College Station, Texas 77843.

REFERENCES

1. R. W. Preisendorfer, *Radiative Transfer on Discrete Spaces* (Pergamon, New York, 1965).
2. A. Messiah, *Quantum Mechanics* (Wiley, New York, 1961), Vol. I.
3. R. M. Lerner and J. D. Summers, "Monte Carlo description of time- and space-resolved multiple forward scatter in natural water," Appl. Opt. **21**, 861–869 (1982); G. I. Marchuk, G. A. Mikhailov, M. A. Nazaraliev, R. A. Darbinjan, B. A. Kargin, and B. S. Elepov, *The Monte Carlo Methods in Atmospheric Optics* (Springer-Verlag, Berlin, Heidelberg, 1980); G. N. Plass and G. W. Kattawar, "Monte Carlo calculations of radiative transfer in the Earth's atmosphere-ocean system: I. Flux in the atmosphere and ocean," J. Phys. Oceanogr. **2**, 139–145 (1972).
4. W. R. Martin, "The application of the finite element method to the neutron transport equation," Ph.D. dissertation (University of Michigan, Ann Arbor, Mich., 1976).
5. R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems* (Wiley, New York, 1967).
6. E. Isaacson and H. B. Keller, *Analysis of Numerical Methods* (Wiley, New York, 1966).
7. E. M. Sparrow and R. D. Cess, *Radiation Heat Transfer* (McGraw-Hill, New York, 1978).
8. J. Tessendorf, "Radiative transfer as a sum over paths," Phys. Rev. A **35**, 872–878 (1987).
9. R. P. Feynmann and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965); F. Langouche, D. Roekaerts, and E. Tirapequi, *Functional Integration and Semiclassical Expansions* (Reidel, Boston, Mass., 1982).
10. J. Tessendorf, "Comparison between data and small-angle approximations for the in-water solar radiance distribution," J. Opt. Soc. Am. A **5**, 1410–1418 (1988).
11. L. D. Faddeev and A. A. Slavnov, *Gauge Fields, An Introduc-*

tion to Quantum Theory (Benjamin/Cummings, Reading, Mass., 1980); H. M. Fried, *Functional Methods and Models in Quantum Field Theory*, special reprint ed. (MIT Press, Cambridge, Mass., 1982).
12. D. J. Amit, *Field Theory, the Renormalization Group, and Critical Phenomena* (McGraw-Hill, New York, 1978).
13. H. M. Fried and J. Tessendorf, "Green's functions at zero viscosity," J. Math. Phys. **25,** 1144–1154 (1984); J. Tessendorf, "Application of a mean-field expansion to stochastically stirred fluids," Ph.D. dissertation (Brown University, Providence, R.I., 1984).
14. C. M. Bender, F. Cooper, and G. S. Guralnik, "Path integral formulation of mean-field perturbation theory," Ann. Phys. **109,** 165–209 (1977); J. Iliopoulos, C. Itzykson, and A. Martin, "Functional methods and perturbation theory," Rev. Mod. Phys. **47,** 165–192 (1975).
15. L. D. Faddeev and V. N. Popov, "Feynman diagrams for the Yang–Mills field," Phys. Lett. B **25,** 29–30 (1967).