Choosing splitting parameters and summation limits in the numerical evaluation of 1-D and 2-D periodic Green’s functions using the Ewald method

Ferhat T. Celepcikay, Donald R. Wilton, David R. Jackson, and Filippo Capolino

Received 26 December 2007; revised 14 April 2008; accepted 5 May 2008; published 30 September 2008.

[1] Accurate and efficient computation of periodic free-space Green’s functions using the Ewald method is considered for three cases: a 1-D array of line sources, a 1-D array of point sources, and a 2-D array of point sources. A limitation on the numerical accuracy when using the “optimum” $E$ parameter (which gives optimum asymptotic convergence) at high frequency is discussed. A “best” $E$ parameter is then derived to overcome these limitations. This choice allows for the fastest convergence while maintaining a specific level of accuracy (loss of significant figures) in the final result. Formulas for the number of terms needed for convergence are also derived for both the spectral and the spatial series that appear in the Ewald method, and these are found to be accurate in almost all cases.

Citation: Celepcikay, F. T., D. R. Wilton, D. R. Jackson, and F. Capolino (2008), Choosing splitting parameters and summation limits in the numerical evaluation of 1-D and 2-D periodic Green’s functions using the Ewald method, Radio Sci., 43, RS6S01, doi:10.1029/2007RS003820.

1. Introduction

[2] In applying numerical full wave methods like the Method of Moments (MoM) or Boundary Integral Equations (BIE) to periodic structures involving conducting or dielectric electromagnetic scatterers, fast and accurate means for evaluating the free-space periodic Green’s function (FSPGF) are often needed. This type of Green’s function arises in a wide variety of applications, ranging from microwaves to optics, to the study of metamaterials and nanostructures. The Ewald method is one of the fastest methods for calculating 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 is used [Jordan et al., 1986], denoted here as $E_{opt}$. (For some applications, such as when using a periodic Green’s function in a MoM solution with full-domain basis functions, one may not wish to have a balanced convergence between the two series, as explained in Mathis and Peterson [1996] and Mathis and Peterson [1998]. However, when using subdomain basis functions and performing the integrations over the basis and testing functions in the spatial domain, the objective is to minimize the computer time of the periodic Green’s function, and this is accomplished by using $E_{opt}$. In this case the singular integrals involved in evaluating the matrix elements can be handled by specially designed numerical quadrature rules [Khayat and Wilton, 2005].)

[3] However, the numerical accuracy of the Ewald method degrades very quickly [Kustepeli and Martin, 2000] with increasing frequency (i.e., when the periodicity becomes large relative to a wavelength). This is due to a catastrophic loss of significant figures in combining the contributions of the two series, wherein the leading terms (and to a lesser extent, other nearby terms) in each series become very large but nearly equal and opposite in sign.

[4] The method proposed and 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 splitting parameter $E$ to avoid undue loss of accuracy. Increasing the $E$ parameter limits the size of the largest terms in both series at the expense of decreasing the convergence rate. 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 value $E_L$ of the Ewald splitting parameter is then obtained based on the number of significant figures $L$ that may be lost. 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 value $L$.

[5] A preliminary and intuitive analysis for the “best” choice of the Ewald splitting parameter was performed in Capolino et al. [2005] for the case of 1-D array of line sources, and in Capolino et al. [2007] for the case of a 1-D array of point sources. The case of a 2D-array of point sources has been treated in detail in Oroskar et al. [2006]. Here we extend the analysis of the last paper to the two other cases, and provide a unified formalism for the choice of the Ewald splitting parameter and the summation limits that is valid for all three cases.

2. Spectral, Spatial, and Ewald Green’s Function Representations

[6] Three different FSPGF cases are considered: a 1-D array of line sources with interelement period $d$, a 2-D array of point sources on a general skewed lattice, and a 1-D array of point sources with interelement spacing $d$. The geometries are depicted along with relevant coordinate systems and geometrical definitions in Figures 1, 2, and 3, respectively. An interelement phase shift along the direction(s) is assumed. Both spatial and spectral representations of each Green’s function exist in the general form

$$G(r, r') = \begin{cases} \sum_{m=-\infty}^{\infty} G_m = \sum_{p=-\infty}^{\infty} \tilde{G}_p, & \text{1-D array} \\ \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G_{m,n} = \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \tilde{G}_{p,q}, & \text{2-D array} \end{cases}$$

where the terms of the spatial representations are

$$G_m = e^{-jk_{0a}md} \frac{1}{4j} H_0^{(2)}(kr_m), \quad 1 \text{--D array of line sources}$$

$$G_m = e^{-jk_{0a}md} e^{-jk R_m}, \quad 1 \text{--D array of point sources}$$

$$G_{m,n} = e^{-jk_{0a}ns} e^{-j R_{mn}}, \quad 2 \text{--D array of point sources}$$

whereas those of the spectral representations are

$$\tilde{G}_p = \frac{1}{d} e^{-jk_{0p}(x-x')} e^{-jk_{0p} \Delta z}, \quad 1 \text{--D array of line sources}$$

$$\tilde{G}_p = \frac{1}{4jd} e^{-jk_{0p}(x-x')} H_0^{(2)}(kr_p \rho), \quad 1 \text{--D array of point sources}$$

$$\tilde{G}_{p,q} = \frac{1}{A} e^{-jk_{pq}} \cdot (\rho - \rho') e^{-jk_{pq} \Delta z}, \quad 2 \text{--D array of point sources}$$

where $\Delta z \equiv |z - z'|$ and an $e^{i \omega t}$ time dependence is assumed and suppressed. In the above, $(x, y, z)$ and $(x', y', z')$ denote the observation and reference-element source points, respectively. For the 1-D cases, the periodicity is along $x$ with a period of $d$ (see Figures 1 and 2), and

$$k_{xp} = k_{x0} + 2\pi p \frac{d}{d}, \quad k_{xp} = k_{xp} = \sqrt{k^2 - k_{xp}^2}$$

$$R_m = \sqrt{(z - z')^2 + (x - x' - md)^2} \quad (1 \text{--D array of line sources}),$$

$$R_m = \sqrt{(z - z')^2 + (y - y')^2 + (x - x' - md)^2} \quad (1 \text{--D array of point sources}),$$

$$\rho = \sqrt{(y - y')^2 + (z - z')^2} \quad (1 \text{--D array of point sources}).$$

For the 2-D array with lattice vectors $s_1, s_2$ (see Figure 3), the parameters are defined as

$$\rho = x\hat{x} + y\hat{y}, \quad \rho' = x'\hat{x} + y'\hat{y}, \quad \rho_{mn} = ms_1 + ns_2,$$

$$R_{mn} = \sqrt{(z - z')^2 + |\rho - \rho' - \rho_{mn}|^2};$$

$$k_{0a} = k_{x0} \hat{x} + k_{y0} \hat{y},$$

$$k_{pq} = k_{0a} + \frac{2p\pi(s_2 \times \hat{z})}{A} + \frac{2q\pi(z \times s_1)}{A},$$

$$k_{pq} = \sqrt{k^2 - k_{pq}^2} \cdot k_{pq}, \quad A = \hat{z} \cdot (s_1 \times s_2).$$

The transverse phasing wave vector $k_{0a} = \hat{x} k \sin \theta_0 \cos \phi_0 + \hat{y} k \sin \theta_0 \sin \phi_0$ defines the interelement phasing for the 2-D
array in terms of the propagation angles $\theta_0$, $\phi_0$ of the (0, 0) Floquet mode. For 1-D arrays, this quantity becomes the scalar phasing $k_x = k \cos \theta_0$, where $\theta_0$ is measured with respect to the $x$-axis. Physically the FSPGF is the time-harmonic scalar potential produced by the associated array of phased sources.

$[7]$ For a lossless medium the square root for the wave numbers $k_{rp}$ and $k_{zp}$ is that which gives a positive real number or a negative imaginary number, corresponding to waves that propagate outward or decay away from the sources, respectively. For a lossy medium the wave numbers are complex and it suffices to require the imaginary part of the wave numbers to be negative. Henceforth, it will be assumed that the medium is lossless, but the formulas can be extended to the lossy case.

$[8]$ When employing the Ewald method for the evaluation of the FSPGF, the Green’s function is expressed as a sum of two series $[Ewald, 1921]$ of the form

$$ G(r, r') = \begin{cases} 
\sum_{m=-\infty}^{\infty} G^E_m \sum_{n=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} G^E_{p,q} & \text{1 - D array} \\
\sum_{m=-\infty}^{\infty} G^E_{m,n} + \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} G^E_{p,q} & \text{2 - D array}.
\end{cases} $$

(4)

The terms that appear in the three different cases may be found in Capolino et al. [2005], Capolino et al. [2007], and Oroskar et al. [2006], for the 1-D array of line sources, the 1-D array of point sources, and the 2-D array of point sources, respectively. Summarizing, the terms of the modified spatial representations are

$$ G^E_m = \frac{1}{4\pi} e^{-jk_0mz} \left[ \sum_{q=0}^{\infty} \left( \frac{k}{2E} \right)^q E_{q+1}(R_m E^2) \right] $$

1 - D array of line sources,

$$ G^E_m = \frac{1}{2} \frac{e^{-jk_0mz}}{4\pi R_m} \left[ e^{-jkRe} \text{erfc} \left( R_mE - j \frac{k}{2E} \right) + e^{jkRe} \text{erfc} \left( R_mE + j \frac{k}{2E} \right) \right] $$

1 - D array of point sources,

$$ G^E_{m,n} = \frac{1}{2} \frac{e^{-jk_0mnR}}{4\pi R_{mn}} \left[ e^{-jkRn} \text{erfc} \left( R_{mn}E - j \frac{k}{2E} \right) + e^{jkRn} \text{erfc} \left( R_{mn}E + j \frac{k}{2E} \right) \right] $$

2 - D array of point sources,

(5)

where $\text{erfc}(z)$ is the complementary error function and $E_q(z)$ denotes the exponential integral function of order
Table 1. Definition of Parameters Appearing in the Iterative Equation for Determining $E_L$

|          | $\zeta$ | $K$ | $c_1$ | $F(\zeta)$ |
|----------|---------|-----|-------|-----------|
| 1-D array of line sources | $k$ | $kR_0$ | $10^44\pi|G^{est}|$ | $c_1\zeta^2$ |
| Spectral | $k_0$ | $k_0\Delta x$ | $10^42\sqrt{\pi}k_0\Delta x|G^{est}|$ | $\zeta(\zeta^2 + \frac{k_0^2}{\zeta})$ |
| 1-D array of point sources | $k$ | $kR_0$ | $10^44\pi^2 R_0|G^{est}|$ | $c_1(\frac{k}{k_0})(\zeta^2 + \frac{k_0^2}{\zeta})$ |
| Spectral | $k_0$ | $k_0\rho_0^2$ | $10^44\pi^2|G^{est}|$ | $c_1\zeta^2$ |
| 2-D array of point sources | $k$ | $kR_{00}$ | $10^44\pi^2 R_{00}|G^{est}|$ | $c_1(\frac{k}{k_0})(\zeta^2 + \frac{k_0^2}{\zeta})$ |
| Spectral | $k_{00}$ | $k_{00}\Delta x$ | $10^42\sqrt{\pi}k_{000}|G^{est}|$ | $\zeta(\zeta^2 + \frac{k_{00}^2}{\zeta})$ |

$q$. The terms of the modified spectral representations are

$$\tilde{G}_p^{E} = \frac{1}{2d} e^{jk_{x0}(x-x')} \left[ e^{-jk_{x0}\Delta x} \text{erfc} \left( \frac{jk_{x0}}{2E} - \Delta \zeta \right) + e^{jk_{x0}\Delta x} \text{erfc} \left( \frac{jk_{x0}}{2E} + \Delta \zeta \right) \right]$$

1 - D array of line sources,

$$\tilde{G}_p^{E} = \frac{1}{4\pi d} e^{-jk_{x0}(x-x')} \left[ \sum_{q=0}^{\infty} \left( \rho E \right)^2 \frac{(-1)^q}{q!} E_{q+1} \left( \frac{-k_{x0}^2}{4E^2} \right) \right]$$

1 - D array of point sources,

$$\tilde{G}_p^{E} = \frac{1}{2A} e^{-jk_{x0}(\rho - \rho')} \left[ e^{-jk_{x0}\Delta x} \text{erfc} \left( \frac{jk_{x0}}{2E} - \Delta \zeta \right) + e^{jk_{x0}\Delta x} \text{erfc} \left( \frac{jk_{x0}}{2E} + \Delta \zeta \right) \right]$$

2 - D array of point sources. (6)

For the 1-D array of point sources, the exponential integral function that appears may have a negative argument, depending on the frequency. In this case the argument is interpreted as being infinitesimally above the branch cut of the exponential integral function on the negative real axis, corresponding to an infinitesimal amount of loss.

3. Splitting Parameter ($E$)

In equations (5) and (6) the spatial and spectral series both involve a “splitting” parameter $E$. The “optimum value” $E_{opt}$ for the splitting parameter [Jordan et al., 1986] balances the asymptotic rate of convergence of the spatial and spectral series, and consequently minimizes the overall number of terms needed to calculate the total FSPGE. The optimum value is found to be

$$E_{opt} = \begin{cases} \frac{\sqrt{\pi}}{d}, & 1 - \text{D array} \\ \frac{\sqrt{\pi}}{A}, & 2 - \text{D array}. \end{cases}$$

(7)

However, numerical difficulties are encountered when the lattice separations (periods) become large relative to a wavelength. This was first discovered in Kustepeli and Martin [2000], and subsequently also discussed in Capolino et al. [2005], Oroskar et al. [2006], and Capolino et al. [2007]. This happens because, for large arguments, both the complementary error function and the exponential integrals contribute terms of the form $\text{exp}[\frac{1}{(2E)^2}]$ that produce extremely large initial terms (and to a lesser extent, large nearby terms) in both the spatial and the spectral series. The resulting series then each converge to very large, but nearly equal and oppositely signed values, resulting in a total sum of moderate value but with a catastrophic loss of significant figures when the two series are combined. Exponential overflow is another potential concern as well, due to the large initial values in each series.

[10] To circumvent the problem, it is desirable to limit the size of the largest terms of each series by choosing an $E$ value larger than the “optimum” value, which reduces the maximum values of both the complementary error function and the exponential integrals. As a result, one avoids loss of accuracy in adding the two series, and a more accurate result for the total Green’s function is obtained [Kustepeli, and Martin, 2000] at the expense of slower convergence.
In the following sections, a recipe for finding the best choice for \( E \), called \( E_L \), that achieves the fastest convergence under the constraint of limiting the loss of significant figures to \( L \) digits, is obtained for a general source and observation point, for all three cases.

4. Choice of the Splitting Parameter

The strategy is to limit the size of the largest terms relative to the value of the total Green’s function, with the largest terms in each series being the initial (0) or (0, 0) terms in 1-D or 2-D, respectively. The value of \( E = E_L \) is obtained by enforcing the following conditions:

\[
\begin{aligned}
|G^E_0|, |G^E_0,0|, 1 - \text{D array} &< \alpha 10^L |G^{est}| \\
|G^E_{0,0}|, |G^E_{0,0,0}|, 2 - \text{D array} &
\end{aligned}
\]

(8)

where \( G^{est} \) is a closed-form estimate of the FSPGF, and the integer parameter \( L \) indicates (roughly) the number of significant figures one is willing to sacrifice in the calculation. For a strict bound, the factor \( \alpha \) in (8) should be chosen as 1/2 to ensure that each of the two initial terms in (8) (one from the spatial series and one from the spectral series) contributes no more than half the total error limit. However, it is found in each of the three cases that one of the initial terms (either the spatial one or the spectral one, depending on the case) is significantly larger than the other, and a factor of \( \alpha = 1 \) therefore becomes more appropriate, and this is adopted here. It suffices to use a rough estimate of the Green’s function on the right-hand side of (8), which may be obtained by examining the most nearly singular terms from both the spatial and the spectral series. (This is an improvement over the previous derivations, as in Oroskar et al. [2006], where only the spatial term was used.) This yields the following magnitude estimate for the overall Green’s function:

\[
|G^{est}| \approx \begin{cases} 
\max \left( \frac{|H_0^{(2)}(kR_m)|}{4}, \frac{1}{2|k_{zp}|d} \right), & 1 - \text{D array of line sources} \\
\max \left( \frac{1}{4\pi R_m}, \frac{|H_0^{(2)}(k_{zp})|}{4d} \right), & 1 - \text{D array of point sources} \\
\max \left( \frac{1}{4\pi R_{mn}}, \frac{1}{2d|k_{zp}|} \right), & 2 - \text{D array of point sources} 
\end{cases}
\]

(9)

In (9), the indices \( p, q, m, n \) that produce the smallest values for \( R_m, R_{mn}, k_{zp}, k_{zp}, \) and \( k_{zp} \) must be determined; although this may be done analytically, it is also very easy to find these values from a simple numerical search.

To proceed with the derivation of \( E_L \), we replace the terms on the left hand side of (8) by their high-frequency asymptotic estimates. The complementary error function terms are estimated by using the asymptotic relation

\[
\text{erfc}(z) \sim \frac{e^{-z^2}}{\sqrt{\pi z}},
\]

valid for large arguments. In addition, in (5) and (6) series of exponential integral functions appear. The simplest way to asymptotically evaluate these series is to re-cast them back into their integral forms. The integral forms are equation (15) in Capolino et al. [2007] and equation (11) in Capolino et al. [2005], which are reproduced below:

\[
\int_{1/(2E)^2}^{\infty} \frac{e^{-\frac{u^2}{2} + (k_{zp})^2 u}}{u} du = \sum_{q=0}^{\infty} \left( \frac{(-1)^q}{q!} \right) (\rho E)^{2q} E_{q+1} \left( \frac{-k_{zp}^2}{4E^2} \right)
\]

(11)

\[
\int_{E}^{\infty} \frac{e^{-R_s^2 + \frac{s^2}{2}} ds}{s} = \frac{1}{2} \sum_{q=0}^{\infty} \left( \frac{k}{2E} \right)^{2q} \frac{1}{q!} E_{q+1} (R_s^2 E^2)
\]

(12)

The integrals that appear in the above identities may be asymptotically evaluated for \( k \to \infty \) via integration by
The resulting value is the smallest value of $E$ that ensures that the largest term in both the spectral and the spatial series (see equation (8)) is limited in magnitude to avoid losing more than $L$ significant figures when the two series are combined. At low or moderate frequency, where the initial terms of the spectral and spatial series are not large, $E_L = E_{opt}$ since $E_{opt}$ will be the largest of the three terms in (17).

5. Number of Terms Needed for Convergence

[15] Having determined the “best” value of the $E$ parameter $E_L$ as a function of frequency, the next goal is to determine how many terms should be summed in each series (modified spatial and modified spectral) in (4) to achieve convergence. (One could, of course, check convergence as the series are summed, but we eventually want to select between the Ewald and alternative methods for computing the FSPGF that may be more efficient in some situations; for that an a priori estimate of the number of series terms and their relative computational cost is required.) Recall that a given value of $L$, the number of significant figures sacrificed in the calculation, has been assumed. For the resulting value of $E = E_L$ we must then 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 series index limits $\pm P$, $\pm Q$, $\pm M$, and $\pm N$ for the series indices $p$, $q$, $m$, and $n$, respectively ($Q$ and $N$ only apply for the 2-D geometry). If the accuracy of the arithmetic is limited to $T$ significant figures due to the machine precision (or, more likely in practice, limited by the accuracy of the complementary error function and exponential integrals), the value of $S$ specified should be limited to $S < T - L$ in order to avoid unnecessary computation.

[16] Owing to the Gaussian convergence, a rough estimate of the truncation error for both the spectral and spatial series is obtained by using the sum of the largest of the first neglected terms along each principal sum index in the series. Limiting the convergence error in each series to one half of the total, we thus require that

$$\begin{align*}
|G_{M+1}^E| &+ |G_{M-1}^E|, \\
|G_{M+1}^E| &+ |G_{M-1}^E| + |G_{0,N+1}^E| + |G_{0,-N-1}^E|, \\
|G_{P+1}^E| &+ |G_{P-1}^E|, \\
|G_{P+1}^E| &+ |G_{P-1}^E| + |G_{0,Q+1}^E| + |G_{0,-Q-1}^E|, \\
\end{align*}$$

$$\begin{align*}
1 - D \text{ array} &< 10^{-5}|G^{opt}| \left(\frac{1}{2}\right)^\beta \\
2 - D \text{ array} &< 10^{-5}|G^{opt}| \left(\frac{1}{2}\right)^\beta.
\end{align*}$$

where $\beta$ is a constant.
Table 2. Definition of Parameters Appearing in the Summation Limit Equation

| Array Type          | Type     | Parameter | Estimation |
|---------------------|----------|-----------|------------|
| 1-D array of line sources | Spatial  | $R_{ax,E}$ | 0          |
|                     | Spectral | $\frac{|k_{sp}|}{2E}$ | $jE\Delta z$ |
| 1-D array of point sources | Spatial  | $R_{ax,E}$ | $\frac{k}{2E}$ |
|                     | Spectral | $\frac{|k_{sp}|}{2E}$ | 0          |
| 2-D array of point sources | Spatial  | $R_{ax,E}$ | $\frac{k}{2E}$ |
|                     | Spectral | $\frac{|k_{sp}|}{2E}$ | $jE\Delta z$ |

The error in stopping the summations is approximated in the above equations as the sum of the two (four) values that give the largest contributions to the summed values for 1-D (2-D) arrays. The factor $\beta$ is introduced to allow for an empirical adjustment of the summation limits. Using $\beta = 1$ corresponds to a strict error bound based on the assumed asymptotic approximations, and therefore usually represents a worst-case error bound. However, factors of $\beta = 2$ and $\beta = 4$ have been found to work well for the 1-D and 2-D cases, respectively. If we assume that the contributions of the positively indexed terms on the LHS of inequalities (18) and (19) are dominant, the choice of the limits amounts to choosing $M, N, P, Q$ such that

$$
|G_{M+1}^{E}|, |G_{P+1}^{E}| < 10^{-5}|G| \left(\frac{1}{4}\right) \beta,
$$

1-D array

$$
|G_{M+1,0}^{E}|, |\hat{G}_{P+1,0}^{E}|, |\hat{G}_{0,Q+1}^{E}| < 10^{-5}|G| \left(\frac{1}{8}\right) \beta,
$$

2-D array

The dominance of the positively indexed terms may be assumed without loss of generality, by placing absolute values on the spatial displacements and phasing wave numbers, as seen in (22)–(25). Using the asymptotic estimates of Table A1, case 1b and case 2b, which assume large values of $M, N$ and $P, Q$ for the terms in (20) above, yields a transcendental equation of the form

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

where $\Delta = \xi^2 + c_2^2$, and the constants $c_2$ and $W$ in (21) are defined in Table 2.

Equation (21) must be solved numerically for $\xi$, which, as seen in Table 2, is the only term involving the indices. Equation (21) may be efficiently solved iteratively for the parameter $\Delta$ as outlined in Oroskar et al. [2006].

Once $\Delta$ and thus $\xi$ are determined, the indexed quantities defined in the $\xi$ column of Table 2 may be used to determine the index limits $m = M + 1, p = P + 1$ for 1-D arrays, and the index pairs $(m, n) = (M + 1, 0), (0, N + 1)$ and $(p, q) = (P + 1, 0), (0, Q + 1)$ for the 2-D array.

For the 1-D point source array the results are

$$
P = \text{Int} \left( \frac{k_{0,0}}{2\pi} \frac{d}{2\pi} \sqrt{\frac{k^2 + (2E\xi)^2}{\Delta}} \right).
$$

For the 2-D point source array the result is (restricting the result to the case of the rectangular lattice $s_1 = a\hat{x}, s_2 = b\hat{y}$ for simplicity):

$$
P = \text{Int} \left( \frac{k_{0,0}}{2\pi} \frac{a}{2\pi} \sqrt{\frac{k^2 + (2E\xi)^2 - k_{0,0}^2}{\Delta}} \right),
$$

$Q = \text{Int} \left( \frac{k_{0,0}}{2\pi} \frac{b}{2\pi} \sqrt{\frac{k^2 + (2E\xi)^2 - k_{0,0}^2}{\Delta}} \right)$. 

[19] The results for the case of a 1-D line source array are

$$
M = \text{Int} \left( \frac{x - x'}{d} + \frac{1}{d} \sqrt{\frac{\xi^2}{\Delta} - (z - z')^2} \right),
$$

$$
P = \text{Int} \left( \frac{k_{0,0}}{2\pi} \frac{d}{2\pi} \frac{1}{\sqrt{k^2 + (2E\xi)^2}} \right).
$$

(22)
One-Dimensional Line Source Array: $G_0^L$ and $\tilde{G}_0^L$ Obtained Using $E_{opt}$ and $E_L$, Compared With $G$, the Exact Value of the Green’s Function*  

| $d/\lambda_0$ | $E_L$ | $G_0^L$ Using $E_{opt}$ | $G_0^L$ Using $E_L$ | $G_0^L$ Using $E_{opt}$ | $G_0^L$ Using $E_L$ | $G$ Exact |
|-------------|-------|-----------------------|--------------------|-----------------------|--------------------|------------|
| 10.5        | 20.58585 | 5.8979E+146           | 12.427708          | 5.9065E+146          | 75.871153        | 4.802E-002 |
| 5.5         | 11.04567 | 1.5269E+038           | 34.849919          | 1.5351E+038          | 115.39839        | 0.1477325  |
| 4.5         | 9.038645  | 5.193E+024            | 47.653662          | 5.2355E+024          | 129.29622        | 0.1585821  |
| 3.5         | 7.013830  | 1.493E+14             | 70.176014          | 1.063E+014           | 147.89049        | 0.1619304  |
| 2.5         | 4.981864  | 1.356E+06             | 116.96534          | 1.395E+06            | 175.14480        | 0.1584406  |

*Results are shown for various frequencies (period relative to a wavelength) with $L = 3$.

\[
M = \text{Int} \left( \frac{|x-x'|}{a} + \frac{1}{a} \sqrt{\left( \frac{x}{E} \right)^2 - (y-y')^2 - (z-z')^2} \right),
\]

\[
N = \text{Int} \left( \frac{|y-y'|}{b} + \frac{1}{b} \sqrt{\left( \frac{y}{E} \right)^2 - (x-x')^2 - (z-z')^2} \right),
\]

(25)

In these final results, noninteger solutions of (21) for the summation limits should be rounded up to the next larger integer to obtain $M + 1$, etc., or equivalently, rounded down to obtain $M$, etc., as assumed in (22)–(25), in which the Int function truncates to the next lower integer.

[20] One note regarding (22)–(25) 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, yielding a complex value for the summation limit. This occurs because the asymptotic approximation used to estimate term magnitudes becomes invalid when the series actually needs only a few terms to converge, corresponding to a summation limit of zero or one. The problem is circumvented by always using a summation limit that is at least unity.

6. Results

[21] In this section results are presented for the three cases: 1-D line sources, 1-D point sources, and 2-D point sources. For all results, free-space conditions are assumed ($k = k_0$ and $\lambda = \lambda_0$). For the 1-D cases, $d = 0.5$ m. The 2-D results are shown for a square lattice ($a = b = 0.5$ m). As a result, the optimum splitting parameter is $E_{opt} = 3.5449$ for all three cases. A zero progressive phase shift is assumed (all source elements are in phase). The reference source element is located at the origin and the observation point is located at $(0, 0, \Delta z)$ with the vertical distance from the source plane set at $\Delta = 0.05$ m. For brevity’s sake, only the magnitude of the Green’s function terms are shown in the results.

[22] For each case, three tables are shown. For the 1-D array of line sources, Tables 3a–3c are shown; for the 1-D array of point sources, Tables 4a–4c are shown; and for the 2-D array of point sources, Tables 5a–5c are shown. The first tables in each case (part (a)) illustrate that the values of $G_0^L$ and $\tilde{G}_0^L$ become enormous at high frequency. For these tables the number of lost significant digits is set to $L = 3$. (The calculations of the special functions were performed with sufficient accuracy to ensure that $T > S + L$ in all cases.) The value of $E_L$ limits the size of the largest of the $(0, 0)$ terms, namely $G_{0,0}^L$, to a value on the order of $10^3$ times the exact Green’s function $G$ (denoted “$G$ Exact” in the tables), as expected. The numerically exact Green’s function has been calculated using a pure spectral method, shown in (3).

[23] The second tables shown for each case (part (b)) show the values of $G_{0,0}^L$ and $\tilde{G}_{0,0}^L$ using $E_L$ obtained for

### Table 3a. One-Dimensional Line Source Array: $G_0^L$ and $\tilde{G}_0^L$ for Various Values of $L$, Keeping the Frequency Fixed at $d = 5.5\lambda_0$  

| $L$ | $E_{opt}$ | $E_L$ | $G_0^L$ Using $E_{opt}$ | $G_0^L$ Using $E_L$ | $G_{opt}$ Using $E_{opt}$ | $G_{opt}$ Using $E_L$ | $G$ Exact |
|-----|-----------|-------|------------------------|---------------------|-------------------------|----------------------|------------|
| 1   | 3.544906  | 15.795788 | 1.0881778              | 0.2137888           | 0.147732                | 0.147732             | 2.01       |
| 2   | 3.544906  | 12.793900 | 11.564017              | 2.9467840           | 0.147732                | 0.147732             | 4.32       |
| 3   | 3.544906  | 11.045671 | 115.39839              | 34.849919           | 0.147732                | 0.147732             | 6.60       |
| 4   | 3.544906  | 9.8719462 | 114.25974              | 391.50175           | 0.147732                | 0.147732             | 6.60       |
| 5   | 3.544906  | 9.0139223 | 133.30598              | 428.78952           | 0.147732                | 0.147732             | 6.60       |
| 6   | 3.544906  | 8.3508864 | 125314.12              | 46237.312           | 0.147732                | 0.147732             | 6.60       |

*Also shown is $S_{act}$ the actual number of significant figures achieved when using the calculated summation limits.
Table 4a. One-Dimensional Point Source Array: $G_E^L$ and $G_0^L$ Obtained Using $E_{opt}$ and $E_L$, Compared With $G$, the Exact Value of the Green’s Function\textsuperscript{a}

\begin{tabular}{cccccccc}
\hline
$d/\lambda_0$ & $E_L$ & $G_0^L$ & $G_0^L$ Using $E_{opt}$ & $G_0^L$ Using $E_L$ & $G_0^L$ Using $E_{opt}$ & $G_0^L$ Using $E_L$ & $G$ Exact \\
\hline
10.5 & 19.93035 & 1.181E+147 & 304.98277 & 1.1830E+147 & 1803.7876 & 1.3718050 \\
5.5 & 10.51319 & 3.070E+038 & 592.50339 & 3.0869E+038 & 1863.0479 & 1.8099252 \\
4.5 & 8.556454 & 1.047E+025 & 730.06634 & 1.05567E+025 & 1869.2320 & 1.7899326 \\
3.5 & 6.595795 & 2.127E+013 & 948.12325 & 2.1571E+013 & 1870.9950 & 1.7072650 \\
2.5 & 4.645090 & 2.791E+006 & 1344.7252 & 2.87365E+006 & 1866.7076 & 1.5862856 \\
\hline
\end{tabular}

\textsuperscript{a}Results are shown for various frequencies (period relative to a wavelength) with $L = 3$.

Table 4b. One-Dimensional Point Source Array: $G_0^L$ and $G_0^L$ for Various Values of $L$, Keeping the Frequency Fixed at $d = 5.5\lambda_0$

\begin{tabular}{cccccc}
\hline
$L$ & $E_{opt}$ & $E_L$ & $G_0^L$ Using $E_{opt}$ & $G_0^L$ Using $E_L$ & $G$ Exact \\
\hline
1 & 3.544906 & 14.70097 & 20.70892 & 4.545456 & 1.8099522 \\
2 & 3.544906 & 12.07276 & 195.6083 & 53.04047 & 1.8099522 \\
3 & 3.544906 & 10.51319 & 1863.048 & 592.5034 & 1.8099522 \\
4 & 3.544906 & 9.452852 & 18043.82 & 6467.390 & 1.8099522 \\
5 & 3.544906 & 8.670248 & 176616.1 & 69600.95 & 1.8099522 \\
6 & 3.544906 & 8.060855 & 1739616.1 & 741610.1 & 1.8099522 \\
\hline
\end{tabular}

\*Also shown is $S_{act}$, the actual number of significant figures achieved when using the calculated summation limits.

Table 4c. One-Dimensional Point Source Array: Calculated and Actual Values of the Summation Limits for Various Values of $S_{spec}$, Keeping the Frequency Fixed at $d = 5.5\lambda_0$, Using $\beta = 2$ as a Factor\textsuperscript{a}

\begin{tabular}{cccccccc}
\hline
$L$ & $P_{cal}$ & $P_{act}$ & $M_{cal}$ & $M_{act}$ & $S_{act}$ \\
\hline
1 & 5 & 5 & 0 & 0 & 0.21 \\
2 & 5 & 5 & 0 & 0 & 0.21 \\
3 & 6 & 6 & 0 & 0 & 0.67 \\
4 & 6 & 6 & 0 & 0 & 0.67 \\
5 & 7 & 7 & 0 & 0 & 7.23 \\
6 & 7 & 7 & 0 & 0 & 7.23 \\
\hline
\end{tabular}

\*Also shown is $S_{act}$, the actual number of significant figures achieved when using the calculated summation limits.

Table 5a. Two-Dimensional Point Source Array: $G^L_{0,0}$ and $G^L_{0,0}$ Obtained Using $E_{opt}$ and $E_L$, Compared With $G$, the Exact Value of the Green’s Function\textsuperscript{a}

\begin{tabular}{ccccccccc}
\hline
$a/\lambda_0$ & $E_L$ & $G^L_{0,0}$ Using $E_{opt}$ & $G^L_{0,0}$ Using $E_L$ & $G^L_{0,0}$ Using $E_{opt}$ & $G^L_{0,0}$ Using $E_L$ & $G$ Exact \\
\hline
10.5 & 19.93034 & 1.1795E + 147 & 51.74205 & 1.1830E+147 & 1803.7876 & 0.4739999 \\
5.5 & 10.51319 & 3.0538E + 038 & 189.1693 & 3.0869E+038 & 1863.0479 & 2.6124853 \\
4.5 & 8.556454 & 1.0386E + 025 & 286.2788 & 1.05567E+025 & 1869.2320 & 3.5952074 \\
3.5 & 6.595795 & 2.0987E + 014 & 482.4770 & 2.1571E+014 & 1870.9950 & 1.0027102 \\
2.5 & 4.645089 & 2.7137E + 006 & 972.9746 & 2.87365E+006 & 1866.7076 & 1.4841352 \\
\hline
\end{tabular}

Table 5b. Two-Dimensional Point Source Array: $G^L_{0,0}$ and $G^L_{0,0}$ for Various Values of $L$, Keeping the Frequency Fixed at $a = 5.5\lambda_0$

\begin{tabular}{cccccccc}
\hline
$L$ & $E_{opt}$ & $E_L$ & $G^L_{0,0}$ Using $E_{opt}$ & $G^L_{0,0}$ Using $E_L$ & $G$ Exact \\
\hline
1 & 3.544906 & 14.70097 & 20.70892 & 0.956796 & 2.6124583 \\
2 & 3.544906 & 12.07276 & 195.6083 & 14.43552 & 2.6124583 \\
3 & 3.544906 & 10.51319 & 1863.048 & 189.1693 & 2.6124583 \\
4 & 3.544906 & 9.452852 & 18043.82 & 2323.698 & 2.6124583 \\
5 & 3.544906 & 8.670248 & 176616.1 & 27471.14 & 2.6124583 \\
6 & 3.544906 & 8.060855 & 1739616.1 & 316496.5 & 2.6124583 \\
\hline
\end{tabular}

Table 5c. Two-Dimensional Point Source Array: Calculated and Actual Values of the Summation Limits for Various Values of $S_{spec}$, Keeping the Frequency Fixed at $a = 5.5\lambda_0$, Using $\beta = 4$ as a Factor\textsuperscript{a}

\begin{tabular}{cccccccc}
\hline
$L$ & $P_{cal}$ & $Q_{cal}$ & $P_{act}$ & $Q_{act}$ & $M_{cal}$ & $N_{cal}$ & $M_{act}$ & $N_{act}$ & $S_{act}$ \\
\hline
1 & 5 & 5 & 5 & 5 & 0 & 0 & 0 & 0 & 2.07 \\
2 & 5 & 5 & 5 & 5 & 0 & 0 & 0 & 0 & 2.07 \\
3 & 6 & 6 & 6 & 6 & 0 & 0 & 0 & 0 & 4.53 \\
4 & 6 & 6 & 6 & 6 & 0 & 0 & 0 & 0 & 4.53 \\
5 & 6 & 6 & 7 & 7 & 0 & 0 & 0 & 0 & 7.09 \\
6 & 7 & 7 & 7 & 7 & 0 & 0 & 0 & 0 & 7.09 \\
\hline
\end{tabular}

\*Also shown is $S_{act}$, the actual number of significant figures achieved when using the calculated summation limits.
Table A1. Definitions of the Integrals and Parameters Appearing in the Four Cases, and the Final Results of the Asymptotic Evaluation

| Case 1a | Case 1b | Case 2a | Case 2b |
|---------|---------|---------|---------|
| $I = \int \frac{e^{-(R,s)^2 + (\frac{s}{2})^2}}{s} ds$ | $\int \frac{e^{-(R,s)^2 + (\frac{s}{2})^2}}{s} ds$ | $\int \frac{e^{\pi^2 + (k_s)^2}}{s} ds$ | $\int \frac{e^{\pi^2 + (k_s)^2}}{s} ds$ |
| $s_0$ | $E$ | $-1/(2E)^2$ | $1/(2E)^2$ |
| $\Omega$ | $k^2$ | $(k \rho)^2$ | $- (k \rho)^2$ |
| $f(s)$ | $\frac{e^{-(R,s)^2}}{s}$ | $\frac{e^{\pi^2}}{s}$ | $\frac{e^{\pi^2}}{s}$ |
| $g(s)$ | $\frac{(2s)^2}{2(R_sE)^2}$ | $\frac{(2s)^2}{2(R_sE)^2}$ | $\frac{(2s)^2}{2(R_sE)^2}$ |
| $I_-$ | $\frac{e^{-(R_sE)^2 + (\frac{s}{2})^2}}{2(\frac{s}{2})^2}$ | $\frac{e^{-(R_sE)^2 + (\frac{s}{2})^2}}{2(\frac{s}{2})^2}$ | $\frac{e^{-(R_sE)^2 + (\frac{s}{2})^2}}{2(\frac{s}{2})^2}$ |
| $I_+$ | $\frac{e^{-(R_sE)^2 + (\frac{s}{2})^2}}{2(\frac{s}{2})^2}$ | $\frac{e^{-(R_sE)^2 + (\frac{s}{2})^2}}{2(\frac{s}{2})^2}$ | $\frac{e^{-(R_sE)^2 + (\frac{s}{2})^2}}{2(\frac{s}{2})^2}$ |

Table A1 demonstrates the accuracy of the asymptotic evaluation for the Ewald sums, where $S_{act}$ and $S_{spec}$ represent the actual and specified number of significant digits, respectively. The tables summarize the asymptotic limits for various cases, including 1-D and 2-D arrays of point sources, with 3 and 2-D cases respectively. The accuracy of the proposed formula is verified through numerical simulations and results illustrate the accuracy of the proposed formula.

7. Conclusions

The Ewald method is a very efficient method for calculating the periodic free-space Green’s function for three different cases: a 1-D array of line sources, a 1-D array of point sources, and a 2-D array of point sources. However, as noted in Kustepeli and Martin [2000] the method suffers from accuracy problems at high frequency due to a loss of significant figures that occurs from a cancellation when combining the spectral and spatial series that appear in the method. The method proposed here determines the “best” value of the splitting parameter $E$ that appears in the method to yield the fastest convergence of the Ewald sums while limiting the number of lost significant digits to a specified level $L$. The derivation has been presented in a unified fashion so that all three cases are treated together, with Table 1 giving the parameters needed for the different cases. The predicted loss of significant digits is verified through numerical simulations and the results illustrate the accuracy of the proposed formula.

Appendix A

In this appendix we summarize the asymptotic evaluation of the exponential-integral series that appear
in the 1-D point-source spectral series and the 1-D line source spectral series. The asymptotic evaluation of the terms involving the complementary error function that appear in the 1-D point-source spectral series, the 1-D line source spectral series, and the 2-D point-source spatial and spectral series, uses (10) and is straightforward.

[28] In performing the asymptotic evaluations of the exponential-integral series that appear in (5) and (6), the series have been cast back into their integral forms, namely the forms shown in equation (15) in Capolino et al. [2007] and equation (11) in Capolino et al. [2005]. These are the forms appearing in (11) and (12), respectively. In the following, case 1 denotes the 1-D line source array and case 2 denotes the 1-D point-source array. For each of these two cases, part (a) is defined as the subcase involving the determination of the splitting parameter, whereas part (b) is the subcase involving the determination of the summation limit. To determine the splitting parameter, either \( k \) or \( k_{\phi} \) is the large parameter in the asymptotic evaluation. To determine the summation limits, either \( R_\Omega \) or the wave number \( k_{\phi} \) is the large parameter in the asymptotic evaluation. In (11), we made the substitution \( s = -u \) for case 2a. For case 2b we then used the notational change \( s = u \) for consistency. (Note that in (11) the upper limit denotes the complex point at infinity, which has been taken as positive infinity for the derivation in case 2b and negative infinity for the derivation in case 2a).

[29] In all four cases, each of the integral expressions can be put into the form

\[
I(\Omega) = \int_{x_0}^{\infty} f(s) e^{i g(s)} ds, \tag{A1}
\]

where \( \Omega \) is a real positive parameter that becomes large. The terms \( x_0, \ Omega, \) and the functions \( f(s) \) and \( g(s) \) are defined in Table A1 for each of the four cases. For large \( \Omega \) we use the integration-by-parts asymptotic approximation [Felsen and Marcuvitz, 1994; Bleistein and Handelsman, 1986] to obtain

\[
I(\Omega) \sim -\frac{f(s_0)}{\Omega g'(s_0)} e^{i g(s_0)}. \tag{A2}
\]

Table A1 summarizes for each of the four cases the integration limits, either \( \Omega_0 \) or \( x_0 \), and all of the parameters that appear in (A1).

References

Bleistein, N., and R. A. Handelsman (1986), *Asymptotic Expansions of Integrals*, chap. 3, pp. 69–76, Dover, New York.

Capolino, F., D. R. Wilton, and W. A. Johnson (2005), Efficient computation of the 2D Green’s function for 1D periodic structures using the Ewald method, *IEEE Trans. Antennas Propag.*, 53, 2977–2984.

Capolino, F., D. R. Wilton, and W. A. Johnson (2007), Efficient computation of the 3D Green’s function for the Helmholtz operator for a linear array of point sources using the Ewald method, *J. Comput. Phys.*, 223, 250–261.

Ewald, P. P. (1921), Die berechnung optischer und elektrostatischer gitterpotentiale, *Ann. Phys.*, 64, 253–287.

Felsen, L. B., and N. Marcuvitz (1994), *Radiation and Scattering of Waves*, chap. 4, p. 376, IEEE Press, Piscataway, N. J. Jordan, K. E., G. R. Richter, and P. Sheng (1986), An efficient numerical evaluation of the Green’s function for the Helmholtz operator on periodic structures, *J. Comput. Phys.*, 63, 222–235.

Khayat, M. A., and D. R. Wilton (2005), Numerical evaluation of singular and near singular potential integrals, *IEEE Trans. Antennas Propag.*, 53, 3180–3190.

Kustepeli, A., and A. Q. Martin (2000), On the splitting parameter in the Ewald method, *Microwave Guided Wave Lett.*, 10, 168–170.

Mathis, A. W., and A. F. Peterson (1996), A comparison of acceleration procedures for the two-dimensional periodic Green’s function, *IEEE Trans. Antennas Propag.*., 44, 567–571.

Mathis, A. W., and A. F. Peterson (1998), Efficient electromagnetic analysis of a doubly infinite array of rectangular apertures, *IEEE Trans. Microwave Theory Tech.*, 46, 46–54.

Oroskar, S., D. R. Jackson, and D. R. Wilton (2006), Efficient computation of the 2D periodic Green’s function using the Ewald method, *J. Comput. Phys.*, 219, 899–911.