Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 219 (2006) 899–911
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Efficient computation of the 2D periodic Green's function using the Ewald method

Siddharth Oroskar, David R. Jackson *, Donald R. Wilton

Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204-4005, USA

Received 12 September 2005; received in revised form 16 May 2006; accepted 28 June 2006
Available online 1 September 2006

Abstract

An efficient computation of the periodic Helmholtz Green's function for a 2D array of point sources using the Ewald method is presented. Limitations on the numerical accuracy when using the "optimum" E parameter (which gives optimum asymptotic convergence) at high frequency are discussed. A "best" E parameter is then derived to overcome these limitations, which allows for the fastest convergence while maintaining a specific level of accuracy (loss of significant figures) in the final result. The actual loss of significant figures has been verified through numerical simulations. Formulas for the number of terms needed for convergence have also been derived for both the spectral and the spatial series that appear in the Ewald method and are found to be accurate in almost all cases.

Keywords: Ewald method; Periodic Green's function; Lattice

1. Introduction

The calculation of the free-space periodic Green's function (FSPGF) is an important problem in physics and engineering [1–3] and the Ewald method [4–6] is a powerful means to efficiently evaluate the FSPGF. In the Ewald method, the FSPGF is expressed as the sum of a modified "spectral" and a modified "spatial" series. The terms of both series possess Gaussian decay, leading to an overall series representation that exhibits a very rapid convergence rate. The convergence rate is optimum when the "optimum" value of the Ewald splitting parameter E is used [5], denoted here as E_{opt}. However, one problem with the Ewald method is that at high frequency (when the periodicity becomes large relative to a wavelength) the numerical accuracy degrades very quickly [6]. This is due to a catastrophic loss of significant figures in the series summation, due to the fact that the $(0,0)$ terms in the two series (and to a lesser extent, other nearby terms) become very large and nearly opposite.

The method studied here limits the size of the largest terms in the series relative to that of the total Green's function by modifying the value of the parameter E to avoid undue loss of accuracy. By increasing the E parameter, the size of the largest terms in the series is limited at the expense of slowing the convergence rate.

* Corresponding author. Tel.: +1 713 743 4426; fax: +1 713 743 4444.
E-mail address: djackson@uh.edu (D.R. Jackson).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2006.06.050

Hence, there is a tradeoff between the size of the largest term allowed, which determines the number of significant figures lost and the series convergence rate. A parameter E_L is then obtained based on a user-defined quantity L, which represents the tolerable loss of significant figures. This "best" value E_L then yields the fastest convergence of the Ewald series while limiting the loss of significant figures to the user-defined limit.

The spatial form of the FSPGF for a 2D periodic array on a general skewed lattice as shown in Fig. 1 is given as (an $e^{j\omega t}$, $j = \sqrt{-1}$ time dependence is assumed and suppressed)

$$G(\mathbf{r},\mathbf{r}') = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} e^{-j\mathbf{k}_{t00}\cdot\boldsymbol{\rho}_{mn}}\left(\frac{e^{-jkR_{mn}}}{4\pi R_{mn}}\right), \tag{1}$$

where

$$R_{mn} = |\mathbf{r} - \mathbf{r}' - m\mathbf{s}_1 - n\mathbf{s}_2|,$$

where R_{mn} is the distance between the observation point at $\mathbf{r} = (x, y, z)$ and the (m, n)th periodic source point located at $\mathbf{r}' + m\mathbf{s}_1 + n\mathbf{s}_2$, $\mathbf{k}_{t00} = \hat{\mathbf{x}}k\sin\theta_0\cos\phi_0 + \hat{\mathbf{y}}k\sin\theta_0\sin\phi_0$ is the transverse phasing wave-vector and $\rho_{mn} = m\mathbf{s}_1 + n\mathbf{s}_2$ is the position vector of the (m, n)th source point in the periodic lattice relative to the reference source in the $(0, 0)$ cell, denoted as $\mathbf{r}' = (x', y', z')$. The lattice vectors of the array are $\mathbf{s}_1$ and $\mathbf{s}_2$. Physically the FSPGF is the time-harmonic scalar potential produced by an array of phased source points lying on the infinite lattice at $\mathbf{r}' + m\mathbf{s}_1 + n\mathbf{s}_2$.

When employing the Ewald method for the evaluation of the FSPGF, the Green's function is expressed as a sum of two series [4–6] so that

$$G(\mathbf{r},\mathbf{r}') = G_{\text{spectral}}(\mathbf{r},\mathbf{r}') + G_{\text{spatial}}(\mathbf{r},\mathbf{r}'). \tag{2}$$

The spectral series $G_{\text{spectral}}(\mathbf{r},\mathbf{r}')$ is given by

$$G_{\text{spectral}}(\mathbf{r},\mathbf{r}') = \frac{1}{A}\sum_{p=-\infty}^{\infty}\sum_{q=-\infty}^{\infty}\frac{e^{-j\mathbf{k}_{tpq}\cdot(\rho-\rho')}}{4jk_{zpq}}\left[e^{-jk_{zpq}|z-z'|}\operatorname{erfc}\left(\frac{jk_{zpq}}{2E} - |z-z'|E\right) + e^{jk_{zpq}|z-z'|}\operatorname{erfc}\left(\frac{jk_{zpq}}{2E} + |z-z'|E\right)\right]. \tag{3}$$

The spatial series $G_{\text{spatial}}(\mathbf{r},\mathbf{r}')$ is given by

$$G_{\text{spatial}}(\mathbf{r},\mathbf{r}') = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{e^{-j\mathbf{k}_{t00}\cdot\rho_{mn}}}{8\pi R_{mn}}\left[e^{-jkR_{mn}}\operatorname{erfc}\left(R_{mn}E - \frac{jk}{2E}\right) + e^{jkR_{mn}}\operatorname{erfc}\left(R_{mn}E + \frac{jk}{2E}\right)\right]. \tag{4}$$

In these equations,

$A = |\mathbf{s}_1 \times \mathbf{s}_2|$ cross sectional area of each lattice cell with edge vectors $\mathbf{s}_1$, $\mathbf{s}_2$,

$\mathbf{k}_{tpq}$ transverse wave-vector of the (p, q)th mode,

$\mathbf{k}_{zpq}$ $\sqrt{k^2 - \mathbf{k}_{tpq}^2}$, where $k = \frac{2\pi}{\lambda}$,

$\boldsymbol{\rho}$, $\boldsymbol{\rho}'$ projections of the source and observation points onto the x–y plane,

$\mathbf{r}$, $\mathbf{r}'$ position vectors for the observation and source points.

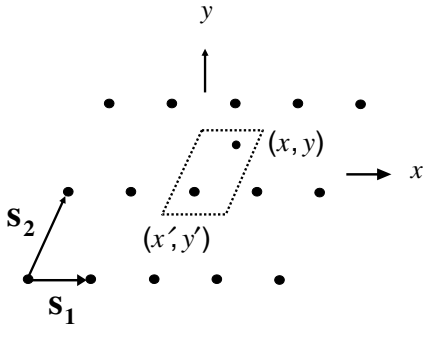

Fig. 1. A periodic lattice of source points [4] is shown with the observation point at (x, y, z).

2. Splitting parameter (E)

As can be seen from the above formulae, the spectral and spatial series are written in terms of a complementary error function involving a "splitting" parameter E. The parameter E controls the convergence rate of the two series. A larger E makes the spatial series G_{spatial} converge faster while a smaller E makes the spectral series G_{spectral} converge faster. The splitting parameter E is an arbitrary number and its "optimum value" E_{opt} is used to balance the asymptotic convergence rate between these two series [5]. It can be shown that this has the effect of minimizing the total number of terms needed in the calculation of the Green's function. The "optimum" E parameter [5] that results in the same asymptotic rate of decay for the G_{spectral} and G_{spatial} series is

$$E_{\text{opt}} = \sqrt{\frac{\pi}{A}}. \tag{5}$$

(The above formula was misprinted in the original article [5], where the π factor was mistakenly taken outside of the square root.) With this choice of E, the rate of exponential decay is the same for both series, and also the coefficients in front of the $(m,n) = (p,q)$ terms in the two series are asymptotically equal. It can be shown that this leads to the minimum overall number of terms needed for the calculation of the two Ewald series [5], and is thus normally the best choice.

However, numerical difficulties are encountered when the lattice separations (periods) becomes large relative to a wavelength. This happens because for large arguments, the complementary error function erfc(z) is approximately $e^{-z^2}/(\sqrt{\pi}z)$ [7]. For large lattice spacings relative to a wavelength, the imaginary part of the argument of erfc becomes large. The first several terms of both the spatial and the spectral series thus have very large values, and each series converges to very large, nearly equal but oppositely signed values. The two series essentially cancel each other, resulting in a sum of moderate value but often with a catastrophic loss of significant figures. To avoid this problem, it is desirable to limit the size of the largest terms of both series. This results in choosing an E value that is greater than the "optimum" value. By increasing E beyond the optimum value, one obtains smaller values for the imaginary part of the argument of the complementary error function. As a result, one avoids loss of accuracy in the addition of the two series and a more accurate result for the total Green's function is obtained [6], at the expense of slower convergence.

In the following sections, a formula for the best E, called E_L, which achieves the best convergence under the constraint of limiting the loss of significant figures to L digits, is obtained for the general non-planar case ($z \neq z'$). This value of E is the smallest value beyond the optimum value that is still large enough to limit to loss of significant figures to L digits.

3. Choice of the splitting parameter

The goal is to limit the size of the largest terms relative to the value of the total Green's function. The largest term in each series arises from the $(0,0)$ terms. We choose the value of $E = E_L$ by enforcing the following conditions:

$$|G_{00,\text{spectral}}| < \alpha 10^L |G| \tag{6}$$

and

$$|G_{00,\text{spatial}}| < \alpha 10^L |G|, \tag{7}$$

where G is the value (or estimate) of the FSPGF and the parameter L indicates (roughly) how many significant figures one is willing to sacrifice in the calculation.

The factor α on the RHS should be chosen as 1/2 for a worst-case error bound, assuming that the error is equal in the two terms on the LHS of (6) and (7). If one of the terms is much larger than the other (when using $E = E_L$), a factor of $\alpha = 1$ is more appropriate. This turns out to be the case, as will be demonstrated later; hence, a factor of $\alpha = 1$ is used here. Roughly speaking, the magnitude of the overall Green's function is

$$|G| \approx \frac{1}{4\pi R_{00}}. \tag{8}$$

This approximation is a reasonable order-of-magnitude estimate unless the distance between the source and observation points becomes comparable to the distance between the source point and the boundary of the unit cell (so that image terms are important).

Consider first the G_{spectral} (modified spectral) series. Using the asymptotic form of the complementary error function [5] at high frequency and defining $\Delta z = |z - z'|$, we have

$$\left| \frac{1}{4Aj k_{z00}\sqrt{\pi}} e^{\left(\frac{k_{z00}}{2E}\right)^2 - (\Delta z E)^2} \left[\frac{1}{\frac{jk_{z00}}{2E} - \Delta z E} + \frac{1}{\frac{jk_{z00}}{2E} + \Delta z E} \right] \right| < \frac{10^L}{4\pi R_{00}}. \tag{9}$$

By solving the above equation for E, we obtain the restriction that

$$E > E_{\text{spect}} = \frac{k_{z00}}{2x_1}, \tag{10}$$

where x_1 satisfies the transcendental equation

$$x_1^2 - \left(\frac{A_1}{x_1}\right)^2 = \ln F_1(x_1), \tag{11}$$

where

$$A_1 = \frac{k_{z00}\Delta z}{2} \tag{12}$$

with $\Delta z = |z - z'|$, and

$$F_1(x_1) = \frac{c_a}{x_1}\left(x_1^2 + \left(\frac{A_1}{x_1}\right)^2\right) \tag{13}$$

with

$$c_a = \frac{10^L k_{z00} A}{2\sqrt{\pi} R_{00}}. \tag{14}$$

The term E_{spect} represents the minimum value of E that will satisfy Eq. (6). The solution to the transcendental Eq. (11) can be easily obtained by a variety of standard methods. One method is iteration, using the iterative formula

$$x_1^{i+1} = \sqrt{\frac{\ln F_1^i + \sqrt{(\ln F_1^i)^2 + 4A_1^2}}{2}} \tag{15}$$

with

$$F_1^i = \left[\frac{c_a}{x_1^i}\left(x_1^{i2} + \left(\frac{A_1}{x_1^i}\right)^2\right)\right]. \tag{16}$$

Eq. (15) comes directly from applying the quadratic formula to (11). Typically only a few iterations are required for convergence when starting with the initial guess $x_1^0 = \sqrt{\ln c_a}$.

The analysis for the G_{spatial} series is similar. In this case we obtain

$$\left| \frac{1}{8\pi R_{00}\sqrt{\pi}} e^{\left(\frac{k}{2E}\right)^2 - (R_{00}E)^2} \left[\frac{1}{R_{00}E - \frac{jk}{2E}} + \frac{1}{R_{00}E + \frac{jk}{2E}} \right] \right| < \frac{10^L}{4\pi R_{00}}. \tag{17}$$

Solving the above equation for E, we obtain the restriction that

$$E > E_{\text{spat}} = \frac{k}{2x_2}, \tag{18}$$

where x_2 is a solution of

$$x_2^2 - \left(\frac{A_2}{x_2}\right)^2 = \ln F_2(x_2), \tag{19}$$

where

$$F_2(x_2) = \frac{c_b x_2}{A_2}\left(x_2^2 + \left(\frac{A_2}{x_2}\right)^2\right) \tag{20}$$

with

$$A_2 = \frac{R_{00}k}{2} \tag{21}$$

and

$$c_b = 10^L\sqrt{\pi}. \tag{22}$$

The term E_{spat} represents the minimum value of E that will satisfy Eq. (7). An iterative formula for the solution of (19) is

$$x_2^{i+1} = \sqrt{\frac{\ln F_2^i + \sqrt{(\ln F_2^i)^2 + 4A_2^2}}{2}}, \tag{23}$$

where

$$F^i = \left[\frac{c_b}{A_2}x_2^i\left(x_2^{i2} + \left(\frac{A_2}{x_2^i}\right)^2\right)\right]. \tag{24}$$

Convergence is usually rapid when starting with the initial guess $x_2^0 = \sqrt{\ln c_b}$.

The best overall splitting parameter is then given by

$$E_L = \max\left(E_{\text{opt}}, \frac{k_{z00}}{2x_1}, \frac{k}{2x_2}\right). \tag{25}$$

This value is the smallest value of E beyond the optimum value (and thus corresponds to the minimum number of total terms required in the Ewald summation) that will still ensure that the largest terms in both the spectral and the spatial series are limited to the required level to avoid losing more than L significant figures when the two series are added together.

4. Number of terms needed for convergence

Having determined the "best" value of the E parameter E_L as a function of frequency, our next goal is to determine how many terms are needed for convergence. We recall that a given value of L has been assumed, which is the number of significant figures that are sacrificed in the calculation. For a given value of $E = E_L$, we next wish to determine how many terms in each series are needed to guarantee convergence of the Green's function to S significant figures. A method is developed here to calculate the summation limits P and Q for the spectral series and M and N for the spatial series. If the machine precision is T significant figures, the value of S that is specified should be limited to $S < T - L$.

As before, the overall magnitude of the Green's function is approximated as in (8). First, consider the summation limits for M and P. For the two series, we require that

$$|G_{\text{spatial},(M+1,0)}| + |G_{\text{spatial},(-M-1,0)}| + |G_{\text{spatial},(0,N+1)}| + |G_{\text{spatial},(0,-N-1)}| < 10^{-S}|G|\left(\frac{1}{2}\right) \tag{26}$$

and

$$|G_{\text{spectral},(P+1,0)}| + |G_{\text{spectral},(-P-1,0)}| + |G_{\text{spectral},(0,Q+1)}| + |G_{\text{spectral},(0,-Q-1)}| < 10^{-S}|G|\left(\frac{1}{2}\right). \tag{27}$$

The error in stopping the summations has been approximated in the above equations by the sum of the four values that give the largest contributions outside the rectangle of summed values. The 1/2 factors on the right-hand sides are present because the error in each series is limited to 1/2 the total error. It is next assumed that the contributions from the first two terms on the LHS of (26) are roughly equal, and similarly for the third and fourth terms. The error from the first two terms is limited to 1/2 the total (from all four terms) in (26), and similarly for (27). This yields the four equations

$$|G_{\text{spatial},(M+1,0)}| < 10^{-S}|G|\left(\frac{1}{8}\right), \tag{28}$$

$$|G_{\text{spatial},(0,N+1)}| < 10^{-S}|G|\left(\frac{1}{8}\right), \tag{29}$$

$$|G_{\text{spectral},(P+1,0)}| < 10^{-S}|G|\left(\frac{1}{8}\right), \tag{30}$$

$$|G_{\text{spectral},(0,Q+1)}| < 10^{-S}|G|\left(\frac{1}{8}\right). \tag{31}$$

First, consider the G_{spatial} series. Using the asymptotic approximation for the complementary error function for the (m,n) term, we have, with either $(m,n) = (M+1,0)$ or $(m,n) = (0,N+1)$,

$$\left|\left(\frac{1}{8\pi R_{mn}\sqrt{\pi}}\right)\left[\frac{e^{-\left((R_{mn}E)^2-\left(\frac{k}{2E}\right)^2\right)}2R_{mn}E}{(R_{mn}E)^2+\left(\frac{k}{2E}\right)^2}\right]\right| < 10^{-S}\left(\frac{1}{4\pi R_{00}}\right)\left(\frac{1}{8}\right). \tag{32}$$

Denote

$$x_3 = R_{mn}E \tag{33}$$

and

$$F = \frac{k}{2E}, \tag{34}$$

where F is a normalized frequency term. Then the above equation reduces to the form

$$\left|\frac{e^{-(x_3^2-F^2)}}{x_3^2+F^2}\right| < c_3, \tag{35}$$

where

$$c_3 = 10^{-S}\left(\frac{\sqrt{\pi}}{R_{00}E}\right)\left(\frac{1}{8}\right)\beta. \tag{36}$$

The factor of β has been introduced as an adjustment factor. Using $\beta = 1$ corresponds directly to the solution of (32) and usually represents a worst-case error bound. However, a factor of $\beta = 4$ was found to work well in almost all cases. Hence, based on this observation,

$$c_3 = 10^{-S}\left(\frac{\sqrt{\pi}}{R_{00}E}\right)\left(\frac{1}{2}\right). \tag{37}$$

Solving (35) for x_3 results in

$$x_3 = \sqrt{\Delta - F^2}, \tag{38}$$

where Δ satisfies

$$\frac{e^{-\Delta}}{\Delta} = W \tag{39}$$

with $W = c_3 e^{-2F^2}$ (the details are shown in the Appendix). The transcendental Eq. (39) can be solved using any standard method. An iterative method is discussed in the Appendix. We then have the criterion

$$R_{mn} > \frac{x_3}{E} \tag{40}$$

or

$$|m\mathbf{s}_1 + n\mathbf{s}_2 - \mathbf{r} + \mathbf{r}'| > \frac{x_3}{E}. \tag{41}$$

The summation limits are denoted as M and N, where $-M \leqslant m \leqslant M$ and $-N \leqslant n \leqslant N$. In determining the limits, we impose the condition that the convergence criterion be reached for both $m = M + 1$ and $n = 0$ and also for $m = 0$ and $n = N + 1$. The above equation can be solved for the general skewed case, although the result is unwieldy. Specializing to the rectangular case for simplicity ($\boldsymbol{s}_1 = \hat{\boldsymbol{x}}a$ and $\boldsymbol{s}_2 = \hat{\boldsymbol{y}}b$), we obtain the summation limits M and N as

$$M = \mathrm{Int}\left(\frac{x - x'}{a} + \frac{1}{a}\sqrt{\left(\frac{x_3}{E}\right)^2 - (y - y')^2 - (z - z')^2}\right) \tag{42}$$

and

$$N = \mathrm{Int}\left(\frac{y - y'}{b} + \frac{1}{b}\sqrt{\left(\frac{x_3}{E}\right)^2 - (x - x')^2 - (z - z')^2}\right), \tag{43}$$

where a and b are the lattice dimensions in the x and y directions, respectively, and $\mathrm{Int}(x)$ denotes the integer part of x.

Next consider the G_{spectral} series, with either $(p, q) = (P + 1, 0)$ or $(p, q) = (0, Q + 1)$. The asymptotic approximation of the complementary error function yields the result

$$\left|\left(\frac{1}{4\sqrt{\pi}A\alpha_{pq}}\right)\mathrm{e}^{-\left(\left(\frac{\gamma_{pq}}{2E}\right)^2 + (E\Delta z)^2\right)}\left[\frac{1}{\frac{\gamma_{pq}}{2E} - E\Delta z} + \frac{1}{\frac{\gamma_{pq}}{2E} + E\Delta z}\right]\right| < 10^{-S}\left(\frac{1}{4\pi R_{00}}\right)\left(\frac{1}{8}\right), \tag{44}$$

where

$$\alpha_{pq} = \mathrm{j}k_{zpq} \tag{45}$$

and

$$\Delta z = |z - z'|. \tag{46}$$

If we denote

$$x_4 = \frac{\alpha_{pq}}{2E} \tag{47}$$

and

$$Z_E = E\Delta z = E|z - z'|, \tag{48}$$

where Z_E is a normalized vertical displacement term, then the above equation reduces to the form

$$\left|\frac{\mathrm{e}^{-(x_4^2 + Z_E^2)}}{x_4^2 - Z_E^2}\right| < c_4, \tag{49}$$

where

$$c_4 = 10^{-S}\left(\frac{1}{\sqrt{\pi}}\right)\left(\frac{EA}{R_{00}}\right)\left(\frac{1}{8}\right)\beta. \tag{50}$$

As in the spatial case, the adjustment factor β is taken to be 4, and hence

$$c_4 = 10^{-S}\left(\frac{1}{\sqrt{\pi}}\right)\left(\frac{EA}{R_{00}}\right)\left(\frac{1}{2}\right). \tag{51}$$

Solving the above expression for x_4, we obtain the following (see the Appendix):

$$x_4 = \sqrt{\Delta + Z_E^2}, \tag{52}$$

where Δ satisfies (39) with $W = c_4 e^{2Z_E^2}$. We then require that

$$\alpha_{pq} > 2Ex_4 \tag{53}$$

or

$$|\mathbf{k}_{tpq}| > \sqrt{k^2 + (2Ex_4)^2}. \tag{54}$$

The transverse wavenumber is given by

$$\mathbf{k}_{tpq} = \left(\frac{2\pi}{A}\right)[-p(\hat{\mathbf{z}} \times \mathbf{s}_2) + q(\hat{\mathbf{z}} \times \mathbf{s}_1)] + \mathbf{k}_{t00}. \tag{55}$$

The summation limits are denoted as P and Q, where $-P \leqslant p \leqslant P$ and $-Q \leqslant q \leqslant Q$. In determining the limits, we impose that the convergence criterion be reached for both $p = P + 1$ and $q = 0$ and also for $p = 0$ and $q = Q + 1$. Solving for the summation limits P and Q for a rectangular lattice, we obtain

$$P = \mathrm{Int}\left(-\frac{k_{x0}a}{2\pi} + \frac{a}{2\pi}\sqrt{k^2 + (2Ex_4)^2 - k_{y0}^2}\right) \tag{56}$$

and

$$Q = \mathrm{Int}\left(-\frac{k_{y0}b}{2\pi} + \frac{b}{2\pi}\sqrt{k^2 + (2Ex_4)^2 - k_{x0}^2}\right). \tag{57}$$

One note regarding Eqs. (42), (43) and (56), (57) should be made in connection with the square roots. Depending on the geometry of the problem and the specified convergence accuracy, it may occur that the argument of one of the square roots is negative. This implies that the asymptotic analysis employed predicts that the necessary value of the corresponding summation limit is less than zero. The lower limit must always be greater than or equal to zero, however. Furthermore, since the analysis that predicts the summation limits is based on asymptotic approximations, occasionally the analysis predicts a limit that is smaller than is actually required. This most commonly occurs when the prediction is for a limit of zero (for example, see the discussion in connection with Table 5 below). To help avoid this problem, one strategy that can be implemented is to always choose the summation limits (M, N, P, Q) to be equal to or larger than 1.

5. Results

The Ewald method involves the use of the complementary error function for the calculation of G. For values of L greater than 6 in our study, the Ewald method suffered from unpredictable round-off errors due to the limitations on the accuracy of the complementary error function software used. Due to large magnitudes of the arguments used in the complementary error function, errors were obtained from this in addition to the errors resulting from cancellations. To avoid this problem, the loss of significant digits was limited to a maximum value of $L = 6$ in the study.

For all results, free-space conditions are assumed ($k = k_0$ and $\lambda = \lambda_0$). Results are shown for a square lattice ($a = b$) with $k_{x0} = k_{y0} = 0$. Table 1 illustrates that the values of $G_{00,\mathrm{spectral}}$ differ by many orders of magnitude using E_{opt} and E_L at different frequencies. For illustration, we take the number of significant digits lost to be $L = 3$. Also shown in the table for convenience is $G_{\mathrm{pure,spectral}}$, which is the numerically-exact Green's function calculated using a pure spectral method, namely

$$G_{\mathrm{pure,spect}}(\mathbf{r}, \mathbf{r}') = \frac{1}{A}\sum_{p=-\infty}^{\infty}\sum_{q=-\infty}^{\infty}\frac{1}{2\mathrm{j}k_{zpq}} e^{-\mathrm{j}\mathbf{k}_{tpq}\cdot(\boldsymbol{\rho}-\boldsymbol{\rho}')} e^{-\mathrm{j}k_{zpq}|z-z'|}. \tag{58}$$

Similarly, Table 2 illustrates that differences of many orders of magnitude exist between the values of $G_{00,\mathrm{spatial}}$ using E_{opt} and E_L at different frequencies. Once again we take the number of lost significant digits to be $L = 3$.

Table 1
$G_{00,\text{spectral}}$ obtained using E_{opt} and E_L, compared with $G_{\text{pure,spectral}}$, for a periodic cell of dimension $a = b = 0.5$ m with $x = y = x' = y' = 0$ and $|z - z'| = 0.05$ m

a/λ	E_{opt}	E_L	$G_{00,\text{spectral}}$ using E_{opt}	$G_{00,\text{spectral}}$ using E_L	$G_{\text{pure,spectral}}$
10	3.5449	19.3974	5.46E+138	54.678	2.8541
5	3.5449	9.6349	1.23E+032	225.50	1.7847
4	3.5449	7.5821	2.27E+019	364.90	3.9148
3	3.5449	5.6205	1.31E+010	659.64	3.0937
2	3.5449	3.6945	4025.72	1537.3	3.9978

Clearly, the value of E_L is limiting the size of the $(0,0)$ terms in each series, which is what results in more accurate results due to less round-off error.

Next, Table 3 shows the values of $G_{00,\text{spectral}}$ and $G_{00,\text{spatial}}$ using E_L for different values of L, keeping the frequency fixed. It can be seen that the largest of the $(0,0)$ terms, namely $G_{00,\text{spatial}}$, has a magnitude that is on the order of 10^L times as large as the total Green's function, as expected.

In Fig. 2, we consider the variation of E_L with respect to the free-space wavelength for various parameters L. We obtain a different curve of E_L vs. free-space wavelength λ for each L, and these curves are compared to the fixed value E_{opt}, also included in the figure as a horizontal line.

To test the validity of our approach, we next consider some cases where we compare the values of L obtained theoretically to the values of L obtained from the *actual* loss of significant figures, obtained numerically. $L_{\text{theoretical}}$ is the input value of L used in the calculation of E_L. The value L_{actual} is the value of L obtained by using the same value of E_L and comparing the value of the total Green's function G_{tot} obtained by the Ewald method with the pure-spectral Green's function $G_{\text{pure,spectral}}$ (assuming the pure spectral Green's function to be the accurate value, since this method, while slowly convergent, does not suffer from the same loss-of-significance problem that the Ewald method does). Since the accuracy of the complementary error function used in the program was always more than six significant figures [8], we deliberately contaminate the (m,n) terms of the spectral and spatial series in the sixth decimal place using a random complex noise, which ensures that the arithmetic is accurate to exactly six significant figures. We then obtain the value of L_{actual} using the formula

$$L_{\text{actual}} = 6 - \left|\log_{10}\left|\frac{G_{\text{pure,spectral}} - G_{\text{tot}}}{G_{\text{pure,spectral}}}\right|\right|. \tag{59}$$

Table 2
$G_{00,\text{spatial}}$ obtained using E_{opt} and E_L compared with $G_{\text{pure,spectral}}$, for a periodic cell of dimension $a = b = 0.5$ m with $x = y = x' = y' = 0$ and $|z - z'| = 0.05$ m

a/λ	E_{opt}	E_L	$G_{00,\text{spatial}}$ using E_{opt}	$G_{00,\text{spatial}}$ using E_L	$G_{\text{pure,spectral}}$
10	3.5449	19.3974	5.48E+138	1807.62	2.8541
5	3.5449	9.6349	1.24E+032	1866.28	1.7847
4	3.5449	7.5821	2.31E+019	1870.75	3.9148
3	3.5449	5.6205	1.36E+010	1869.78	3.0937
2	3.5449	3.6945	4437.39	1861.66	3.9978

Table 3
$G_{00,\text{spectral}}$ and $G_{00,\text{spatial}}$ for a periodic cell of dimensions $a = b = 0.5$ m with $x = y = x' = y' = 0$ and $|z - z'| = 0.05$ m, keeping the frequency fixed such that $\frac{a}{\lambda} = \frac{b}{\lambda} = 5$

L	E_{opt}	E_L	$G_{00,\text{spectral}}$ using E_L	$G_{00,\text{spatial}}$ using E_L	$G_{\text{pure,spectral}}$
1	3.5449	13.5192	1.14919	21.07691	1.7847
2	3.5449	11.0693	17.2311	196.7255	1.7847
3	3.5449	9.6348	225.499	1866.286	1.7847
4	3.5449	8.6632	2768.04	18053.94	1.7847
5	3.5449	7.9469	32707.6	176646.9	1.7847
6	3.5449	7.3894	376678.0	1739668.0	1.7847

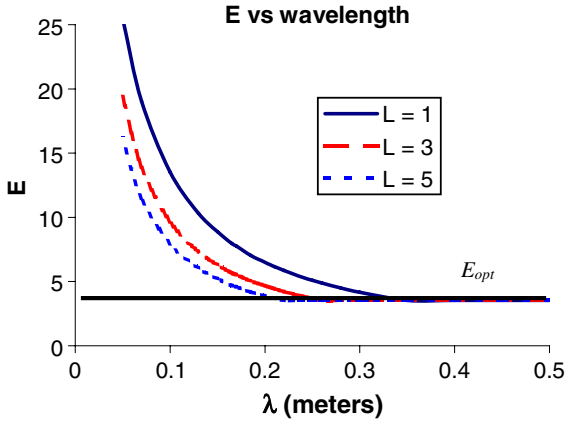

Fig. 2. E_L vs. free-space wavelength λ for $a = b = 0.5$ m, $k_{x0} = k_{y0} = 0$, $x' = y' = z' = 0$, $x = y = 0$ and $|z - z'| = 0.05$ m.

Fig. 3 shows a typical result. We have a square lattice with sides $a = 0.5$ m, $b = 0.5$ m. The frequency is fixed at 3 GHz such that $\frac{a}{\lambda} = \frac{b}{\lambda} = 5$. The phasing wavevectors k_{x0} and k_{y0} are taken to be zero while the position of the observation point is fixed at $x = 0$, $y = 0$ and $|z - z'| = 0.05$ m. Fig. 3 shows results for this case. It is seen that the agreement between the actual and theoretical values of L is quite good.

If the number of significant digits desired for convergence S_{spec} is specified, then we can calculate the summation limits for the spectral series, P_{cal} and Q_{cal} and the summation limits for the spatial series, M_{cal} and N_{cal}, using the formulae derived previously for the limits of P, Q, M and N. These four values are then verified by comparing their values with P_{act}, Q_{act}, M_{act} and N_{act} obtained from numerical convergence studies. The term S_{act} denotes the actual number of significant digits that the Ewald method has converged to, using the formula

$$S_{\text{act}} = -\log_{10}\left|\frac{G_{\text{pure,spectral}} - G_{\text{tot}}}{G_{\text{pure,spectral}}}\right|, \tag{60}$$

where $G_{\text{tot}} = G_{\text{spectral}} + G_{\text{spatial}}$ is the value obtained from the Ewald method after summing the two series using P_{cal}, Q_{cal}, M_{cal} and N_{cal}.

For the spectral and the spatial series, the adjustment factor $\beta = 1$ works in all cases but is excessively conservative as can be seen in Table 4. If we assume a factor of $\beta = 4$, as explained previously, we obtain Table 5. The second row ($S_{\text{spec}} = 2$) indicates that the actual number of significant figures obtained is less than that specified for this particular value of S_{spec}. The value $\beta = 4$ is used henceforth in the following results.

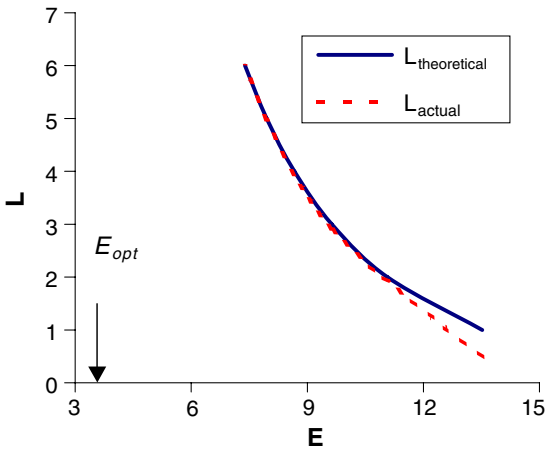

Fig. 3. A comparison between $L_{\text{theoretical}}$ and L_{actual} for different values of E for $a = b = 0.5$ m, $\frac{a}{\lambda} = \frac{b}{\lambda} = 5$, $k_{x0} = k_{y0} = 0$, $x' = y' = z' = 0.0$, $x = y = 0$, and $|z - z'| = 0.05$ m.

Table 4
S_{spec} and S_{act} for a periodic cell of dimensions $a = b = 0.5$ m, with $x = y = x' = y' = 0$ and $|z - z'| = 0.05$ m, keeping the frequency fixed such that $\frac{a}{\lambda} = \frac{b}{\lambda} = 0.5$, using $\beta = 1$

S_{spec}	P_{cal}, Q_{cal}	P_{act}, Q_{act}	M_{cal}, N_{cal}	M_{act}, N_{act}	S_{act}
1	0, 0	0, 0	0, 0	0, 0	**1.49**
2	1, 1	1, 1	1, 1	1, 1	**6.15**
3	1, 1	1, 1	1, 1	1, 1	**6.15**
4	1, 1	1, 1	1, 1	1, 1	**6.15**
5	1, 1	1, 1	1, 1	1, 1	**6.15**

Table 5
S_{spec} and S_{act} for a periodic cell of dimensions $a = b = 0.5$ m, with $x = y = x' = y' = 0$ and $|z - z'| = 0.05$ m, keeping the frequency fixed such that $\frac{a}{\lambda} = \frac{b}{\lambda} = 0.5$, using $\beta = 4$

S_{spec}	P_{cal}, Q_{cal}	P_{act}, Q_{act}	M_{cal}, N_{cal}	M_{act}, N_{act}	S_{act}
1	0, 0	0, 0	0, 0	0, 0	**1.49**
2	0, 0	1, 1	0, 0	1, 1	**1.49**
3	1, 1	1, 1	1, 1	1, 1	**6.15**
4	1, 1	1, 1	1, 1	1, 1	**6.15**
5	1, 1	1, 1	1, 1	1, 1	**6.15**

Table 6
S_{spec} and S_{act} for a periodic cell of dimensions $a = b = 0.5$ m, with $x = y = x' = y' = 0$ and $|z - z'| = 0.05$ m, keeping the frequency fixed such that $\frac{a}{\lambda} = \frac{b}{\lambda} = 5$, using $\beta = 4$

S_{spec}	P_{cal}, Q_{cal}	P_{act}, Q_{act}	M_{cal}, N_{cal}	M_{act}, N_{act}	S_{act}
1	5, 5	5, 5	0, 0	0, 0	**3.19**
2	5, 5	5, 5	0, 0	0, 0	**3.19**
3	5, 5	5, 5	0, 0	0, 0	**3.19**
4	6, 6	6, 6	0, 0	0, 0	**5.88**
5	6, 6	6, 6	0, 0	0, 0	**5.88**

Table 6 shows a case for a higher frequency such that $\frac{a}{\lambda} = \frac{b}{\lambda} = 5$. Again, the other dimensions are kept the same as in Table 5. The agreement between the actual and specified values of S is good, especially for larger values of S, with $S_{act} > S_{spec}$ in all cases.

The agreement between S_{spec} and S_{act} has also been tested for various other wavevectors, periodic cell sizes and horizontal and vertical positions of the observation point, and good agreement has been found.

6. Conclusion

The Ewald method is a very efficient method for calculating the periodic free-space Green's function, but the method suffers from accuracy problems at high frequency, as noted in [6], due to a loss of significant figures that occurs from a cancellation when adding the two series, spectral and spatial, that appear in the method. The method proposed here determines the "best" value of the parameter E that appears in the method in order to obtain the fastest convergence of the Ewald sum, while limiting the number of significant digits that are lost to a specified level L. In particular, the method determines the "best" value $E = E_L$ that yields the fastest convergence while limiting the size of the largest $(0,0)$ terms in the spatial and the spectral series relative to the numerical value of the total Green's function, so as to limit cancellation error. Although the overall convergence is not as fast as when using the "optimum" value $E = E_{opt}$, the loss of significant digits is kept to a tolerable level. Many orders of magnitude difference obtained between the values of $G_{00,spectral}$ and $G_{00,spatial}$ using E_{opt} and E_L at high frequencies is evident. For higher frequencies, $E_L > E_{opt}$. However $E_L = E_{opt}$ for frequencies below a certain threshold that depends on L. For a fixed frequency, the value of E_L increases as L decreases. The predicted loss of significant digits is verified through numerical simulations and the results illustrate the accuracy of the proposed formula.

Approximate expressions for the summation limits required to achieve a specified convergence accuracy for both the spectral as well as the spatial series have also been formulated and tested for different cases. The specified number of significant digits desired for convergence, S_{spec}, is compared with the actual number of significant digits that the series have converged to, S_{act}, and found to be in good agreement in almost all cases, thereby validating the formulas.

Appendix

Consider the transcendental equation

$$\frac{e^{-(x^2-K^2)}}{x^2+K^2} = c.$$

For the spatial case,

$$x = x_3, \quad K^2 = F^2, \quad \text{and} \quad c = c_3 = 10^{-S}\left(\frac{\sqrt{\pi}}{R_{00}E}\right)\left(\frac{1}{2}\right).$$

For the spectral case,

$$x = x_4, \quad K^2 = -Z_E^2 \quad \text{and} \quad c = c_4 = 10^{-S}\left(\frac{1}{\sqrt{\pi}}\right)\left(\frac{EA}{R_{00}}\right)\left(\frac{1}{2}\right).$$

The transcendental equation is rewritten as

$$e^{-x^2}e^{K^2} = cx^2 + cK^2.$$

Let $A = ce^{-K^2}$ and $B = ce^{-K^2}K^2$. Then the above equation reduces to

$$e^{-x^2} = Ax^2 + B.$$

Denoting $C = \frac{B}{A}$, we obtain

$$e^{-x^2} = A(x^2 + C).$$

Let

$$(x^2 + C) = \Delta.$$

Hence, $x^2 = \Delta - C$ and

$$e^{-\Delta}e^{C} = A\Delta$$

or

$$\frac{e^{-\Delta}}{\Delta} = W,$$

where $W = ce^{-2K^2}$.

Three cases are considered:

Case 1: $W \leqslant 0.36$.

Taking logarithms on both sides, we have

$$-\Delta - \ln\Delta = \ln W.$$

For the first iteration, we start with

$$\Delta^0 = -\ln W.$$

The iterative solution to this equation is

$$\Delta^{i+1} = -\ln\Delta^i - \ln W.$$

It has been found numerically that this iteration converges when $W \leqslant 0.36$.

Case 2: $W \geqslant 0.38$

For this case we use

$$\Delta = \frac{1}{W} e^{-\Delta}.$$

For the first iteration, we start with

$$\Delta^0 = \frac{1}{W}.$$

The iterative solution is then

$$\Delta^{i+1} = \frac{1}{W} e^{-\Delta^i}.$$

It has been found numerically that this iteration converges when $W \geqslant 0.38$.

Case 3: $0.36 < W < 0.38$.

In this "iterative-failure" region neither iterative method converges, However, by numerical solution, we obtain the approximate value of $\Delta = 0.997$.

As W approaches the limiting values of 0.36 or 0.38, the number of iterations needed increases significantly. As a remedy, the solution of (39) for values within a particular region surrounding the iterative-failure region, e.g., $0.3 < W < 0.5$, may be computed and stored in a look-up table.

References

[1] R.E. Jorgenson, R. Mittra, Efficient calculation of the free-space periodic Green's function, IEEE Trans. Antennas Propagat. 38 (May) (1990) 633–642.
[2] S. Singh, W.F. Richards, J.R. Zineckar, D.R. Wilton, Accelerating the convergence of series representing the free space periodic Green's function, IEEE Trans. Antennas Propagat. 38 (1990) 1958–1962.
[3] R.M. Shubair, Y.L. Chow, Efficient computation of the periodic Green's function in layered dielectric media, IEEE Trans. Microwave Theory Techn. 41 (3) (1993) 498–502.
[4] P.P. Ewald, Die berechnung optischer und electrostatischer gitterpotentiale, Annal. Phys. 64 (1921) 253–287.
[5] K.E. Jordan, G.R. Richter, P. Sheng, An efficient numerical evaluation of the Green's function for the Helmholtz operator on periodic structures, J. Comput. Phys. 63 (1986) 222–235.
[6] A. Kustepeli, A.Q. Martin, On the splitting parameter in the Ewald method, Microwave Guided Wave Lett., IEEE 10 (May) (2000) 168–170.
[7] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables, Dover, NY, 1965.
[8] S. Zhang, J. Jin, Computation of Special Functions, 1996, pp. 620–624 (Chapter 16).