Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media

Knut Stamnes, S-Chee Tsay, Warren Wiscombe, and Kolf Jayaweera

We summarize an advanced, thoroughly documented, and quite general purpose discrete ordinate algorithm for time-independent transfer calculations in vertically inhomogeneous, nonisothermal, plane-parallel media. Atmospheric applications ranging from the UV to the radar region of the electromagnetic spectrum are possible. The physical processes included are thermal emission, scattering, absorption, and bidirectional reflection and emission at the lower boundary. The medium may be forced at the top boundary by parallel or diffuse radiation and by internal and boundary thermal sources as well. We provide a brief account of the theoretical basis as well as a discussion of the numerical implementation of the theory. The recent advances made by ourselves and our collaborators—advances in both formulation and numerical solution—are all incorporated in the algorithm. Prominent among these advances are the complete conquest of two ill-conditioning problems which afflicted all previous discrete ordinate implementations: (1) the computation of eigenvalues and eigenvectors and (2) the inversion of the matrix determining the constants of integration. Copies of the FORTRAN program on microcomputer diskettes are available for interested users.

I. Introduction

The discrete ordinate method for radiative transfer is commonly ascribed to Chandrasekhar.[1] Computer implementations of that method were, however, plagued by numerical difficulties[2] to such an extent that researchers made little use of it. The purpose of this paper is to alert the community to a new numerical implementation of the discrete ordinate method for vertically inhomogeneous layered media which is free of these difficulties and to give a summary of its equations and its various advanced features. The resulting computer code represents the culmination of years of effort[3-9] on the part of ourselves and our collaborators to make it the finest algorithm available. Our intent is that the code be so well documented, so versatile, and so error-free that other researchers can easily and safely use it both in data analysis and as a component of large models.

Knut Stamnes is with University of Tromsø, Institute of Mathematical & Physical Sciences, Auroral Observatory, N-9001 Tromsø, Norway; W. Wiscombe is with NASA Goddard Space Flight Center, Greenbelt, Maryland 20771; the other authors are with University of Alaska—Fairbanks, Physics Department & Geophysical Institute, Fairbanks, Alaska 99775-0800.

Received 7 August 1987.

The problem to be solved is the transfer of monochromatic radiation in a scattering, absorbing, and emitting plane-parallel medium with a specified bidirectional reflectivity at the lower boundary. Section II summarizes the equations and boundary conditions. Section III discusses the numerical implementation of the theory, in particular, (a) how to compute eigenvalues and eigenvectors reliably and efficiently and (b) how to avoid fatal overflows and ill-conditioning in the matrix inversion needed to determine the constants of integration. Reference 9 provides a more detailed account as well as documentation, test problems, and a listing of the FORTRAN code.

II. Theory

A. Basic Equations

The purpose of this section is to present the basic radiative transfer formulas with a minimum of definition and explanation to establish notation and conventions. (For more comprehensive discussions of the subject, the reader is referred to textbooks[1,10] and review articles.[11-15])

The equation describing the transfer of monochromatic radiation at frequency ν through a plane-parallel medium is given by[1]

$$\mu \frac{du_\nu(\tau_\nu,\mu,\phi)}{d\tau} = u_\nu(\tau_\nu,\mu,\phi) - S_\nu(\tau_\nu,\mu,\phi), \quad (1)$$

where $u_\nu(\tau_\nu,\mu,\phi)$ is the specific intensity along direction μ,ϕ at optical depth τ_ν measured perpendicular to the surface of the medium (ϕ is the azimuthal angle, and μ is the cosine of the polar angle). S_ν is the source function given by

$$S_\nu(\tau_\nu,\mu,\phi) = \frac{\omega_\nu(\tau_\nu)}{4\pi}\int_0^{2\pi} d\phi' \int_{-1}^{1} d\mu' P_\nu(\tau_\nu,\mu,\phi;\mu',\phi') \times u_\nu(\tau_\nu,\mu',\phi') + Q_\nu(\tau_\nu,\mu,\phi), \quad (2)$$

where $\omega_\nu(\tau_\nu)$ is the single-scattering albedo, and $P_\nu(\tau_\nu,\mu,\phi;\mu',\phi')$ is the phase function. For thermal emission in local thermodynamic equilibrium (LTE), the source term Q_ν is

$$Q_\nu^{(\text{thermal})}(\tau_\nu) = [1-\omega_\nu(\tau_\nu)]B_\nu[T(\tau_\nu)], \quad (3)$$

where $B_\nu(T)$ is the Planck function at frequency ν and temperature T. If the usual diffuse–direct distinction is made (Ref. 1, p. 22), so that u_ν in Eqs. (1) and (2) describes the diffuse radiation only, then for a parallel beam incident in direction μ_0,ϕ_0 on a nonemitting medium,

$$Q_\nu^{(\text{beam})}(\tau_\nu,\mu,\phi) = \frac{\omega_\nu(\tau_\nu)I_0}{4\pi} P_\nu(\tau_\nu,\mu,\phi;-\mu_0,\phi_0)\exp(-\tau_\nu/\mu_0), \quad (4)$$

where $\mu_0 I_0$ is the incident flux. We take

$$Q_\nu(\tau_\nu,\mu,\phi) = Q_\nu^{(\text{thermal})}(\tau_\nu) + Q_\nu^{(\text{beam})}(\tau_\nu,\mu,\phi).$$

(Many radiative transfer models consider one or the other source term but not both.)

To simplify the notation we omit the ν subscript in the following. By expanding the phase function $P(\tau, \cos\theta)$ in a series of $2N$ Legendre polynomials and the intensity in a Fourier cosine series,[1,9]

$$u(\tau,\mu,\phi) = \sum_{m=0}^{2N-1} u^m(\tau,\mu)\cos m(\phi_0-\phi), \quad (5)$$

we find that Eqs. (1) and (2) are replaced by $2N$ independent equations (one for each Fourier component)

$$\mu\frac{du^m(\tau,\mu)}{d\tau} = u^m(\tau,\mu) - \int_{-1}^{1} D^m(\tau,\mu,\mu')u^m(\tau,\mu')d\mu' - Q^m(\tau,\mu) \quad (6a)$$

$$(m = 0,1,2,\ldots,2N-1),$$

where

$$D^m(\tau,\mu,\mu') = \frac{\omega(\tau)}{2}\sum_{l=m}^{2N-1}(2l+1)g_l^m(\tau)P_l^m(\mu)P_l^m(\mu') \quad (6b)$$

$$Q^m(\tau,\mu) = X_0^m(\tau,\mu)\exp(-\tau/\mu_0) + \delta_{m0}Q^{(\text{thermal})}(\tau), \quad (6c)$$

$$X_0^m(\tau,\mu) = \frac{\omega(\tau)I_0}{4\pi}(2-\delta_{m0})\sum_{l=0}^{2N-1}(-1)^{l+m}(2l+1) \times g_l^m(\tau)P_l^m(\mu)P_l^m(\mu_0), \quad (6d)$$

$$\delta_{m0} = 1 \text{ if } m = 0 \text{ (0 otherwise)}, g_l^m(\tau) = g_l(\tau)\frac{(l-m)!}{(l+m)!},$$

$$g_l(\tau) = \frac{1}{2}\int_{-1}^{1} P_l(\cos\theta)P(\tau,\cos\theta)d\cos\theta.$$

$P_l(\cos\theta)$ is the Legendre polynomial, $P_l^m(\mu)$ is the associated Legendre polynomial, and θ is the angle between the direction vectors before and after scattering. Solutions to Eq. (6) give the azimuthal components, and then Eq. (5) gives the complete azimuthal dependence of the intensity. [When there is no beam source, the sum in Eq. (5) reduces to the $m = 0$ term, and μ_0, ϕ_0 are irrelevant.]

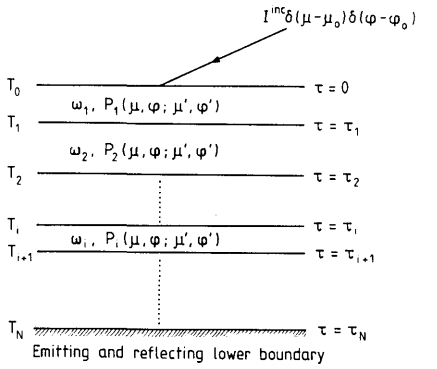

Fig. 1. Schematic illustration of a multilayered medium.

B. Discrete Ordinate Approximation—Matrix Formulation

The discrete ordinate approximation to (6) can be written[3,4]

$$\mu_i\frac{du^m(\tau,\mu_i)}{d\tau} = u^m(\tau,\mu_i) - \sum_{\substack{j=-N\\ j\neq 0}}^{N} w_j D^m(\tau,\mu_i,\mu_j) \times u^m(\tau,\mu_j) - Q^m(\tau,\mu_i) \qquad (i = \pm 1,\ldots\pm N), \quad (7a)$$

where μ_i and w_i are quadrature points and weights.

Since the single-scattering albedo $\omega(\tau)$ and the phase function $P(\tau,\mu,\phi;\mu',\phi')$ are functions of τ in a vertically inhomogeneous medium, Eqs. (7) constitute (for each m) a system of $2N$ coupled differential equations with nonconstant coefficients for which analytic solutions do not exist. To obtain analytic solutions, the medium is assumed to consist of L adjacent homogeneous layers in which the single-scattering albedo and the phase function are taken to be constant within each layer (but allowed to vary from layer to layer, see Fig. 1), and the thermal source term is approximated by a polynomial in τ. For the time being, we shall consider a single homogeneous layer for which $\tau_{p-1} \leq \tau \leq \tau_p$. The τ arguments of D and X will be omitted since D and X are by assumption independent of τ in any one layer. We will also omit the p subscripts, which are implicit on all quantities, until later.

For any layer in Fig. 1, then, we write Eq. (7) in matrix form as

$$\begin{vmatrix} \frac{d\mathbf{u}^+}{d\tau} \\ \frac{d\mathbf{u}^-}{d\tau} \end{vmatrix} = \begin{vmatrix} -\alpha & -\beta \\ \beta & \alpha \end{vmatrix} \begin{vmatrix} \mathbf{u}^+ \\ \mathbf{u}^- \end{vmatrix} + \begin{vmatrix} \tilde{\mathbf{Q}}^+ \\ \tilde{\mathbf{Q}}^- \end{vmatrix}, \quad (7b)$$

where[3,16]

$$\mathbf{u}^\pm = [u^m(\tau,\pm\mu_i)] \qquad i = 1,\ldots,N,$$

$$\tilde{\mathbf{Q}}^\pm = \mathbf{M}^{-1}\mathbf{Q}^\pm,$$

$$\mathbf{Q}^{\pm} = Q^m(\tau,\pm\mu_i) \qquad i = 1,\ldots,N,$$

$$\mathbf{M} = \{\mu_i\,\delta_{ij}\} \qquad i,j = 1,\ldots,N,$$

$$\alpha = \mathbf{M}^{-1}(\mathbf{D}^{+}\mathbf{W} - \mathbf{I}),$$

$$\beta = \mathbf{M}^{-1}\mathbf{D}^{-}\mathbf{W},$$

$$\mathbf{W} = (w_i\delta_{ij}) \qquad i,j = 1,\ldots,N,$$

$$\mathbf{D}^{+} = [D^m(\mu_i,\mu_j)] = [D^m(-\mu_i,-\mu_j)] \quad i,j = 1,\ldots,N,$$

$$\mathbf{D}^{-} = [D^m(-\mu_i,\mu_j)] = [D^m(\mu_i,-\mu_j)] \quad i,j = 1,\ldots,N.$$

C. Quadrature Rule

The use of Gaussian quadrature is essential because it makes phase function renormalization unnecessary, i.e.,

$$\sum_{\substack{j=-N\\ j\neq 0}}^{N} w_j D^0(\tau,\mu_i,\mu_j) = \sum_{\substack{i=-N\\ i\neq 0}}^{N} w_i D^0(\tau,\mu_i,\mu_j) = \omega(\tau), \tag{7c}$$

implying that energy is conserved in the computation.[17]

The quadrature points and weights of the double-Gauss scheme adopted here satisfy $\mu_{-j} = -\mu_j$ and $w_{-j} = w_j$. Double-Gauss simply refers to a quadrature rule suggested by Sykes[18] in which the Gaussian formula is applied separately to the half-ranges $-1 < \mu < 0$ and $0 < \mu < 1$. The main advantage of this double-Gauss scheme is that the quadrature points (in even orders) are distributed symmetrically around $|\mu| = 0.5$ and clustered both toward $|\mu| = 1$ and $\mu = 0$, whereas in the Gaussian scheme for the complete range, $-1 < \mu < 1$, they are clustered toward $\mu = \pm 1$. The clustering toward $\mu = 0$ will give superior results near the boundaries where the intensity varies rapidly around $\mu = 0$. A half-range scheme is also preferred since the intensity is discontinuous at the boundaries. Another advantage is that upward and downward fluxes and average intensities are obtained immediately without further approximations.

D. Basic Solution

Equation (7b) is a system of $2N$-coupled ordinary differential equations with constant coefficients. These coupled equations are linear, and our goal is to uncouple them by using well-known methods of linear algebra.

Seeking solutions to the homogeneous version ($\tilde{\mathbf{Q}} = \mathbf{O}$) of Eq. (7b) of the form

$$\mathbf{u}^{\pm} = \mathbf{G}^{\pm}\exp(-k\tau), \tag{8a}$$

we find

$$\begin{vmatrix} \alpha & \beta \\ -\beta & -\alpha \end{vmatrix} \begin{vmatrix} \mathbf{G}^{+} \\ \mathbf{G}^{-} \end{vmatrix} = k \begin{vmatrix} \mathbf{G}^{+} \\ \mathbf{G}^{-} \end{vmatrix}, \tag{8b}$$

which is a standard algebraic eigenvalue problem of order $2N \times 2N$ determining the eigenvalues k and the eigenvectors $\mathbf{G}^{\pm}$.

Because of the special structure of the matrix in Eq. (8b), the eigenvalues occur in pairs ($\pm k$), and the order of the algebraic eigenvalue problem [Eq. (8b)] may be reduced by a factor of 2 as follows.[3] Breaking Eq. (8b) into

$$\alpha\mathbf{G}^{+} + \beta\,\mathbf{G}^{-} = k\mathbf{G}^{+},$$

$$\beta\mathbf{G}^{+} + \alpha\mathbf{G}^{-} = -k\mathbf{G}^{-},$$

and then adding and subtracing these two equations, we find

$$(\alpha - \beta)\,(\mathbf{G}^{+} - \mathbf{G}^{-}) = k(\mathbf{G}^{+} + \mathbf{G}^{-}), \tag{8c}$$

$$(\alpha + \beta)\,(\mathbf{G}^{+} + \mathbf{G}^{-}) = k(\mathbf{G}^{+} - \mathbf{G}^{-}). \tag{8d}$$

Combining Eqs. (8c) and (8d), we obtain

$$(\alpha - \beta)\,(\alpha + \beta)\,(\mathbf{G}^{+} + \mathbf{G}^{-}) = k^2(\mathbf{G}^{+} + \mathbf{G}^{-}), \tag{8e}$$

which completes the reduction of the order. Following Stamnes and Swanson[3] we solve Eq. (8e) to obtain eigenvalues and eigenvectors $(\mathbf{G}^{+} + \mathbf{G}^{-})$. We then use Eq. (8d) to determine $(\mathbf{G}^{+} - \mathbf{G}^{-})$.

For beam sources $Q^{(\mathrm{beam})}(\tau,\mu) = X_0(\mu)\exp(-\tau/\mu_0)$ [Eq. (6d)], it is easily verified that a particular solution of Eq. (7) is (omitting m superscript)

$$u(\tau,\mu_i) = Z_0(\mu_i)\exp(-\tau/\mu_0). \tag{9a}$$

For thermal sources the emitted radiation is isotropic, so that $Q^m = 0$, $m > 0$, $Q(\tau) \equiv Q^0(\tau) = (1 - \omega)B(\tau)$. By approximating the Planck function for each layer by a polynomial in τ,[3,19] i.e.,

$$B[T(\tau)] = \sum_{l=0}^{K} b_l\tau^l, \tag{9b}$$

we find that the particular solution may also be expressed as a polynomial in τ:

$$u(\tau,\mu_i) = \sum_{l=0}^{K} Y_l(\mu_i)\tau^l. \tag{9c}$$

The coefficients $Z_0(\mu_i)$ and $Y_l(\mu_i)$ are readily determined by solving systems of linear algebraic equations (see Refs. 3 and 9 for details).

The general solution to Eq. (7a) consists of a linear combination, with coefficients C_j, of all the homogeneous solutions plus the particular solutions for beam and thermal emission sources (omitting the m superscript):

$$u(\tau,\mu_i) = \sum_{j=-N}^{N} C_j G_j(\mu_i)\exp(-k_j\tau) + \delta_{m0}\sum_{l=0}^{K} Y_l(\mu_i)\tau^l. \tag{10a}$$

The k_j and the $G_j(\mu_i)$ are the eigenvalues and eigenvectors obtained as described in Sec. III.A, the μ_i the quadrature angles, and the C_j the constants of integration. We have for convenience included the beam solution in the sum by defining for $j = 0$

$$C_0G_0(\mu_i) \equiv Z_0(\mu_i), \tag{10b}$$

$$k_0 \equiv 1/\mu_0. \tag{10c}$$

E. Boundary Conditions

Below we shall limit ourselves to a linear-in-optical-depth dependence of the Planck function within each layer, since this will give sufficient accuracy for most cases of practical interest. We shall discuss the inclusion of boundary conditions in some detail, since a proper formulation of this aspect of the problem (in terms of the bidirectional reflectivity) is rarely discussed in the open literature.

In general, Eq. (1) must be solved subject to the boundary conditions

$$u(\tau = 0,-\mu,\phi) = u_\infty(\mu,\phi), \tag{11a}$$

$$u(\tau = \tau_L,+\mu,\phi) = u_g(\mu,\phi), \tag{11b}$$

where u_∞ and u_g are the intensities incident at the top and bottom boundaries, respectively, and τ_L is the total optical depth.

Quite general boundary conditions can be accommodated in the discrete ordinate method. To be specific, we shall assume that the medium is illuminated from above by known diffuse radiation (incident parallel beams are treated as pseudosources) and that the bottom boundary has a known bidirectional reflectivity. Equation (11b) may now be rewritten as

$$u(\tau_L,+\mu,\phi) = \epsilon(\mu)B(T_g) + \frac{1}{\pi}\int_0^{2\pi} d\phi' \int_0^1 \rho_d(\mu,\phi;-\mu',\phi')u(\tau_L,-\mu',\phi')\mu' d\mu' + \frac{\mu_0}{\pi} I_0 \exp(-\tau_L/\mu_0)\rho_d(\mu,\phi;-\mu_0,\phi_0), \tag{12}$$

where $\epsilon(\mu)$ is the directional emissivity and T_g is the temperature of the bottom boundary; $\rho_d(\mu,\phi;-\mu',\phi')$ is the bidirectional reflectivity, and I_0 is the incident beam intensity at the upper boundary. Kirchhoff's law

$$\epsilon(\mu) + \frac{1}{\pi}\int_0^{2\pi} d\phi' \int_0^1 \rho_d(\mu,\phi;-\mu',\phi')\mu' d\mu' = 1$$

is applied to calculate the emissivity from the reflectivity. Note that some authors include the $(1/\pi)$ factor in ρ_d; we find our definition more useful, however, since when ρ_d is independent of angle it reduces to an albedo.

Below we will assume that the bidirectional reflectivity is a function only of the angle θ between the incident and the reflected radiation (i.e., there are no preferred directions at the lower boundary). Then it can be expanded in a series of $2N$ Legendre polynomials:

$$\rho_d(\mu,\phi;-\mu',\phi') = \sum_{l=0}^{2N-1} (2l+1)h_l P_l(\cos\theta), \tag{13a}$$

where

$$\cos\theta = -\mu\mu' + (1-\mu^2)^{1/2}(1-\mu'^2)^{1/2}\cos(\phi'-\phi). \tag{13b}$$

Using the addition theorem for spherical harmonics, we find

$$\rho_d(\mu,\phi;-\mu',\phi') = \sum_{m=0}^{2N-1} \rho_d^m(\mu,-\mu')\cos m(\phi'-\phi), \tag{13c}$$

$$\rho_d^m(\mu,-\mu') = (2-\delta_{0m})\sum_{l=m}^{2N-1}(2l+1)h_l^m P_l^m(-\mu')P_l^m(\mu), \tag{14}$$

where

$$h_l^m = h_l\frac{(l-m)!}{(l+m)!},$$

$$h_l = \frac{1}{2}\int_{-1}^{1} P_l(\cos\theta)\rho_d(\cos\theta)d\cos\theta. \tag{15}$$

Substituting Eq. (13c) into Eq. (12) and using Eq. (5), we find that each Fourier component must satisfy the bottom boundary condition

$$u^m(\tau_L,+\mu) = u_g^m(\mu), \tag{16a}$$

where

$$u_g^m(u) = \delta_{0m}\epsilon(\mu)B(T_g) + (1+\delta_{0m})\int_0^1 \mu'\rho_d^m(\mu,-\mu')u^m(\tau_L,-\mu')d\mu' + \frac{\mu_0}{\pi} I_0 \exp(-\tau_L/\mu_0)\,\rho_d^m(\mu,-\mu_0). \tag{16b}$$

In a multilayered medium we must also require the intensity to be continuous across layer interfaces.[7] Thus Eq. (7) must satisfy boundary and continuity conditions as follows:

$$u_1^m(0,-\mu_i) = u_\infty^m(-\mu_i), \qquad i = 1,\ldots,N, \tag{17a}$$

$$u_p^m(\tau_p,\mu_i) = u_{p+1}^m(\tau_p,\mu_i), \qquad i = \pm1,\ldots,\pm N; \quad p = 1,\ldots,L-1; \tag{17b}$$

$$u_L^m(\tau_L,+\mu_i) = u_g^m(\mu_i), \qquad i = 1,\ldots,N, \tag{17c}$$

where

$$u_g^m(\mu_i) = \delta_{0m}\epsilon(\mu_i)B(T_g) + (1+\delta_{0m})\sum_{j=1}^{N} w_j\mu_j\rho_d^m(\mu_i,-\mu_j)u^m(\tau_L,-\mu_j) + \frac{\mu_0}{\pi} I_0 \exp(-\tau_L/\mu_0)\,\rho_d^m(\mu_i,-\mu_0). \tag{17d}$$

Evaluation of the integral in Eq. (16b) by numerical quadrature as in Eq. (17d) requires a quadrature rule which integrates separately over the downward and upward directions. As noted previously, the double-Gauss rule adopted here satisfies this requirement.

When discussing boundary conditions, it is convenient to write the discrete ordinate solution in the following form ($k_{jp} > 0$ and $k_{-jp} = -k_{jp}$):

$$u_p(\tau,\mu_i) = \sum_{j=1}^{N}[C_{jp}G_{jp}(\mu_i)\exp(-k_{jp}\tau) + C_{-jp}G_{-jp}(\mu_i)\exp(+k_{jp}\tau)] + R_p(\tau,\mu_i), \tag{18}$$

where the sum contains the homogeneous solution involving the unknown coefficients (the C_{jp}) to be determined, and $R_p(\tau,\mu_i)$ is the particular solution given by [Eq. (10a)]

$$R_p(\tau,\mu_i) = Z_0(\mu_i)\exp(-\tau/\mu_0) + \delta_{m0}[Y_0(\mu_i) + Y_1\tau]. \tag{19}$$

Insertion of Eq. (18) into Eqs. (17a)–(17c) yields (omitting the m superscript)

$$\sum_{j=1}^{N} [C_{j1}G_{j1}(-\mu_i) + C_{-j1}G_{-j1}(-\mu_i)]$$

$$= u_\infty(-\mu_i) - R_p(0,-\mu_i), \qquad i = 1,\ldots,N \tag{20a}$$

$$\sum_{j=1}^{N} \{C_{jp}G_{jp}(\mu_i)\exp(-k_{jp}\tau_p) + C_{-jp}G_{-jp}(\mu_i)\exp(k_{jp}\tau_p) - [C_{j,p+1}G_{j,p+1}(\mu_i)\exp(-k_{j,p+1}\tau_p) + C_{-j,p+1}G_{-j,p+1}(\mu_i)\exp(k_{j,p+1}\tau_p)]\}$$

$$= [R_{p+1}(\tau_p,\mu_i) - R_p(\tau_p,\mu_i)], \qquad i = \pm 1,\ldots \pm N;\ p = 1,\ldots,L-1 \tag{20b}$$

$$\sum_{j=1}^{N} [C_{jL}r_{jL}(\mu_i)\exp(-k_{jL}\tau_L) + C_{-jL}r_{-jL}(\mu_i)\exp(k_{jL}\tau_L)]$$

$$= \gamma(\tau_L,\mu_i), \qquad i = 1,\ldots,N, \tag{20c}$$

where

$$\gamma(\tau_L,\mu_i) = \delta_{0m}\epsilon(\mu_i)B(T_g) - R_L(\tau_L,+\mu_i) + \frac{\mu_0}{\pi} I_0 \exp(-\tau_L/\mu_0)\,\rho_d(\mu_i,-\mu_0) + (1+\delta_{0m})\sum_{j=1}^{N} \rho_d(\mu_i,-\mu_j)w_j\mu_j R_L(\tau_L,-\mu_j), \tag{21a}$$

$$r_{jL}(\mu_i) = G_{jL}(\mu_i) - (1+\delta_{0m}) \times \sum_{n=1}^{N} \rho_d(\mu_i,-\mu_n)w_n\mu_n G_{jL}(-\mu_n). \tag{21b}$$

Equations (20a)–(20c) constitute a $(2N \times L) \times (2N \times L)$ system of linear algebraic equations from which the $2N \times L$ unknown coefficients, the C_{jp} ($j = \pm 1,\ldots,\pm N$; $p = 1,\ldots,L$) are determined. To solve this system of equations, we take advantage of the fact that the coefficient matrix is a $(6N-1)$ diagonal band matrix.[7]

III. Numerical Implementation

A. Computation of Eigenvalues and Eigenvectors

Stamnes and Swanson[3] obtained the eigenvalues and eigenvectors of Eq. (8e) using subroutines in EISPACK[20] for a general real asymmetric matrix. Since such matrices may have complex eigenvalues/eigenvectors, the EISPACK routines use complex arithmetic. Yet invariably the eigenvalues returned were purely real as they should be.

As discussed by Stamnes *et al.*,[8] the advantage of solving the eigenvalue problem involving the asymmetric matrix $(\alpha-\beta)(\alpha+\beta)$ in Eq. (8e) is that only one matrix multiplication is necessary. Nakajima and Tanaka[16] showed that the eigenvalue problem can be reduced to finding eigenvalues and eigenvectors of the product of two symmetric matrices, one of which is positive definite. Stamnes *et al.*[8] used the Cholesky decomposition of this positive definite matrix to convert the eigenvalue problem into one involving only one symmetric matrix. But the transformation to symmetric matrices (whose eigensolutions can be found faster and by using real arithmetic) and subsequent solution procedures[8,16] introduce matrix operations in which the effect of rounding error can seriously degrade the results. Stamnes *et al.*[8] compared the original Stamnes/Swanson method with the new procedure based on the Cholesky decomposition and with Nakajima and Tanaka's symmetrizing procedure. They summarize the outcome of these comparisons as follows:

(1) In single precision the Stamnes/Swanson procedure is more accurate than the Cholesky procedure, which in turn is slightly more accurate than the Nakajima/Tanaka procedure.

(2) The Cholesky procedure is slightly faster than Nakajima/Tanaka's, which is slightly faster than Stamnes/Swanson's.

Based on the findings, they decided to speed up the Stamnes/Swanson procedure since it is the most accurate. By eliminating complex arithmetic in the EISPACK subroutines, the speed became comparable with that of the Cholesky approach. We have adopted this faster algorithm in our code.

B. Scaling Transformation

As discussed by Stamnes and Conklin,[7] to avoid numerical ill-conditioning, it is necessary to remove the positive exponentials in Eq. (20) (remember $k_{jp} > 0$ by convention). This is achieved by the scaling transformation[7]

$$C_{+jp} = \tilde{C}_{+jp}\exp(k_{jp}\tau_{p-1}), \tag{22a}$$

$$C_{-jp} = \tilde{C}_{-jp}\exp(-k_{jp}\tau_p). \tag{22b}$$

Inserting Eqs. (22) in Eqs. (20) and solving for the $\tilde{C}_{jp}$ instead of the C_{jp}, we find that all the exponential terms in the coefficient matrix have the form $\exp[-k_{jp}(\tau_p - \tau_{p-1})]$; i.e., all the exponentials in the coefficient matrix have negative arguments ($k_{jp} > 0$, $\tau_p > \tau_{p-1}$). Consequently, numerical ill-conditioning is avoided implying that our numerical scheme for finding the $\tilde{C}_{jp}$ is unconditionally stable for arbitrary total optical depths and arbitrary individual layer thicknesses. [To solve Eqs. (20) for the $\tilde{C}_{jp}$ we use the LINPACK[21] routine SGBSL designed to solve a banded system of linear algebraic equations.]

C. Scaled Solutions—Angular Distributions

Since only the homogeneous solution is affected, we start by introducing the scaling transformation (22) into Eq. (18), i.e.,

$$u_p(\tau,\mu_i) = \sum_{j=1}^{N} \{\tilde{C}_{jp}G_{jp}(\mu_i)\exp[-k_{jp}(\tau - \tau_{p-1})] + \tilde{C}_{-jp}G_{-jp}(\mu_i)\exp[-k_{jp}(\tau_p - \tau)]\}. \tag{23}$$

Since $k_{jp} > 0$ and $\tau_{p-1} \le \tau \le \tau_p$, all exponentials in Eq. (23) have negative arguments as they should to avoid fatal overflows.

For a slab of thickness τ_L, we may solve Eq. (1) formally to obtain ($\mu > 0$)

$$u(\tau,+\mu) = u(\tau_L,+\mu)\exp[-(\tau_L-\tau)\mu] + \int_{\tau}^{\tau_L} S(t,+\mu)\exp[-(t-\tau)/\mu]\frac{dt}{\mu}, \quad (24a)$$

$$u(\tau,-\mu) = u(0,-\mu)\exp(-\tau/\mu) + \int_{0}^{\tau} S(t,-\mu)\exp[-(\tau-t)/\mu]\frac{dt}{\mu}, \quad (24b)$$

where we have again omitted the m superscript. These two equations show that if we know the source function $S(t,\pm\mu)$, we can find the intensity at arbitrary angles by integrating the source function. The discrete ordinate solutions are used to derive explicit expressions for the source function, which can be integrated analytically.[3–5,9] This procedure is sometimes referred to as the iteration of the source-function technique, although as pointed out by Kourganoff[22] and discussed in some detail by Stamnes,[5] this procedure essentially amounts to an interpolation.

In a multilayered medium we may now integrate Eq. (24a) layer by layer[5] to obtain the following expression for the upward intensity at arbitrary τ and μ:

$$u_p(\tau,+\mu) = u(\tau_L,+\mu)\exp[-(\tau_L-\tau)/\mu] + \sum_{n=p}^{L}\left\{\sum_{j=1}^{N}\left[\tilde{C}_{-jn}\frac{G_{-jn}(+\mu)}{1-k_{jn}\mu}E_{-jn}(\tau,+\mu) + \sum\tilde{C}_{+jn}\frac{G_{+jn}(+\mu)}{1+k_{jn}\mu}E_{+jn}(\tau,+\mu)\right] + \frac{G_{0n}}{1+\mu/\mu_0}E_{0n}(\tau,+\mu) + \delta_{m0}[V_{0n}(\mu)F_{0n}(\tau,\mu) + b_{1n}F_{1n}(\tau,\mu)]\right\}, \quad (25a)$$

where the scaling given by Eqs. (22) has been incorporated and

$$E_{-jn}(\tau,+\mu) = \exp[-(k_{jn}\Delta\tau_n - \delta\tau/\mu)] - \exp[-(\tau_n-\tau)/\mu], \quad (25b)$$

$$F_{0n}(\tau,+\mu) = \exp(-\delta\tau/\mu) - \exp[-(\tau_n-\tau)/\mu], \quad (25c)$$

$$F_{1n}(\tau,+\mu) = (\tau_{n-1}+\mu)\exp(-\delta\tau/\mu) - (\tau_n+\mu)\exp[-(\tau_n-\tau)/\mu], \quad (25d)$$

with

$$\delta\tau = \tau_{n-1} - \tau \text{ for } n > p \text{ and } \delta\tau = 0 \text{ for } n = p;$$

$$\Delta\tau_n = \tau_n - \tau_{n-1} \text{ for } n > p \text{ and } \Delta\tau_p = \tau_p - \tau \text{ for } n = p,$$

$$E_{+jn}(\tau,+\mu) = \exp[-(\tau_{n-1}-\tau)/\mu] - \exp\{-[k_{jn}(\tau_n-\tau_{n-1}) - (\tau-\tau_n)/\mu]\}, \quad (25e)$$

for $n > p$ and

$$E_{+jp}(\tau,+\mu) = \exp[-k_{jp}(\tau-\tau_{p-1})] - \exp\{-[k_{jp}(\tau_p-\tau_{p-1}) + (\tau_p-\tau)/\mu]\}. \quad (25f)$$

Similarly, we find that the downward intensity becomes

$$u_p(\tau,-\mu) = u(0,-\mu)\exp(-\tau/\mu) + \sum_{n=1}^{p}\left\{\sum_{j=1}^{N}\left[\tilde{C}_{-jn}\frac{G_{-jn}(-\mu)}{1+k_{jn}\mu}E_{-jn}(\tau,-\mu) + \tilde{C}_{+jn}\frac{G_{+jn}(-\mu)}{1-k_{jn}\mu}E_{+jn}(\tau,-\mu)\right] + \frac{G_{0n}(-\mu)}{1-\mu/\mu_0}E_{0n}(\tau,-\mu) + \delta_{m0}[V_{0n}(\mu)F_{0n}(\tau,-\mu) + b_{1n}F_{1n}(\tau,-\mu)]\right\}, \quad (26a)$$

where

$$E_{+jn}(\tau,-\mu) = \exp[-(k_{jn}\Delta\tau_n + \delta\tau/\mu)] - \exp[-(\tau-\tau_{n-1})/\mu] \quad (26b)$$

$$F_{0n}(\tau,-\mu) = \exp(-\delta\tau/\mu) - \exp[-(\tau-\tau_{n-1})/\mu] \quad (26c)$$

$$F_{1n}(\tau,-\mu) = (\tau_n-\mu)\exp(-\delta\tau/\mu) - (\tau_{n-1}-\mu)\exp[-(\tau-\tau_{n-1})/\mu], \quad (26d)$$

with

$$\delta\tau = \tau - \tau_n \text{ for } n < p \text{ and } \delta\tau = 0 \text{ for } n = p,$$

$$\Delta\tau_n = \tau_n - \tau_{n-1} \text{ for } n < p \text{ and } \Delta\tau_p = \tau - \tau_{p-1} \text{ for } n = p,$$

$$E_{-jn}(\tau,-\mu) = \exp[-(\tau-\tau_n)/\mu] - \exp\{-[k_{jn}(\tau_n-\tau_{n-1}) + (\tau-\tau_{n-1})/\mu]\}, \quad (26e)$$

for $n < p$ and

$$E_{-jp}(\tau,-\mu) = \exp[-k_{jp}(\tau_p-\tau)] - \exp\{-[k_{jp}(\tau_p-\tau_{p-1}) + (\tau-\tau_{p-1})/\mu]\}. \quad (26f)$$

We see that all exponentials in Eqs. (25b)–(25f) and (26b)–(26f) have negative arguments as they should. This ensures that fatal overflow errors are avoided in the computations.

The G_{jn} terms and $\bar{V}_{0n}(\mu)$ in Eqs. (25a) and (26a) are given by the following explicit expressions (omitting the n subscript):

$$G_j(\mu) = \sum_{\substack{i=-N\\ i\neq 0}}^{N} w_i D(\mu,\mu_i)G_j(\mu_i) \text{ for } j \neq 0; \quad (27a)$$

$$G_0(\mu) \equiv Z_0(\mu) = \sum_{\substack{i=-N\\ i\neq 0}}^{N} w_i D(\mu,\mu_i)Z_0(\mu_i) + X_0(\mu); \quad (27b)$$

$$V_0(\mu) = \sum_{\substack{i=-N\\ i\neq 0}}^{N} w_i D^0(\mu,\mu_i)Y_0(\mu_i) + (1-\omega)b_0; \quad (27c)$$

where b_0 is defined in Eq. (9b). Formulas (27a)–(27c) clearly reveal the interpolatory nature of Eqs. (25) and (26). Although we have generally omitted the m superscript above, we have explicitly written D^0 in Eq. (27c) to indicate that thermal radiation contributes only to the azimuth-independent part of the intensity.

D. Computational Shortcut for Highly Absorbing Media

For highly absorbing media without thermal emission, very little radiation will reach levels for which the absorbing optical depth is greater than, say, 10. In such cases we ignore the medium below this depth and set the surface reflection to zero. This may result in enormous computation savings. For example, in an atmosphere with clouds it prevents unnecessary scattering computations at wavelengths where sunlight never reaches the clouds.

E. Simplified Albedo and Transmissivity Computations

By appealing to the reciprocity principle we may derive simple expressions for the plane albedo $a(\mu)$ and the transmissivity $t(\mu)$ of a vertically inhomogeneous plane-parallel medium lacking thermal sources. Thus

(1) the plane albedo for a given angle of parallel incidence equals the azimuthally averaged reflected intensity $u^0(0,\mu)$ for isotropic illumination of unit intensity incident at the top boundary;

(2) the transmissivity equals the azimuthally averaged transmitted intensity $u^{0*}(0,\mu)$ for isotropic illumination of unit intensity incident at the bottom boundary.

Consequently, for an inhomogeneous medium the following duality relations apply[6]:

$$a(\mu) = u^0(0,\mu);\ a^*(\mu) = u^{0^*}(\tau_L,-\mu); \tag{28}$$

$$t(\mu) = u^{0*}(0,\mu);\ t^*(\mu) = u^0(\tau_L,-\mu); \tag{29}$$

where τ_L is to the optical thickness of the inhomogeneous medium, and the asterisk refers to illumination from below.

Equations (28) and (29) offer substantial computational advantages when only albedo and transmissivity are required. The conventional procedure requires computation of the angular distribution of the diffuse intensity, which must then be integrated over angle to obtain the desired result. This necessitates the computation of a particular solution, which can be quite costly in an inhomogeneous (multilayered) medium, since such a solution is required for each layer and every direction of incidence considered. In contrast, by applying an isotropic boundary condition and using Eqs. (28) and (29), one may solve for all desired angles of parallel beam incidence simultaneously and avoid entirely the computation of a particular solution. (Details about the numerical implementation of this procedure are provided in Refs. 6 and 9.)

IV. Summary

We have described an advanced discrete ordinate algorithm for radiative transfer calculations, including multiple scattering, in vertically inhomogeneous nonisothermal plane-parallel media. This algorithm is the product of over seven years of development by ourselves and our collaborators. It has been designed to be as versatile and general as possible and should find applications from the UV through to the radar region of the electromagnetic spectrum.

The algorithm has the following important features:

(1) It is unconditionally stable for an arbitrarily large number of quadrature angles (streams) and arbitrarily large optical depths.

(2) It can be forced by any combination of parallel beam or diffuse incidence and thermal sources in the medium and at the boundaries.

(3) It allows for an arbitrary bidirectional reflectivity at the lower boundary; directional emissivity is computed from this reflectivity.

(4) It offers rapid computation of bulk albedo and transmissivity when thermal sources are absent.

(5) Unlike the popular doubling method, computing time for individual layers is independent of optical thickness.

(6) Because the solution is analytic, and by using the iteration-of-the-source-function method, intensities can be returned at arbitrary angles and optical depths, unrelated to the computational meshes for these quantities.

(7) It has been thoroughly tested against a wide variety of published solutions.

(8) It is thoroughly documented both in Ref. 9 and in the code itself (with extensive references to published equation numbers).

Copies of the FORTRAN program are available from the third author on microcomputer diskettes (IBM or MacIntosh).

This work was supported by the National Science Foundation through grant DPP 84-06093 to the University of Alaska. One of us (K.S.) would like to acknowledge partial support of this work from the Royal Norwegian Council on Science and Technology.

References

1. S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).
2. K.-N. Liou, "A Numerical Experiment on Chandrasekhar's Discrete-Ordinate Method for Radiative Transfer: Applications to Cloudy and Hazy Atmosphere," J. Atmos. Sci. **30**, 1303 (1973).
3. K. Stamnes and R. A. Swanson, "A New Look at the Discrete Ordinate Method for Radiative Transfer Calculations in Anisotropically Scattering Atmospheres," J. Atmos. Sci. **38**, 387 (1981).
4. K. Stamnes and H. Dale, "A New Look at the Discrete Ordinate Method for Radiative Transfer Calculation in Anisotropically Scattering Atmospheres, II: Intensity Computations," J. Atmos. Sci. **38**, 2696 (1981).
5. K. Stamnes, "On the Computation of Angular Distributions of Radiation in Planetary Atmospheres," J. Quant. Spectrosc. Radiat. Transfer **28**, 47 (1982a).
6. K. Stamnes, "Reflection and Transmission by a Vertically Inhomogeneous Planetary Atmosphere," Planet. Space Sci. **30**, 727 (1982b).
7. K. Stamnes and P. Conklin, "A New Multi-Layer Discrete Ordinate Approach to Radiative Transfer in Vertically Inhomogeneous Atmospheres," J. Quant. Spectrosc. Radiat. Transfer **31**, 273 (1984).
8. K. Stamnes, S.-C. Tsay, and T. Nakajima, "Computation of Eigenvalues and Eigenvectors for Discrete Ordinate and Matrix Operator Method Radiative Transfer," J. Quant. Spectrosc. Radiat. Transfer in press (1988).

9. K. Stamnes, S.-C. Tsay, W. J. Wiscombe, and K. Jayaweera, "An Improved, Numerically Stable Computer Code for Discrete-Ordinate-Method Radiative Transfer in Scattering and Emitting Layered Meda," NASA Report, in press (1988).
10. K. N. Liou, *An Introduction to Atmospheric Radiation* (Academic, Orlando, FL, 1980).
11. J. E. Hansen and L. D. Travis, "Light Scattering in Planetary Atmospheres," Space Sci. Rev. **16,** 527 (1974).
12. W. M. Irvine, "Multiple Scattering in Planetary Atmospheres," Icarus **25,** 175 (1975).
13. J. Lenoble, Ed., *Radiative Transfer in Scattering and Absorbing Atmospheres: Standard Computational Procedures* (A. Deepak, Hampton, VA, 1985).
14. W. J. Wiscombe, "Atmospheric Radiation: 1975–1983," Rev. Geophys. **21,** 957 (1983).
15. K. Stamnes, "The Theory of Multiple Scattering of Radiation in Plane Parallel Atmospheres," Rev. Geophys. **24,** 299 (1986).
16. T Nakajima and M. Tanaka, "Matrix Formulations for the Transfer of Solar Radiation in a Plane-Parallel Atmosphere," J. Quant. Spectrosc. Radiat. Transfer **35,** 13 (1986).
17. W. J. Wiscombe, "The Delta-M Method: Rapid Yet Accurate Radiative Flux Calculations for Strongly Asymmetric Phase Functions," J. Atmos. Sci. **34,** 1408 (1977).
18. J. B. Sykes, "Approximate Integration of the Equation of Transfer," Mon. Not. Roy. Astron. Soc. **11,** 377 (1951).
19. W. J. Wiscombe, "Extension of the Doubling Method to Inhomogeneous Sources," J. Quant. Spectrosc. Radiat. Transfer **16,** 477 (1976).
20. W. R. Cowell, Ed., *Sources and Developments of Mathematical Software* (Prentice Hall, Englewood Cliffs, NJ, 1980).
21. J. J. Dongarra, C. B. Moler, J. R. Bunch, and G. W. Stewart, LINPACK *User's Guide* (SIAM, Philadelphia, 1979).
22. V. Kourganoff, *Basic Methods in Transfer Problems* (Dover, New York, 1963).

NASA *continued from page 2450*

Computing radiation characteristics of phased arrays

One of the most important radiation characteristics of an antenna system is the directivity. The accurate determination of this parameter is essential for the analysis and design of advanced antenna systems. In general, for most commonly used array elements like open-ended waveguides, horns, or microstrip patch antennas, an analytical expression of the $(\cos\theta)^q$ type can be used to tailor actual element patterns properly. Experimentally measured element patterns could include mutual coupling, which may be significant in large arrays. Expressions for the directivity and radiation pattern of a single element and of a rectangluar array using $(\cos\theta)^q$-type element patterns have been reported.

With the advent of monolithic microwave integrated circuits, phased arrays have become key components in the design of advanced antenna systems. Array-fed antennas are used extensively in today's multiple-beam satellite antennas.

A computer program was devised to generalize these results and obtain an efficient numerical technique for computing the directivity and the radiation pattern of the generalized array. The characteristics of the generalized array used in this program include arbitrary element locations; element patterns that are of the $(\cos\theta)^q$ type, analytically describable by other functions, or experimentally measured; and the polarizations and excitations of the elements. The computer program is based on an efficient numerical technique for calculating the radiated power (Romberg integration), directivity, and radiation pattern of a phased array. The formulation is general and takes into account arbitrary element polarization, electric-field-plane and magnetic-field-plane element patterns, element location, and complex element excitation.

One of the advantages of this generalized array formulation is that it does not break down in special cases as might other approximations that use closed forms. Also, the formulation does not limit the observation of the pattern to the far-field zone. This can be useful if the generalized formulation is used with analysis programs for dual-reflector configurations. The program can be modified easily for use in the analysis of reflectors with phased-array feeds.

This computer program is one of the principal research tools at NASA Lewis Research Center for the analysis of advanced space-communication antenna systems. The generalized formulation and computer program provide complete flexibility in the analysis of array configurations and in the accurate analysis of experimental data. The program is written in FORTRAN IV and can be used on an IBM 370 computer.

This program was written by Roberto J. Acosta of Lewis Research Center. Refer to LEW-14460.

Mark III Hypercube-ensemble computers

The Mark III Hypercube concept is being applied in the development of a series of increasingly powerful computers. When fully developed, the Mark III computers will be especially useful for the solution of physical problems with spatially and temporally varying quantities. Disciplines that will benefit from the speed and memory capacity of the Mark III Hypercube include astrophysics, geophysics, chemistry, weather, high-energy physics, applied mechanics, image processing, oil exploration, aircraft design, and microcircuit design.

As in previous Hypercube networks, each of $N = 2^k$ identical concurrent processors is located at a node that corresponds to a corner of a k-dimensional cube. Each processor is connected to its k nearest neighbors along the topological equivalents of the k edges that meet at that corner. The prototype contained 4 nodes; two additional 32-node models were built and are operational. A subsequent 1 gigaflop 128-node model is now under construction and will be operational by the end of 1988.

The Mark III concept includes powerful nodes connected by high-speed communication links. The processor at each node is composed of three subprocessors that are tightly coupled through a 4-Mbyte main node memory on a common bus (see Fig. 3). The peak rate of data transfer between the main memory and each subprocessor is 12 Mbytes/s. To minimize the degradation of performance caused by simultaneous attempts of the subprocessors to gain access to the main node memory, each subprocessor is also provided with a high-speed local memory. The main subprocessor arbitrates the memory contentions.

The input/output control subprocessor includes an MC68020 microprocessor, 64 Kbytes of high-speed local memory, and a direct-memory-access channel-support module. This subprocessor is capable of an input/output-channel transfer rate up to 68 Mbits/s. The main subprocessor is also based on the MC68020 and operates at over 2 million instructions per second. It includes a 32-bit arithmetic and logic unit and a full 32-bit address generator. The local memory of 128 Kbytes has faster access than does the shared main memory; this allows the microprocessor to reach its full execution speed.

The high-speed floating-point subprocessor is a peripheral unit attached to the node board by a daughter-board connector. This subprocessor is responsible for heavy floating-point arithmetic: it can increase the speed of such computations to as much as fifty times that achievable with the floating-point coprocessor that is part of the main subprocessor. The shared main memory is partitioned into four address-interleaved dynamic memory modules, each having its own dynamic memory controller. Circuitry is included to correct 1-bit data errors in memory and to signal (but not to correct) 2-bit

continued on page 2531