

# 1

## Basics

This chapter presents some of the preliminaries to the multilevel fast multipole algorithm (MLFMA). Using Maxwell's equations and boundary conditions, surface integral equations are derived to formulate electromagnetics problems involving metallic and dielectric objects. Discretizations of the surface integral equations with basis and testing functions on triangular domains lead to dense matrix equations, which can be solved iteratively via MLFMA. Numerical integrations on triangular domains, different types of excitations, iterative algorithms, and preconditioning are also discussed.

### 1.1 Introduction

Solving electromagnetics problems is extremely important to analyze electromagnetic interactions of electronic devices with each other and with their environments including living and nonliving objects [1]. A plethora of applications in the areas of antennas [2]–[13], radars [14], optics [15], medical imaging [16], wireless communications [17], nanotechnology [18], metamaterials [19]–[25], photonic crystals [26]–[33], remote sensing, and electronic packaging involve scattering and/or radiation of electromagnetic waves. The following are some examples of popular electromagnetics problems:

- Radiation from antennas, e.g., designing novel antennas and estimating their adverse effects on the human health.
- Indoor and outdoor propagation, e.g., optimizations of wireless electromagnetic sources to cover required areas with minimum power.
- Visualization via optical imaging devices, such as lens systems in microscopy, astronomy, surgery, and biology.
- Transmission through frequency-selective photonic crystals and metamaterials, which are artificial structures having a vast variety of applications, such as invisibility cloaking.
- Detection of ordinary and stealth airborne and naval targets with radar systems using the radar cross section (RCS).

- Scattering from biological structures, such as red blood cells (RBCs), to diagnose various diseases.
- Electromagnetic compatibility of novel electronic devices, such as THz-range circuits.

Solutions of these problems are essential in the design and manufacturing cycles of electronic devices; they enable engineers to explore novel designs and prevent the waste of the resources involved in building prototypes and carrying out laboratory tests.

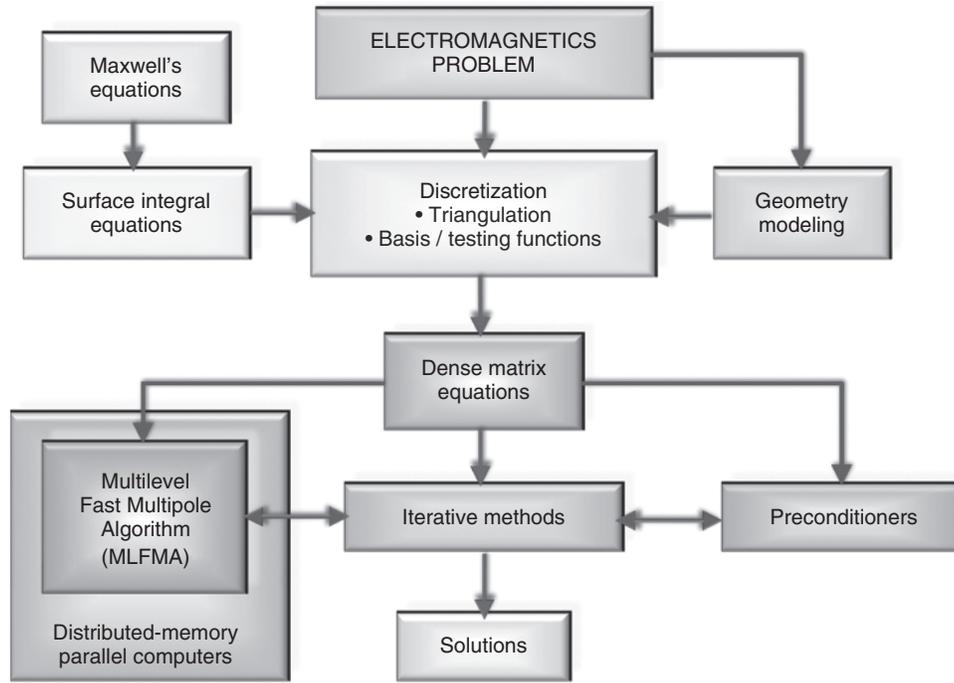
A quick look at a list of emerging technologies shows that electromagnetic fields will play a major role in the future. Wireless communications, e.g., cellular phones, have already become key components of modern life, with continuously increasing markets. Wireless energy transferring systems allowing unplugged electrical devices are now at the prototype stage, and are a major priority of technology companies. Novel designs in nanotechnology will require better imaging systems with subwavelength focusing abilities that can be achieved with metamaterials. Electromagnetic scattering from human tissues will become more and more popular to design diagnostic tools for noninvasive treatments. Hence, electromagnetics simulations will maintain their importance in the future in testing novel devices based on electromagnetic fields and in reducing their design and manufacturing costs.

Electromagnetics problems can be formulated rigorously with Maxwell's equations. Unfortunately, these equations can only be solved analytically for a few canonical objects, such as a sphere [34]. Recently, computational electromagnetics has become a scientifically popular area, where electromagnetic radiation, scattering, and transmission problems in the aforementioned areas are solved with numerical techniques. Mathematical formulations of physical events lead to a set of equations that can be solved numerically by using computers. Thanks to advances in computer technology and solution algorithms, it has become possible to solve many real-life problems involving complicated structures.

Developing a fast and accurate electromagnetics solver requires a well-designed combination of diverse components in many areas, such as wave theory, integral equations, numerical techniques, fast algorithms, iterative solvers, preconditioners, parallel implementations, high-performance computing, and computer technology. Accurate simulations of real-life problems usually require the solution of numerical problems involving large numbers of unknowns, which cannot be achieved easily, even when using the most powerful computers. Therefore, one needs to develop special acceleration algorithms, such as MLFMA [35], in order to solve those large-scale problems efficiently on relatively inexpensive computing platforms. This book is about MLFMA and its implementation for fast and accurate solutions of electromagnetics problems.

## 1.2 Simulation Environments Based on MLFMA

Figure 1.1 presents a typical simulation environment based on MLFMA. Surface integral equations are derived from Maxwell's equations without any fundamental approximations and are applied to the modeled geometry represented by surfaces. Simultaneous discretizations of the integral equations and surfaces using a set of basis and testing functions defined on triangles lead to dense matrix equations. These matrix equations can be solved iteratively, where the required matrix-vector multiplications (MVMs) are performed efficiently by MLFMA. Optionally, convergence of iterations can be accelerated with preconditioners. MLFMA is employed on parallel computers, particularly those with distributed-memory architectures, to



**Figure 1.1** A typical simulation environment based on MLFMA.

solve very large problems. Finally, solutions, e.g., scattered or radiated electromagnetic fields, are obtained after a certain number of iterations.

Although Figure 1.1 illustrates the main focus of this book, many different simulation environments can be constructed using alternative components. For example, volume integral equations [36]–[40] discretized with tetrahedra can be used to formulate problems. Instead of MLFMA, one can use the fast Fourier transform (FFT) [41]–[46] or the adaptive integral method [47],[48] to accelerate MVMs. MLFMA can be further accelerated via asymptotic techniques [49],[50]. Besides, in the context of parallelization, it can be employed on shared-memory architectures. Figure 1.1 merely depicts one of the simulation environments that have passed the time test for the accurate and efficient solutions of electromagnetics problems.

### 1.3 From Maxwell's Equations to Integro-Differential Operators

Maxwell's equations for linear, isotropic, and homogeneous (simple) media can be written as

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\mu \frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t} \quad (1.1)$$

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \epsilon \frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} + \mathbf{J}(\mathbf{r}, t) \quad (1.2)$$

$$\nabla \cdot \mathbf{E}(\mathbf{r}, t) = \frac{1}{\epsilon} \rho_e(\mathbf{r}, t) \quad (1.3)$$

$$\nabla \cdot \mathbf{H}(\mathbf{r}, t) = 0, \quad (1.4)$$

where  $\epsilon$  and  $\mu$  are the permittivity and permeability of the medium,  $\mathbf{E}$  and  $\mathbf{H}$  are the electric and magnetic fields,  $\mathbf{J}$  is the electric current density, and  $\rho_e$  is the electric charge density, which is related to the electric current via the continuity equation as

$$\nabla \cdot \mathbf{J}(\mathbf{r}, t) = -\frac{\partial \rho_e(\mathbf{r}, t)}{\partial t}. \quad (1.5)$$

In general, electromagnetic fields and sources (currents and charges) depend on position  $\mathbf{r} = \hat{x}x + \hat{y}y + \hat{z}z$  and time  $t$ . Assuming time-harmonic electromagnetic fields and sources with  $e^{-i\omega t}$  time dependence, Maxwell's equations can be rewritten as

$$\nabla \times \mathbf{E}(\mathbf{r}) = i\omega\mu\mathbf{H}(\mathbf{r}) \quad (1.6)$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = -i\omega\epsilon\mathbf{E}(\mathbf{r}) + \mathbf{J}(\mathbf{r}) \quad (1.7)$$

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{1}{\epsilon} \rho_e(\mathbf{r}) \quad (1.8)$$

$$\nabla \cdot \mathbf{H}(\mathbf{r}) = 0, \quad (1.9)$$

where

$$\nabla \cdot \mathbf{J}(\mathbf{r}) = i\omega\rho_e(\mathbf{r}). \quad (1.10)$$

Note that, using the phasor notation,

$$f_s(\mathbf{r}, t) = \text{Re}\{f_s(\mathbf{r}) \exp(-i\omega t)\} \quad (1.11)$$

$$\mathbf{f}_v(\mathbf{r}, t) = \text{Re}\{\mathbf{f}_v(\mathbf{r}) \exp(-i\omega t)\} \quad (1.12)$$

$$\frac{\partial f_s(\mathbf{r}, t)}{\partial t} = -i\omega \text{Re}\{f_s(\mathbf{r}) \exp(-i\omega t)\} \quad (1.13)$$

for a scalar  $f_s$  and a vector  $\mathbf{f}_v$ .

Applying the equivalence principle on dielectric objects introduces magnetic sources, in addition to electric sources. Maxwell's equations can be extended by including the magnetic current density  $\mathbf{M}$  and the magnetic charge density  $\rho_m$  as

$$\nabla \times \mathbf{E}(\mathbf{r}) = i\omega\mu\mathbf{H}(\mathbf{r}) - \mathbf{M}(\mathbf{r}) \quad (1.14)$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = -i\omega\epsilon\mathbf{E}(\mathbf{r}) + \mathbf{J}(\mathbf{r}) \quad (1.15)$$

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{1}{\epsilon} \rho_e(\mathbf{r}) \quad (1.16)$$

$$\nabla \cdot \mathbf{H}(\mathbf{r}) = \frac{1}{\mu} \rho_m(\mathbf{r}), \quad (1.17)$$

where

$$\nabla \cdot \mathbf{M}(\mathbf{r}) = i\omega\rho_m(\mathbf{r}) \quad (1.18)$$

is the continuity equation for the magnetic sources.

A set of Helmholtz equations for the electric and magnetic fields can be directly derived from Maxwell's equations. Taking the curl of (1.14) and using (1.15), (1.16), and the identity

$$\nabla \times \nabla \times \mathbf{f}_v = \nabla \nabla \cdot \mathbf{f}_v - \nabla^2 \mathbf{f}_v, \quad (1.19)$$

one can obtain

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) = i\omega\mu\nabla \times \mathbf{H}(\mathbf{r}) - \nabla \times \mathbf{M}(\mathbf{r}) \quad (1.20)$$

$$\nabla \nabla \cdot \mathbf{E}(\mathbf{r}) - \nabla^2 \mathbf{E}(\mathbf{r}) = i\omega\mu[-i\omega\epsilon\mathbf{E}(\mathbf{r}) + \mathbf{J}(\mathbf{r})] - \nabla \times \mathbf{M}(\mathbf{r}) \quad (1.21)$$

$$\frac{1}{\epsilon}\nabla\rho_e(\mathbf{r}) - \nabla^2 \mathbf{E}(\mathbf{r}) = \omega^2\mu\epsilon\mathbf{E}(\mathbf{r}) + i\omega\mu\mathbf{J}(\mathbf{r}) - \nabla \times \mathbf{M}(\mathbf{r}) \quad (1.22)$$

$$\nabla^2 \mathbf{E}(\mathbf{r}) + k^2 \mathbf{E}(\mathbf{r}) = -i\omega\mu\mathbf{J}(\mathbf{r}) + \frac{1}{\epsilon}\nabla\rho_e(\mathbf{r}) + \nabla \times \mathbf{M}(\mathbf{r}), \quad (1.23)$$

where  $k = \omega\sqrt{\epsilon\mu} = 2\pi/\lambda$  is the wavenumber and  $\lambda$  is the wavelength. Similarly, taking the curl of (1.15) and using (1.14), (1.17), and (1.19),

$$\nabla \times \nabla \times \mathbf{H}(\mathbf{r}) = -i\omega\epsilon\nabla \times \mathbf{E}(\mathbf{r}) + \nabla \times \mathbf{J}(\mathbf{r}) \quad (1.24)$$

$$\nabla \nabla \cdot \mathbf{H}(\mathbf{r}) - \nabla^2 \mathbf{H}(\mathbf{r}) = -i\omega\epsilon[i\omega\mu\mathbf{H}(\mathbf{r}) - \mathbf{M}(\mathbf{r})] + \nabla \times \mathbf{J}(\mathbf{r}) \quad (1.25)$$

$$\frac{1}{\mu}\nabla\rho_m(\mathbf{r}) - \nabla^2 \mathbf{H}(\mathbf{r}) = \omega^2\mu\epsilon\mathbf{H}(\mathbf{r}) + i\omega\epsilon\mathbf{M}(\mathbf{r}) + \nabla \times \mathbf{J}(\mathbf{r}) \quad (1.26)$$

$$\nabla^2 \mathbf{H}(\mathbf{r}) + k^2 \mathbf{H}(\mathbf{r}) = -i\omega\epsilon\mathbf{M}(\mathbf{r}) + \frac{1}{\mu}\nabla\rho_m(\mathbf{r}) - \nabla \times \mathbf{J}(\mathbf{r}). \quad (1.27)$$

By writing the fields in terms of the vector and scalar potentials as

$$\mathbf{E}(\mathbf{r}) = i\omega\mathbf{A}_m(\mathbf{r}) - \nabla\phi_e(\mathbf{r}) - \frac{1}{\epsilon}\nabla \times \mathbf{A}_e(\mathbf{r}) \quad (1.28)$$

$$\mathbf{H}(\mathbf{r}) = i\omega\mathbf{A}_e(\mathbf{r}) - \nabla\phi_m(\mathbf{r}) + \frac{1}{\mu}\nabla \times \mathbf{A}_m(\mathbf{r}) \quad (1.29)$$

and using the Lorentz gauge, i.e.,

$$\nabla \cdot \mathbf{A}_m(\mathbf{r}) = i\omega\epsilon\mu\phi_e(\mathbf{r}) \quad (1.30)$$

$$\nabla \cdot \mathbf{A}_e(\mathbf{r}) = i\omega\mu\epsilon\phi_m(\mathbf{r}), \quad (1.31)$$

a set of Helmholtz equations for the potentials can be derived as

$$\nabla^2 \phi_e(\mathbf{r}) + k^2 \phi_e(\mathbf{r}) = -\frac{1}{\epsilon}\rho_e(\mathbf{r}) \quad (1.32)$$

$$\nabla^2 \phi_m(\mathbf{r}) + k^2 \phi_m(\mathbf{r}) = -\frac{1}{\mu}\rho_m(\mathbf{r}) \quad (1.33)$$

$$\nabla^2 \mathbf{A}_e(\mathbf{r}) + k^2 \mathbf{A}_e(\mathbf{r}) = -\epsilon\mathbf{M}(\mathbf{r}) \quad (1.34)$$

$$\nabla^2 \mathbf{A}_m(\mathbf{r}) + k^2 \mathbf{A}_m(\mathbf{r}) = -\mu\mathbf{J}(\mathbf{r}). \quad (1.35)$$

The Helmholtz equations in (1.32)–(1.35) can be solved for arbitrary current and charge distributions, leading to

$$\phi_e(\mathbf{r}) = \frac{1}{\epsilon} \int d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \rho_e(\mathbf{r}') \quad (1.36)$$

$$\phi_m(\mathbf{r}) = \frac{1}{\mu} \int d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \rho_m(\mathbf{r}') \quad (1.37)$$

$$\mathbf{A}_e(\mathbf{r}) = \epsilon \int d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \mathbf{M}(\mathbf{r}') \quad (1.38)$$

$$\mathbf{A}_m(\mathbf{r}) = \mu \int d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \mathbf{J}(\mathbf{r}'), \quad (1.39)$$

where

$$g(\mathbf{r}, \mathbf{r}') = \frac{\exp(ikR)}{4\pi R} \quad (R = |\mathbf{r} - \mathbf{r}'|) \quad (1.40)$$

is the homogeneous-space Green's function. Using (1.36)–(1.39) in (1.28) and (1.29), electromagnetic fields due to arbitrary current and charge distributions can be written as

$$\begin{aligned} \mathbf{E}(\mathbf{r}) &= ik\eta \int d\mathbf{r}' \left[ \mathbf{J}(\mathbf{r}') + \frac{1}{k^2} \nabla' \cdot \mathbf{J}(\mathbf{r}') \nabla \right] g(\mathbf{r}, \mathbf{r}') \\ &\quad - \int d\mathbf{r}' \nabla g(\mathbf{r}, \mathbf{r}') \times \mathbf{M}(\mathbf{r}') \end{aligned} \quad (1.41)$$

$$\begin{aligned} \mathbf{H}(\mathbf{r}) &= ik\eta^{-1} \int d\mathbf{r}' \left[ \mathbf{M}(\mathbf{r}') + \frac{1}{k^2} \nabla' \cdot \mathbf{M}(\mathbf{r}') \nabla \right] g(\mathbf{r}, \mathbf{r}') \\ &\quad + \int d\mathbf{r}' \nabla g(\mathbf{r}, \mathbf{r}') \times \mathbf{J}(\mathbf{r}'), \end{aligned} \quad (1.42)$$

where  $\eta = \sqrt{\frac{\mu}{\epsilon}}$  is the intrinsic impedance. Finally, by defining the integro-differential operators  $\mathcal{T}$  and  $\mathcal{K}$  as

$$\mathcal{T}\{\mathbf{X}\}(\mathbf{r}) = ik \int d\mathbf{r}' \left[ \mathbf{X}(\mathbf{r}') + \frac{1}{k^2} \nabla' \cdot \mathbf{X}(\mathbf{r}') \nabla \right] g(\mathbf{r}, \mathbf{r}') \quad (1.43)$$

$$\mathcal{K}\{\mathbf{X}\}(\mathbf{r}) = \int d\mathbf{r}' \mathbf{X}(\mathbf{r}') \times \nabla' g(\mathbf{r}, \mathbf{r}'), \quad (1.44)$$

the expressions for the electromagnetic fields can be rewritten as

$$\mathbf{E}(\mathbf{r}) = \eta \mathcal{T}\{\mathbf{J}\}(\mathbf{r}) - \mathcal{K}\{\mathbf{M}\}(\mathbf{r}) \quad (1.45)$$

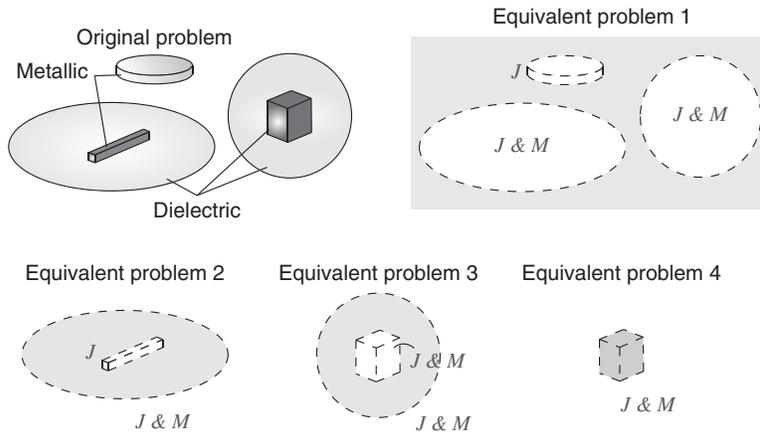
$$\mathbf{H}(\mathbf{r}) = \eta^{-1} \mathcal{T}\{\mathbf{M}\}(\mathbf{r}) + \mathcal{K}\{\mathbf{J}\}(\mathbf{r}). \quad (1.46)$$

Equations (1.45) and (1.46) clearly show the dependence of electromagnetic fields on electromagnetic sources in a simple medium.

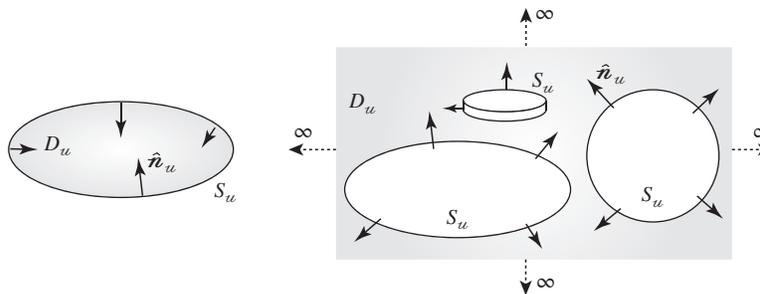
### 1.4 Surface Integral Equations

In general, electromagnetics problems involve three-dimensional objects with arbitrary geometries that can be modeled with perfectly conducting and dielectric (lossy or lossless) regions. As depicted in Figure 1.2, the equivalence principle can be used to decompose the original problem into equivalent problems by employing equivalent currents on surfaces. Equivalent problems are defined for all nonmetallic regions, including the host medium extending to infinity. Then, integral equations can be derived by applying the operators on the currents and enforcing the boundary conditions for the tangential components of the electromagnetic fields.

Consider a homogeneous dielectric region  $D_u$  bounded by a surface  $S_u$ . As depicted in Figure 1.3, this region may extend to infinity. Considering the tangential components of the



**Figure 1.2** Applying the equivalence principle to an electromagnetics problem involving multiple dielectric and metallic regions.



**Figure 1.3** A dielectric region enclosed in a surface and a dielectric region extending to infinity. In each case, the region is bounded by a surface  $S_u$ , which consists of multiple unconnected parts in the second case.

electric and magnetic fields on  $S_u$ , one can derive

$$\begin{aligned} -\eta_u \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathcal{T}_u\{\mathbf{J}_u\}(\mathbf{r}) + \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathcal{K}_u\{\mathbf{M}_u\}(\mathbf{r}) \\ - \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathbf{E}_u^{inc}(\mathbf{r}) = -\hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathbf{E}_u(\mathbf{r}) \end{aligned} \quad (1.47)$$

$$\begin{aligned} -\eta_u^{-1} \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathcal{T}_u\{\mathbf{M}_u\}(\mathbf{r}) - \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathcal{K}_u\{\mathbf{J}_u\}(\mathbf{r}) \\ - \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathbf{H}_u^{inc}(\mathbf{r}) = -\hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathbf{H}_u(\mathbf{r}), \end{aligned} \quad (1.48)$$

where  $\mathbf{r} \in S_u$ ,  $\hat{\mathbf{n}}_u$  is the inward normal unit vector, and  $\mathbf{E}_u^{inc}$  and  $\mathbf{H}_u^{inc}$  are the incident electric and magnetic fields, respectively, created by external sources located in  $D_u$ . The equivalent currents equivalent electric and magnetic currents are defined as

$$\mathbf{J}_u(\mathbf{r}) = \mathcal{I}^{\times n_u}\{\mathbf{H}_u\}(\mathbf{r}) = \hat{\mathbf{n}}_u \times \mathbf{H}_u(\mathbf{r}) \quad (1.49)$$

$$\mathbf{M}_u(\mathbf{r}) = -\mathcal{I}^{\times n_u}\{\mathbf{E}_u\}(\mathbf{r}) = -\hat{\mathbf{n}}_u \times \mathbf{E}_u(\mathbf{r}), \quad (1.50)$$

where

$$\mathcal{I}^{\times n_u}\{\mathbf{X}\}(\mathbf{r}) = \hat{\mathbf{n}}_u \times \mathcal{I}\{\mathbf{X}\}(\mathbf{r}) = \hat{\mathbf{n}}_u \times \mathbf{X}(\mathbf{r}) \quad (1.51)$$

and  $\mathcal{I}$  is the identity operator. The operator  $\mathcal{K}_u$  is commonly separated into principal-value and limit parts (see Section 6.1) as

$$\mathcal{K}_u\{\mathbf{X}\}(\mathbf{r}) = \mathcal{K}_{PV,u}\{\mathbf{X}\}(\mathbf{r}) - \frac{4\pi - \Omega_u}{4\pi} \mathcal{I}^{\times n_u}\{\mathbf{X}\}(\mathbf{r}), \quad (1.52)$$

where  $0 \leq \Omega_u \leq 4\pi$  is the solid angle that is nonzero when the observation point  $\mathbf{r}$  is on the surface. Using (1.52) in (1.47) and (1.48), and rearranging the terms, one can obtain

$$\begin{aligned} -\eta_u \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathcal{T}_u\{\mathbf{J}_u\}(\mathbf{r}) + \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathcal{K}_{PV,u}\{\mathbf{M}_u\}(\mathbf{r}) \\ - \frac{\Omega_u}{4\pi} \mathcal{I}^{\times n_u}\{\mathbf{M}_u\}(\mathbf{r}) = \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathbf{E}_u^{inc}(\mathbf{r}) \end{aligned} \quad (1.53)$$

$$\begin{aligned} -\eta_u^{-1} \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathcal{T}_u\{\mathbf{M}_u\}(\mathbf{r}) - \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathcal{K}_{PV,u}\{\mathbf{J}_u\}(\mathbf{r}) \\ + \frac{\Omega_u}{4\pi} \mathcal{I}^{\times n_u}\{\mathbf{J}_u\}(\mathbf{r}) = \hat{\mathbf{n}}_u \times \hat{\mathbf{n}}_u \times \mathbf{H}_u^{inc}(\mathbf{r}). \end{aligned} \quad (1.54)$$

Equations (1.53) and (1.54) are known as the electric-field integral equation (EFIE) and the magnetic-field integral equation (MFIE), respectively.

As discussed in Chapter 2, EFIE and MFIE each have two forms. In the forms given in (1.53) and (1.54), the tangential components of the electric and magnetic fields are sampled directly since

$$-\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathbf{f}_v = \mathbf{f}_v - \hat{\mathbf{n}} \hat{\mathbf{n}} \cdot \mathbf{f}_v \quad (1.55)$$

for an arbitrary vector  $\mathbf{f}_v$ . Hence, (1.53) and (1.54) are commonly known as tangential integral equations and denoted with T-EFIE and T-MFIE, respectively. On the other hand, it is also

possible to sample a rotated version of the tangential electromagnetic fields. Specifically, (1.53) and (1.54) can be modified as

$$\eta_u \hat{\mathbf{n}}_u \times \mathcal{T}_u \{ \mathbf{J}_u \}(\mathbf{r}) - \hat{\mathbf{n}}_u \times \mathcal{K}_{PV,u} \{ \mathbf{M}_u \}(\mathbf{r}) + \frac{\Omega_u}{4\pi} \mathcal{I} \{ \mathbf{M}_u \}(\mathbf{r}) = -\hat{\mathbf{n}}_u \times \mathbf{E}_u^{inc}(\mathbf{r}) \quad (1.56)$$

$$\eta_u^{-1} \hat{\mathbf{n}}_u \times \mathcal{T}_u \{ \mathbf{M}_u \}(\mathbf{r}) + \hat{\mathbf{n}}_u \times \mathcal{K}_{PV,u} \{ \mathbf{J}_u \}(\mathbf{r}) - \frac{\Omega_u}{4\pi} \mathcal{I} \{ \mathbf{J}_u \}(\mathbf{r}) = -\hat{\mathbf{n}}_u \times \mathbf{H}_u^{inc}(\mathbf{r}), \quad (1.57)$$

which are commonly known as normal integral equations and denoted with N-EFIE and N-MFIE, respectively. It should be emphasized that normal integral equations do not involve electromagnetic fields in normal directions; this convention is used throughout the book.

## 1.5 Boundary Conditions

Solutions of the surface integral equations (EFIE and MFIE) to compute the electric and magnetic currents require the use of appropriate boundary conditions at boundaries. As depicted in Figure 1.4, two scenarios are of particular interest. When two dielectric regions  $D_u$  and  $D_v$  are separated by a surface  $S_{uv}$ , the tangential electric and magnetic fields are continuous across the surface. This can be written as

$$\mathbf{J}_u(\mathbf{r}) = \hat{\mathbf{n}}_u \times \mathbf{H}_u(\mathbf{r}) = -\mathbf{J}_v(\mathbf{r}) = -\hat{\mathbf{n}}_v \times \mathbf{H}_v(\mathbf{r}) \quad (1.58)$$

$$\mathbf{M}_u(\mathbf{r}) = -\hat{\mathbf{n}}_u \times \mathbf{E}_u(\mathbf{r}) = -\mathbf{M}_v(\mathbf{r}) = \hat{\mathbf{n}}_v \times \mathbf{E}_v(\mathbf{r}), \quad (1.59)$$

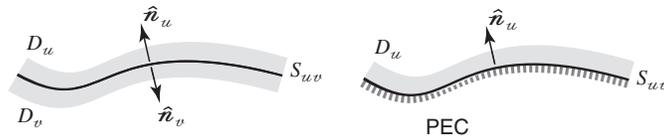
or simply

$$\mathbf{J}_u(\mathbf{r}) = -\mathbf{J}_v(\mathbf{r}) = \mathbf{J}(\mathbf{r}) = \hat{\mathbf{n}} \times \mathbf{H}(\mathbf{r}) \quad (1.60)$$

$$\mathbf{M}_u(\mathbf{r}) = -\mathbf{M}_v(\mathbf{r}) = \mathbf{M}(\mathbf{r}) = -\hat{\mathbf{n}} \times \mathbf{E}(\mathbf{r}) \quad (1.61)$$

with  $\hat{\mathbf{n}} = \hat{\mathbf{n}}_u$ . Hence, the integral equations derived for  $D_u$  and  $D_v$  can be combined in different ways to solve for  $\mathbf{J}$  and  $\mathbf{M}$ .

In many electromagnetics problems, some of the objects can be modeled as perfect electric conductors (PECs). Since electromagnetics fields cannot penetrate into PECs, surface integral equations are not required inside such objects. On a surface separating a dielectric medium



**Figure 1.4** Boundaries between two dielectric regions and on a PEC surface.

from a PEC, the magnetic current is identically zero and the electric current corresponds to the induced electric current, i.e.,

$$\mathbf{J}_u(\mathbf{r}) = \hat{\mathbf{n}} \times \mathbf{H}(\mathbf{r}) = \mathbf{J}^{ind}(\mathbf{r}) \quad (1.62)$$

$$\mathbf{M}_u(\mathbf{r}) = 0. \quad (1.63)$$

As discussed in Chapter 2, (1.62) and (1.63) are also valid for an open PEC with zero volume, since an open surface can be considered as a limiting case of a closed surface.

Finally, finding the equivalent currents  $\mathbf{J}_u$  and  $\mathbf{M}_u$  on the surface  $S_u$  bounding  $D_u$ , the secondary (scattered or radiated) electromagnetic fields can be computed everywhere inside  $D_u$  (and on  $S_u$ ) as

$$\mathbf{E}_u^{sec}(\mathbf{r}) = \eta_u \mathcal{T}_u\{\mathbf{J}_u\}(\mathbf{r}) - \mathcal{K}_{PV,u}\{\mathbf{M}_u\}(\mathbf{r}) + \frac{4\pi - \Omega_u}{4\pi} \mathcal{I}^{\times n_u}\{\mathbf{M}_u\}(\mathbf{r}) \quad (1.64)$$

$$\mathbf{H}_u^{sec}(\mathbf{r}) = \eta_u^{-1} \mathcal{T}_u\{\mathbf{M}_u\}(\mathbf{r}) + \mathcal{K}_{PV,u}\{\mathbf{J}_u\}(\mathbf{r}) - \frac{4\pi - \Omega_u}{4\pi} \mathcal{I}^{\times n_u}\{\mathbf{J}_u\}(\mathbf{r}). \quad (1.65)$$

## 1.6 Surface Formulations

For the solution of electromagnetics problems, various surface formulations can be derived by using diverse combinations of surface integral equations. Some of these formulations are known to be stable and provide accurate results [51]–[83], although their performances may vary significantly in terms of efficiency and accuracy. In general, surface formulations can be categorized into three groups, i.e., tangential, normal, and mixed, depending on their contents. Tangential formulations involve T-EFIE and/or T-MFIE, while normal formulations involve N-EFIE and/or N-MFIE. Mixed formulations are obtained by combining the tangential and normal formulations, and they contain at least one tangential equation and at least one normal equation.

For dielectric objects with homogeneous material properties, surface formulations are derived by properly combining the equations obtained for the inner and outer regions. To avoid numerical internal resonances, it is preferable to derive a combined-field integral equation (CFIE) by linearly combining EFIE and MFIE [62]. For example, a combination of T-EFIE and T-MFIE leads to a T-T-CFIE formulation. Similarly, one can obtain T-N-CFIE (T-EFIE+N-MFIE), N-T-CFIE (N-EFIE+T-MFIE), and N-N-CFIE (N-EFIE+N-MFIE) by combining EFIE and MFIE appropriately. In these formulations, the equations obtained for the inner and outer regions are solved simultaneously to obtain the two sets of unknowns, i.e., electric and magnetic currents. One can also derive various other formulations involving triple combinations, such as TN-N-CFIE (T-EFIE+N-EFIE+N-MFIE), for more stable solutions [68].

The CFIE formulations described above are based on a linear combination of EFIE and MFIE in the same way for each medium, but it is also possible to use different combinations for the inner and outer regions [69]. Alternatively, it is quite popular to linearly combine the inner and outer equations while solving EFIE and MFIE simultaneously. In this class of formulations, the Poggio-Miller-Chang-Harrington-Wu-Tsai (PMCHWT) [52]–[54] and Müller [51] formulations are well known and commonly used in the literature. Both tangential and normal versions of these formulations are possible; however, using a Galerkin discretization,

only the tangential PMCHWT (T-PMCHWT) formulation and the normal Müller formulation (NMF) are stable, whereas the normal PMCHWT and the tangential Müller formulations are usually unstable. Recently, remarkable efforts have been made to further improve dielectric formulations by devising novel combinations of integral equations. For example, the combined tangential formulation (CTF) [74] is similar to the T-PMCHWT formulation, but it involves a careful (and improved) scaling of T-EFIE and T-MFIE. A similar combination of N-EFIE and N-MFIE leads to the combined normal formulation (CNF) [74]. The modified NMF (MNMF) is obtained by normalizing equations in NMF to produce better-conditioned matrix equations [73]. Finally, a mixed formulation called the electric and magnetic current combined-field integral equation (JMCFIE) is derived by combining all four types of integral equations [71],[78],[80],[83].

Surface formulations of dielectric objects tend to be less accurate as the contrast of the object decreases and electromagnetic properties change slightly across dielectric interfaces. There are various applications that involve scattering from low-contrast objects, such as dielectric photonic crystals [29], RBCs in blood plasma [84]–[92], plastic land mines buried in soil [93], and polymeric materials [94]. When the contrast is low, however, conventional surface formulations encounter stability problems, and scattered fields cannot be calculated accurately with them. Those scattering problems can be solved accurately with volume formulations [36]–[39], which are usually stable independent of the contrast. On the other hand, it is also desirable to extend the applicability of surface integral equations to low-contrast problems to be able to take the advantage of surface formulations, which are usually discretized with fewer unknowns compared to volume formulations. The inaccuracy of surface formulations for the solution of low-contrast problems is due to insufficient modeling of radiating parts of equivalent currents defined on objects [95]. By extracting nonradiating currents and solving modified equations only for radiating currents (similar to volume formulations), scattered fields from low-contrast objects can be calculated accurately [96],[97]. These types of formulations are extensively discussed in Chapter 2, Section 2.2.

When the object is a PEC, T-EFIE, T-MFIE, N-EFIE, and N-MFIE can be solved independently to obtain the induced electric current on the surface. Especially, T-EFIE and N-MFIE are commonly used in the literature [52],[57],[61]. However, for closed conductors, those formulations suffer from internal resonances. Therefore, similar to dielectric formulations, it is essential to combine EFIE and MFIE appropriately to derive various CFIE formulations [55]. Conventional CFIE formulations are usually obtained by convex combinations of EFIE and MFIE. However, there are many studies presenting hybrid formulations, where EFIE and MFIE are coupled in diverse ways. For example, EFIE and MFIE can be applied on different parts of the object to efficiently and accurately solve electromagnetics problems involving complicated structures [59],[64],[65],[76],[82]. This type of hybrid formulations are detailed in Chapter 3, Section 3.6.1.

For PEC objects, T-EFIE suffers from the low-frequency breakdown problem (see Chapter 2, Section 2.3.4) when it is discretized with ordinary basis functions [98],[99]. Specifically, matrix equations obtained from T-EFIE become increasingly ill-conditioned as the discretization is refined. The T-N-CFIE formulation is also affected by the low-frequency breakdown since it contains T-EFIE [100],[101], whereas N-MFIE is usually stable unless the frequency is extremely low [102]. To eliminate the low-frequency breakdown problem, one can apply loop-star or loop-tree decomposition methods [103]–[107]. In addition, current and charge decompositions can be used [108],[109] to derive very stable formulations at low

frequencies. Recently, regularization of T-EFIE using Calderon identities has become very popular to tackle the low-frequency breakdown issue [101],[110]–[116].

## 1.7 Method of Moments and Discretization

Electromagnetics problems involving complicated objects can be discretized and solved numerically. By means of a simultaneous discretization of the geometry and surface integral equations, the equivalent currents are expanded in a series of basis functions. Then, the coefficients of the basis functions are calculated by solving dense matrix equations, which are derived by using the method of moments (MOM) [117],[118].

Let  $\mathcal{L}$  be a linear operator applied on an unknown vector function  $\mathbf{f}(\mathbf{r})$  to produce a known vector function  $\mathbf{g}(\mathbf{r})$ , i.e.,

$$\mathcal{L}\{\mathbf{f}\}(\mathbf{r}) = \mathbf{g}(\mathbf{r}). \quad (1.66)$$

Considering surface formulations,  $\mathcal{L}$  can be a combination of the integro-differential  $\mathcal{T}$  and  $\mathcal{K}$  operators as well as the identity operator  $\mathcal{I}$ , while the right-hand-side (RHS) function can be a combination of the incident electromagnetic fields created by external sources. Expanding  $\mathbf{f}(\mathbf{r})$  in a series of known basis functions  $\mathbf{b}_n$  ( $n = 1, 2, \dots, N$ ), i.e.,

$$\mathbf{f}(\mathbf{r}) \approx \sum_{n=1}^N \mathbf{a}[n] \mathbf{b}_n(\mathbf{r}), \quad (1.67)$$

and testing (1.66) using a set of testing functions  $\mathbf{t}_m$  ( $m = 1, 2, \dots, N$ ), one can obtain

$$\int d\mathbf{r} \mathbf{t}_m(\mathbf{r}) \cdot \sum_{n=1}^N \mathbf{a}[n] \mathcal{L}\{\mathbf{b}_n\}(\mathbf{r}) = \int d\mathbf{r} \mathbf{t}_m(\mathbf{r}) \cdot \mathbf{g}(\mathbf{r}) \quad (m = 1, 2, \dots, N). \quad (1.68)$$

In (1.67) and (1.68),  $\mathbf{a}[n]$  represents the  $n$ th unknown coefficient, i.e.,  $n$ th element of the unknown vector  $\mathbf{a}$ . Changing the order of the summation and integration, the equation becomes

$$\sum_{n=1}^N \mathbf{a}[n] \int d\mathbf{r} \mathbf{t}_m(\mathbf{r}) \cdot \mathcal{L}\{\mathbf{b}_n\}(\mathbf{r}) = \int d\mathbf{r} \mathbf{t}_m(\mathbf{r}) \cdot \mathbf{g}(\mathbf{r}) \quad (m = 1, 2, \dots, N), \quad (1.69)$$

which can be written as a matrix equation, i.e.,

$$\sum_{n=1}^N \mathbf{a}[n] \bar{\mathbf{Z}}[m, n] = \mathbf{w}[m] \quad (m = 1, 2, \dots, N), \quad (1.70)$$

where

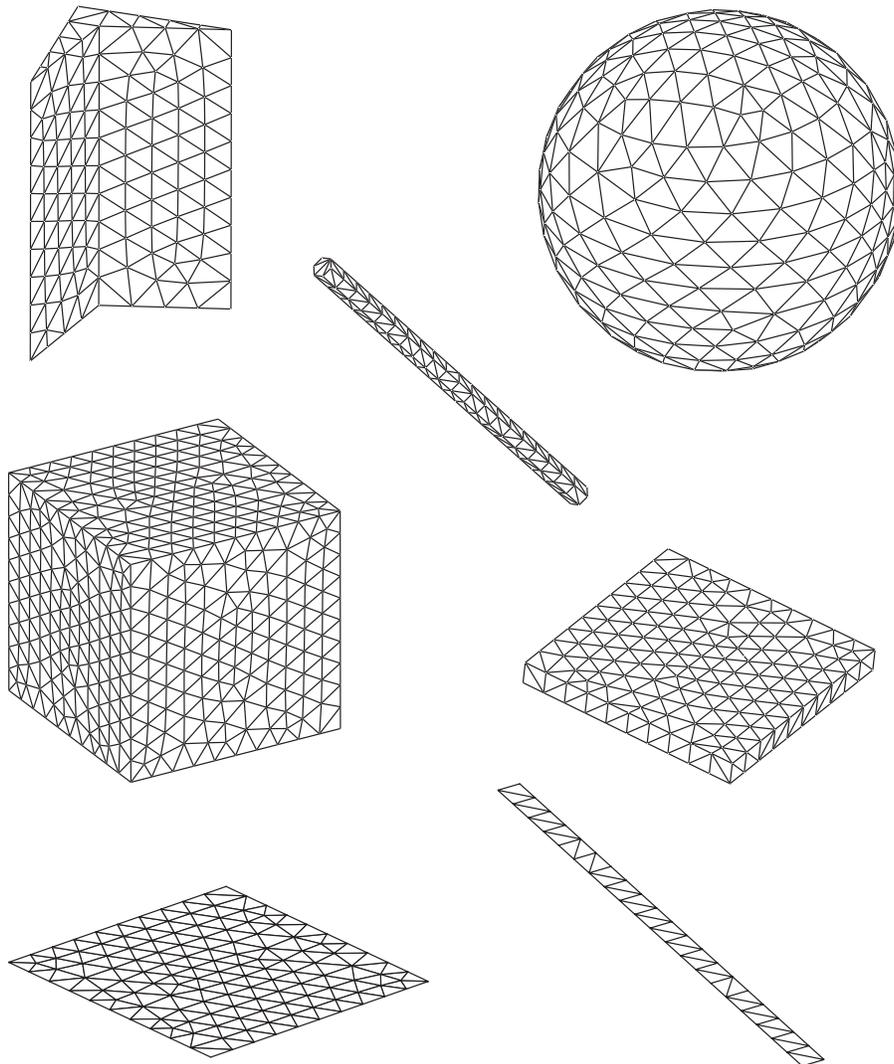
$$\bar{\mathbf{Z}}[m, n] = \int d\mathbf{r} \mathbf{t}_m(\mathbf{r}) \cdot \mathcal{L}\{\mathbf{b}_n\}(\mathbf{r}) \quad (1.71)$$

$$\mathbf{w}[m] = \int d\mathbf{r} \mathbf{t}_m(\mathbf{r}) \cdot \mathbf{g}(\mathbf{r}). \quad (1.72)$$

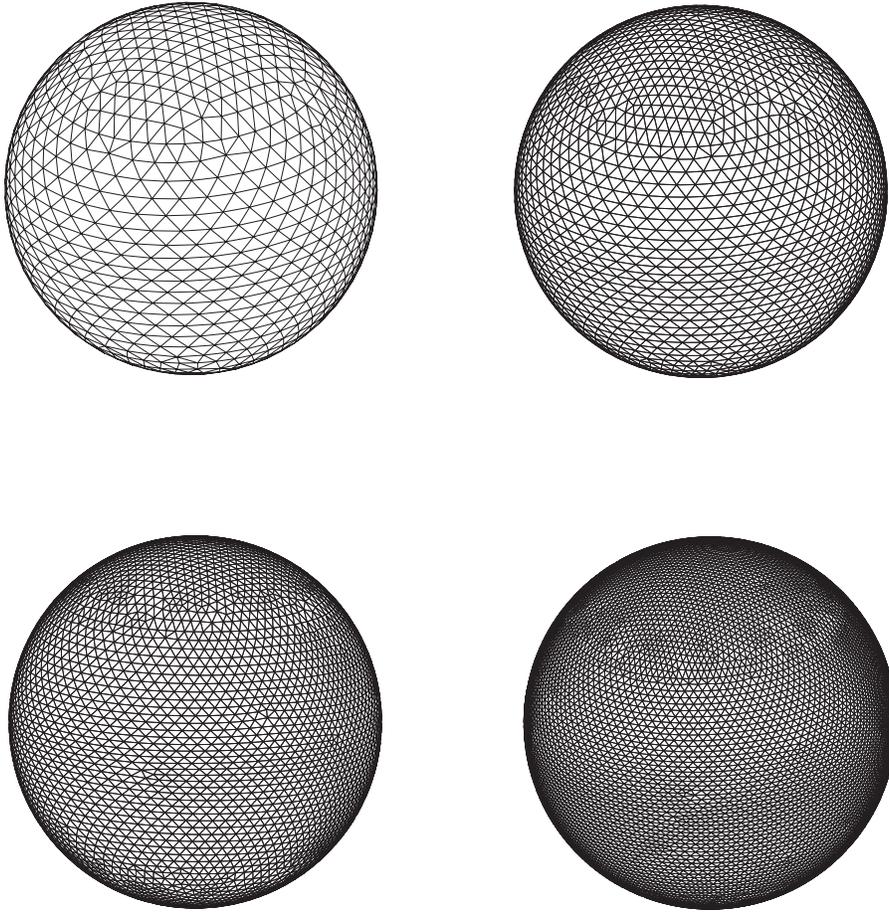
In the context of electromagnetics problems,  $\bar{\mathbf{Z}}$  and  $\mathbf{w}$  are usually called the impedance matrix and the excitation vector, respectively. Furthermore, elements of the impedance matrices

(obtained by the discretization of the integral-equation formulations) can be interpreted as electromagnetic interactions between discretization elements, i.e., basis and testing functions.

Using small planar triangles on which basis and testing functions are defined is a popular approach to discretize three-dimensional objects. Some examples of the triangulation of closed and open objects are illustrated in Figures 1.5, 1.6, and 1.7. In general, the size of triangles should be small with respect to wavelength, but the discretization depends on the nature of the problem, e.g., geometry and excitation, as well as the desired level of accuracy. There are actually two error sources that can appear in the triangulation stage. The first one is the approximation of the sources (and fields) with discrete functions. The second one is the modeling of



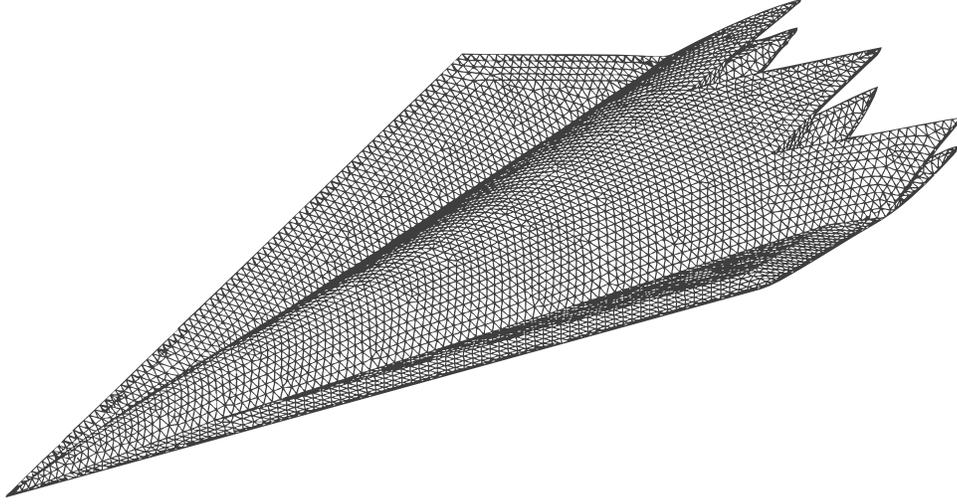
**Figure 1.5** Surface models of various canonical geometries with triangular meshing.



**Figure 1.6** Surface models of a sphere with triangular meshing using various mesh sizes.

curved surfaces with planar elements (e.g., see Figure 1.6). Both types of errors may be reduced by decreasing the element size unless limited by low-frequency breakdowns.

Applying a Galerkin scheme for the discretization of surface formulations, i.e., using the same set of functions to expand the current densities (basis functions) and to test the boundary conditions (testing functions), normal and mixed formulations will contain well-tested identity operators. It is well-known that matrix equations involving well-tested identity operators are diagonally dominant and well-conditioned [74],[119],[120]. However, tangential formulations do not contain well-tested identity operators, and their discretizations may lead to ill-conditioned matrix equations. Therefore, for the efficiency of solutions, normal and mixed formulations are usually preferable, especially for large objects discretized with large numbers of unknowns. On the other hand, errors in the discretization of well-tested identity operators may decrease the accuracy of solutions obtained with normal and mixed formulations [120]–[122]. The error level in these formulations compared to tangential formulations can be significant, especially in conventional implementations employing the



**Figure 1.7** Surface model of a stealth airborne target Flamme with triangular meshing.

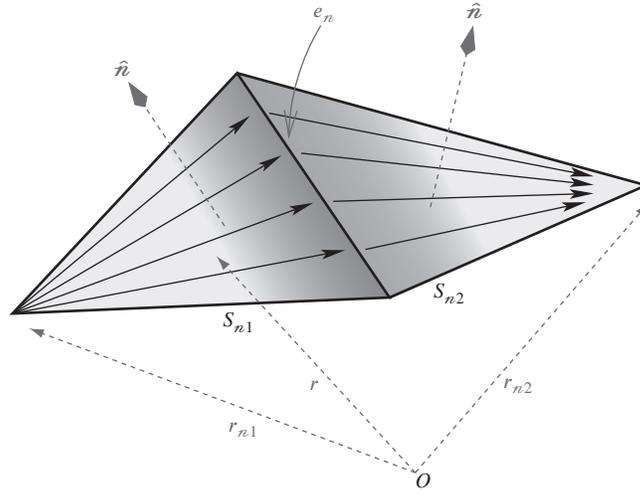
Rao-Wilton-Glisson (RWG) [57] functions. Investigations on PEC objects show that scattered fields obtained with N-MFIE and T-N-CFIE are significantly inaccurate compared to those obtained with T-EFIE [123]–[129]. Similar observations have been made for the solution of dielectric objects [74],[80]. In such cases, it is helpful to improve the discretizations by employing more appropriate, such as higher-order [130]–[133], basis functions in order to obtain accurate results with normal and mixed formulations [74],[127]–[129],[134]–[139]. Recent studies also show that a regularization of the identity operator improves the accuracy of N-MFIE [140].

### 1.7.1 Linear Functions

Linear functions are widely used to discretize surface integral equations. Among such functions, the RWG functions [57] are very popular due to their simplicity, flexibility, and usefulness. Over the years, a plethora of electromagnetics problems involving complicated three-dimensional geometries have been solved with the help of the RWG functions. Spatial distributions of the RWG functions are depicted in Figure 1.8, and can be written as

$$\mathbf{b}_n^{\text{RWG}}(\mathbf{r}) = \begin{cases} \frac{l_n}{2A_{n1}}(\mathbf{r} - \mathbf{r}_{n1}), & \mathbf{r} \in S_{n1} \\ \frac{l_n}{2A_{n2}}(\mathbf{r}_{n2} - \mathbf{r}), & \mathbf{r} \in S_{n2} \\ 0, & \mathbf{r} \notin S_n. \end{cases} \quad (1.73)$$

In (1.73),  $l_n$  represents the length of the main edge  $e_n$ ,  $A_{n1}$  and  $A_{n2}$  are respectively areas of the first ( $S_{n1}$ ) and the second ( $S_{n2}$ ) triangles associated with the edge, and  $S_n = S_{n1} \cup S_{n2}$ . Since they have constant normal (CN) and linear tangential (LT) components at the main edge, the RWG functions can be classified as a CN-LT type. Despite their simplicity, they provide the necessary flexibility to model current and field distributions on arbitrarily complicated surfaces with sufficient accuracy. They are also shown to possess the ability to properly model



**Figure 1.8** Spatial distributions of the RWG functions. Arrows show the direction while the shading indicates the magnitude of the vector distribution. Light and dark colors represent the low and high values, respectively.

the charge distribution, in addition to the current distribution [141]. Therefore, the RWG functions have been extensively used for the discretization of surface integral equations. The RWG functions are divergence conforming and their divergence is finite everywhere, e.g.,

$$\nabla \cdot \mathbf{b}_n^{\text{RWG}}(\mathbf{r}) = \begin{cases} \frac{l_n}{A_{n1}}, & \mathbf{r} \in S_{n1} \\ -\frac{l_n}{A_{n2}}, & \mathbf{r} \in S_{n2} \\ 0, & \mathbf{r} \notin S_n. \end{cases} \quad (1.74)$$

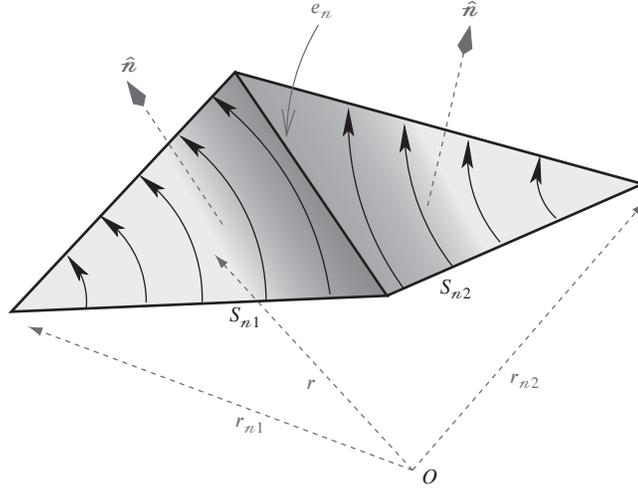
This property can be crucial in some formulations, such as T-EFIE. In addition, the total charge distribution associated with an RWG function is identically zero since

$$\int_{S_n} d\mathbf{r} \nabla \cdot \mathbf{b}_n^{\text{RWG}}(\mathbf{r}) = A_{n1} \frac{l_n}{A_{n1}} - A_{n2} \frac{l_n}{A_{n2}} = 0. \quad (1.75)$$

As extensively discussed in Chapter 2, the RWG functions can be replaced with other simple functions to improve the accuracy of normal and mixed formulations. For example, the curl-conforming  $\hat{\mathbf{n}} \times \text{RWG}$  functions have a spatial distribution shown in Figure 1.9 that can be expressed as

$$\mathbf{b}_n^{\text{nRWG}}(\mathbf{r}) = \hat{\mathbf{n}} \times \mathbf{b}_n^{\text{RWG}}(\mathbf{r}), \quad (1.76)$$

where  $\mathbf{b}_n^{\text{RWG}}(\mathbf{r})$  is the associated RWG function. Obviously, the  $\hat{\mathbf{n}} \times \text{RWG}$  functions are LN-CT (linear normal, constant tangential) type. Implementations of the  $\hat{\mathbf{n}} \times \text{RWG}$  functions can easily be obtained from the implementations of the RWG functions considering the simple rotation relation in (1.76), whereas more efficient implementations can be derived using the curl-conforming property of these functions, as discussed in Chapter 2, Section 2.3.9.



**Figure 1.9** Spatial distribution of the  $\hat{\mathbf{n}} \times \text{RWG}$  functions. Arrows show the direction while the shading indicates the magnitude of the vector distribution. Light and dark colors represent the low and high values, respectively.

In addition to the  $\hat{\mathbf{n}} \times \text{RWG}$  functions obtained via rotation, the RWG functions can be decomposed into the linear-linear (LL) functions [133] to further improve the accuracy of normal and mixed formulations, as detailed in Chapter 2, Section 2.3.10. Spatial distributions of the LL functions are depicted in Figure 1.10, where two kinds of the LL functions are defined on the same edge  $e_n$  with the expressions

$$\mathbf{b}_n^{\text{LL}(1)}(\mathbf{r}) = \begin{cases} \frac{l_n}{4(A_{n1})^2}(\mathbf{r} - \mathbf{r}_{n1}) \cdot [(\mathbf{r}_{n4} - \mathbf{r}_{n1}) \times \hat{\mathbf{n}}](\mathbf{r}_{n3} - \mathbf{r}_{n1}), & r \in S_{n1} \\ \frac{l_n}{4(A_{n2})^2}(\mathbf{r} - \mathbf{r}_{n2}) \cdot [(\mathbf{r}_{n4} - \mathbf{r}_{n2}) \times \hat{\mathbf{n}}](\mathbf{r}_{n3} - \mathbf{r}_{n2}), & r \in S_{n2} \\ 0, & r \notin S_n \end{cases} \quad (1.77)$$

for the LL function of the first kind, and

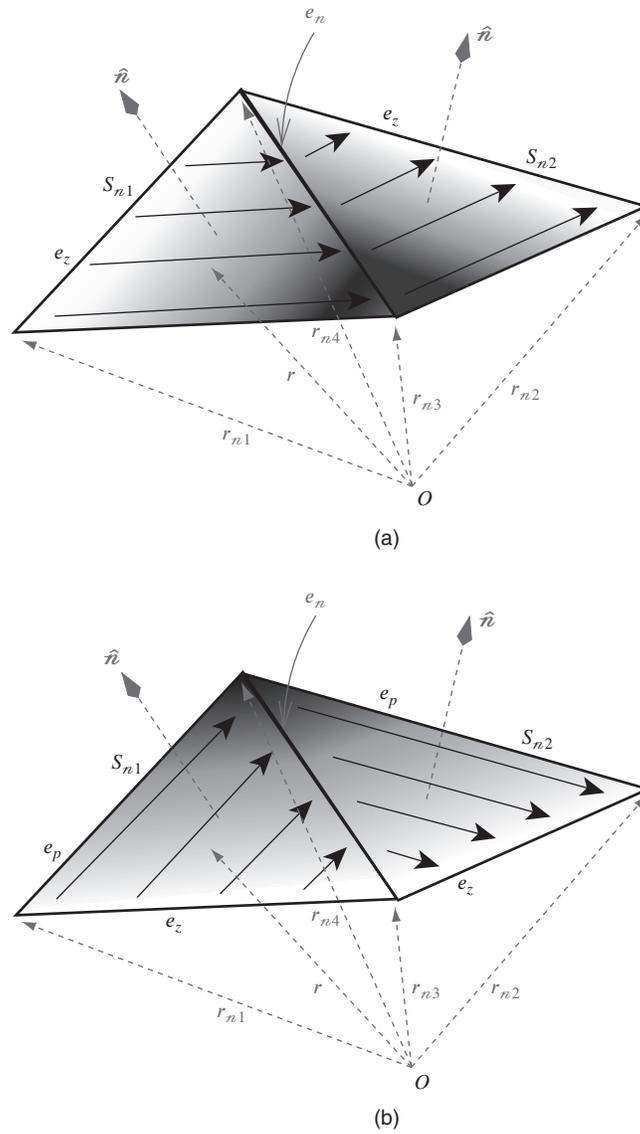
$$\mathbf{b}_n^{\text{LL}(2)}(\mathbf{r}) = \begin{cases} \frac{l_n}{4(A_{n1})^2}(\mathbf{r} - \mathbf{r}_{n1}) \cdot [(\mathbf{r}_{n3} - \mathbf{r}_{n1}) \times \hat{\mathbf{n}}](\mathbf{r}_{n4} - \mathbf{r}_{n1}), & r \in S_{n1} \\ \frac{l_n}{4(A_{n2})^2}(\mathbf{r} - \mathbf{r}_{n2}) \cdot [(\mathbf{r}_{n3} - \mathbf{r}_{n2}) \times \hat{\mathbf{n}}](\mathbf{r}_{n4} - \mathbf{r}_{n2}), & r \in S_{n2} \\ 0, & r \notin S_n \end{cases} \quad (1.78)$$

for the LL function of the second kind. In Figure 1.10, vertex locations  $\mathbf{r}_{n1}$ ,  $\mathbf{r}_{n2}$ ,  $\mathbf{r}_{n3}$ , and  $\mathbf{r}_{n4}$  are labeled as well as the three edges of the triangles  $e_n$ ,  $e_p$ , and  $e_z$  depending on the kind of the LL function. Similar to the  $\hat{\mathbf{n}} \times \text{RWG}$  functions that are derived from the RWG functions, the curl-conforming  $\hat{\mathbf{n}} \times \text{LL}$  functions can be derived from the LL functions as

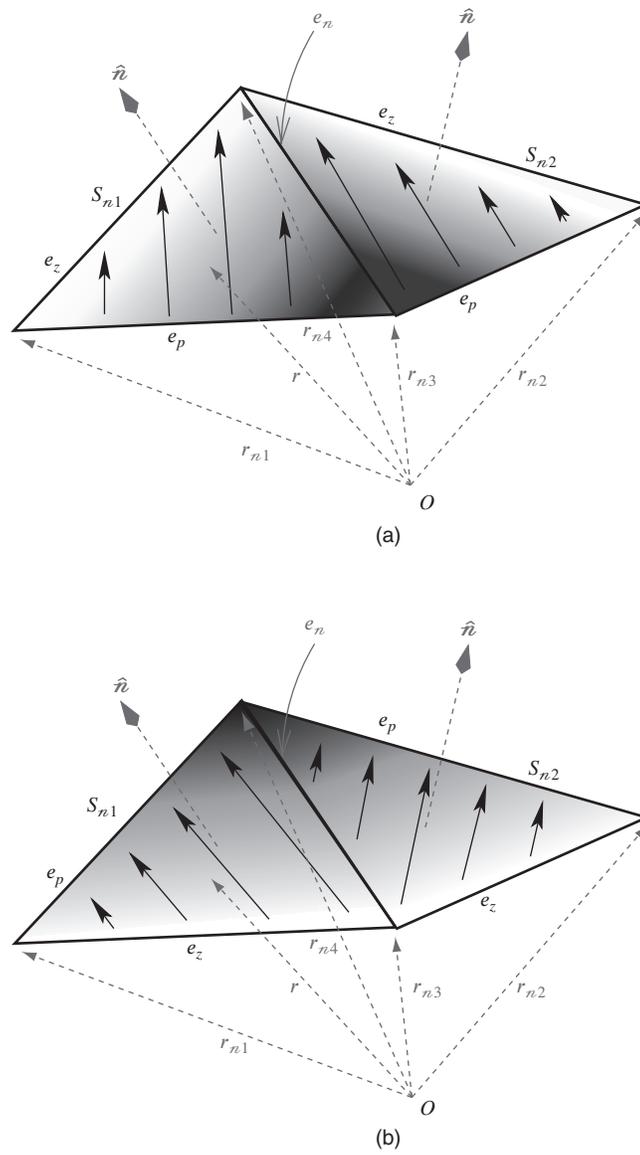
$$\mathbf{b}^{\text{nLL}(1)}(\mathbf{r}) = \hat{\mathbf{n}} \times \mathbf{b}^{\text{LL}(1)}(\mathbf{r}) \quad (1.79)$$

$$\mathbf{b}^{\text{nLL}(2)}(\mathbf{r}) = \hat{\mathbf{n}} \times \mathbf{b}^{\text{LL}(2)}(\mathbf{r}). \quad (1.80)$$

Spatial distributions of the  $\hat{\mathbf{n}} \times \text{LL}$  functions are depicted in Figure 1.11.



**Figure 1.10** (a) First-kind and (b) second-kind LL functions defined at the  $n$ th edge ( $e_n$ ). Arrows show the direction while the shading indicates the magnitude of the vector distribution. Light and dark colors represent the low and high values, respectively, and white corresponds to zero.



**Figure 1.11** (a) First-kind and (b) second-kind  $\hat{n} \times LL$  functions defined at the  $n$ th edge ( $e_n$ ). Arrows show the direction while the shading indicates the magnitude of the vector distribution. Light and dark colors represent the low and high values, respectively, and white corresponds to zero.

In order to understand the relation between the LL and RWG functions (or the  $\hat{\mathbf{n}} \times \text{LL}$  and  $\hat{\mathbf{n}} \times \text{RWG}$  functions), one can consider the general representation of vector functions [130]. The zeroth-order divergence-conforming basis functions for triangular elements, such as the RWG functions, can be written as

$$\mathbf{f}_\beta(\mathbf{r}) = \frac{1}{J_\beta} [\zeta_{\beta+1}(\mathbf{r}) \ell_{\beta-1} - \zeta_{\beta-1}(\mathbf{r}) \ell_{\beta+1}] \quad (\beta = 1, 2, 3), \quad (1.81)$$

where  $J_\beta$  indicates the Jacobian that can be derived for the RWG functions as

$$J_\beta = \frac{2A}{|\ell_\beta|} \quad (1.82)$$

and  $A$  is the area of the triangle. In (1.81),  $\zeta_\beta$  for  $\beta = 1, 2, 3$  represent parent (simplex) coordinates with the dependency relationship

$$\zeta_1(\mathbf{r}) + \zeta_2(\mathbf{r}) + \zeta_3(\mathbf{r}) = 1 \quad (1.83)$$

when  $\mathbf{r}$  is on the triangle. In addition,  $\ell_\beta$  in (1.81) and (1.82) for  $\beta = 1, 2, 3$  represent edge vectors with the dependence

$$\ell_1 + \ell_2 + \ell_3 = 0. \quad (1.84)$$

The RWG functions, which provide three degrees of freedom per triangle, are zeroth-order complete since they can represent any constant vector function on a triangle and their surface divergence is also constant. Given the three RWG functions corresponding to the three edges of a triangle as in (1.81), the set of LL functions associated with the same triangle can be derived as

$$\mathbf{f}_\beta^{(1)}(\mathbf{r}) = \frac{1}{J_\beta} \zeta_{\beta+1}(\mathbf{r}) \ell_{\beta-1} \quad (1.85)$$

$$\mathbf{f}_\beta^{(2)}(\mathbf{r}) = -\frac{1}{J_\beta} \zeta_{\beta-1}(\mathbf{r}) \ell_{\beta+1} \quad (1.86)$$

with the decomposition

$$\mathbf{f}_\beta(\mathbf{r}) = \mathbf{f}_\beta^{(1)}(\mathbf{r}) + \mathbf{f}_\beta^{(2)}(\mathbf{r}) \quad (\beta = 1, 2, 3). \quad (1.87)$$

Note that both of the distributions in (1.81) and (1.85)–(1.86) provide a linear variation for parallel components along the edges  $\beta$ . On the other hand,

$$\nabla \zeta_\beta(\mathbf{r})|_{\zeta_\beta(\mathbf{r})=0} \cdot \mathbf{f}_\beta^{(1)}(\mathbf{r}) = \zeta_{\beta+1}(\mathbf{r}) \quad (1.88)$$

$$\nabla \zeta_\beta(\mathbf{r})|_{\zeta_\beta(\mathbf{r})=0} \cdot \mathbf{f}_\beta^{(2)}(\mathbf{r}) = \zeta_{\beta-1}(\mathbf{r}) \quad (1.89)$$

so that normal components of the LL functions in (1.85) and (1.86) also have a linear variation along edges  $\beta$ , while they are constant for the RWG functions in (1.81), i.e.,

$$\nabla \zeta_\beta(\mathbf{r})|_{\zeta_\beta(\mathbf{r})=0} \cdot \mathbf{f}_\beta(\mathbf{r}) = \zeta_{\beta+1}(\mathbf{r}) + \zeta_{\beta-1}(\mathbf{r}) = 1 - \zeta_\beta(\mathbf{r}) = 1. \quad (1.90)$$

Any linearly varying vector function on a triangle can be represented by a combination of the LL functions since

$$\frac{\mathbf{f}_2^{(1)}(\mathbf{r})}{|\ell_2|} + \frac{\mathbf{f}_2^{(2)}(\mathbf{r})}{|\ell_2|} - \frac{\mathbf{f}_3^{(1)}(\mathbf{r})}{|\ell_3|} - \frac{\mathbf{f}_3^{(2)}(\mathbf{r})}{|\ell_3|} = \frac{\ell_1}{2A} \quad (1.91)$$

$$\frac{\mathbf{f}_3^{(1)}(\mathbf{r})}{|\ell_3|} + \frac{\mathbf{f}_3^{(2)}(\mathbf{r})}{|\ell_3|} - \frac{\mathbf{f}_1^{(1)}(\mathbf{r})}{|\ell_1|} - \frac{\mathbf{f}_1^{(2)}(\mathbf{r})}{|\ell_1|} = \frac{\ell_2}{2A} \quad (1.92)$$

$$\frac{\mathbf{f}_2^{(2)}(\mathbf{r})}{|\ell_2|} - \frac{\mathbf{f}_3^{(1)}(\mathbf{r})}{|\ell_3|} = \frac{\zeta_1(\mathbf{r})\ell_1}{2A} \quad (1.93)$$

$$\frac{\mathbf{f}_3^{(2)}(\mathbf{r})}{|\ell_3|} - \frac{\mathbf{f}_1^{(1)}(\mathbf{r})}{|\ell_1|} = \frac{\zeta_2(\mathbf{r})\ell_2}{2A} \quad (1.94)$$

$$-\frac{\mathbf{f}_3^{(2)}(\mathbf{r})}{|\ell_3|} = \frac{\zeta_2(\mathbf{r})\ell_1}{2A} \quad (1.95)$$

$$\frac{\mathbf{f}_3^{(1)}(\mathbf{r})}{|\ell_3|} = \frac{\zeta_1(\mathbf{r})\ell_2}{2A}. \quad (1.96)$$

On the other hand, the LL functions are not strictly first-order complete since their divergences

$$\nabla \cdot \mathbf{f}_\beta^{(1)}(\mathbf{r}) = \frac{1}{J_\beta} \ell_{\beta-1} \cdot \nabla \zeta_{\beta+1}(\mathbf{r}) = \frac{1}{J_\beta} = \frac{|\ell_\beta|}{2A} \quad (1.97)$$

$$\nabla \cdot \mathbf{f}_\beta^{(2)}(\mathbf{r}) = -\frac{1}{J_\beta} \ell_{\beta+1} \cdot \nabla \zeta_{\beta-1}(\mathbf{r}) = -\frac{1}{J_\beta} = -\frac{|\ell_\beta|}{2A} \quad (1.98)$$

for  $\beta = 1, 2$ , or  $3$  are constants. As indicated in [130], a strictly first-order complete basis requires eight degrees of freedom on the triangle, whereas the LL functions have six degrees of freedom. Nevertheless, the LL functions (and their curl-conforming versions, i.e., the  $\hat{\mathbf{n}} \times \text{LL}$  functions) have the advantage that their implementations do not require higher-order techniques yet they provide significantly higher accuracy than the RWG functions for normal and mixed formulations. This is discussed in Chapter 2, Section 2.3.10.

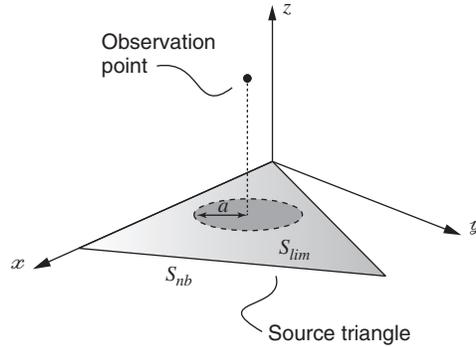
## 1.8 Integrals on Triangular Domains

Elements of the matrices derived from the discretization of surface integral equations involve integrals over the basis and testing functions. Outer integrals over the testing functions can usually be performed via numerical integration techniques, such as Gaussian quadratures [142] and adaptive integration methods, while inner integrals over the basis functions may require singularity extractions [143], [144] before such numerical techniques are applied. In general, considering a basis triangle with a surface of  $S_{nb}$  lying on the  $x$ - $y$  plane, as depicted in Figure 1.12, the integrals to be evaluated can be written as

$$I_{in}^f = \int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') g(\mathbf{r}, \mathbf{r}') \quad (1.99)$$

$$I_{in}^{f\nabla'} = \int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' g(\mathbf{r}, \mathbf{r}'), \quad (1.100)$$

where  $f(\mathbf{r}, \mathbf{r}') = 1$  or  $f(\mathbf{r}, \mathbf{r}') = (x' - x)$  or  $f(\mathbf{r}, \mathbf{r}') = (y' - y)$ . Since the Green's function is singular and its value becomes infinite as  $R = |\mathbf{r} - \mathbf{r}'|$  goes to zero, singularity extraction is required



**Figure 1.12** A source triangle on the  $x$ - $y$  plane and the projection of an observation point.

in (1.99) and (1.100) for accurate computations of the integrals, especially when the observation point is close to the source (basis) triangle. Specifically, the integrals are divided into numerical and analytical parts as

$$I_{in}^f = \frac{1}{4\pi} \int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \left( \frac{\exp(ikR) - 1}{R} \right) + \frac{1}{4\pi} \int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \frac{1}{R} \quad (1.101)$$

$$\begin{aligned} \mathbf{I}_{in}^{f\nabla'} &= \frac{1}{4\pi} \int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' \left( \frac{\exp(ikR) - 1 + 0.5k^2 R^2}{R} \right) \\ &+ \frac{1}{4\pi} \int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' \left( \frac{1}{R} \right) - \frac{k^2}{8\pi} \int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' R. \end{aligned} \quad (1.102)$$

The numerical integrals in (1.101)–(1.102) can be evaluated accurately by using Gaussian quadratures or adaptive integration methods.

### 1.8.1 Analytical Integrals

Consider the scenario depicted in Figure 1.12 involving a source triangle lying on the  $x$ - $y$  plane and an observation point  $\mathbf{r}$ . The analytical integrals can be further divided as

$$\int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \frac{1}{R} = \int_{S_{PV,nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \frac{1}{R} + \int_{S_{lim}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \frac{1}{R} \quad (1.103)$$

$$\begin{aligned} \int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' \left( \frac{1}{R} \right) &= \int_{S_{PV,nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' \left( \frac{1}{R} \right) \\ &+ \int_{S_{lim}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' \left( \frac{1}{R} \right) \end{aligned} \quad (1.104)$$

$$\int_{S_{nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' R = \int_{S_{PV,nb}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' R + \int_{S_{lim}} d\mathbf{r}' f(\mathbf{r}, \mathbf{r}') \nabla' R, \quad (1.105)$$

where  $S_{lim}$  is defined as the infinitesimal circular surface (with radius  $a \rightarrow 0$ ) in the vicinity of the projection of the observation point and  $S_{PV,nb} = S_{nb} - S_{lim}$ . Depending on the location of the observation point,  $S_{lim}$  may not exist. In general, the decompositions in (1.103)–(1.105) allow for rigorous analytical evaluations of the integrals in terms of the geometric quantities associated with the basis triangle and the location of the observation point [144]. These quantities are depicted in Figure 1.13 and can be listed for each edge ( $i = 1, 2, 3$ ) as follows:

- $\hat{u}_i$  represents the outward unit vector perpendicular to the  $i$ th edge on the  $x$ - $y$  plane.
- $R_i^+$  and  $R_i^-$  are the distances between the observation point and the end points of the  $i$ th edge. The superscripts “+” and “–”, referring to the end points of the edge, are determined by the right-hand rule applied on the triangle in the  $z$  direction.
- $R_i^0$  is the distance from the observation point to the  $i$ th edge and can be written as

$$R_i^0 = \sqrt{z^2 + |t_i^0|^2}, \quad (1.106)$$

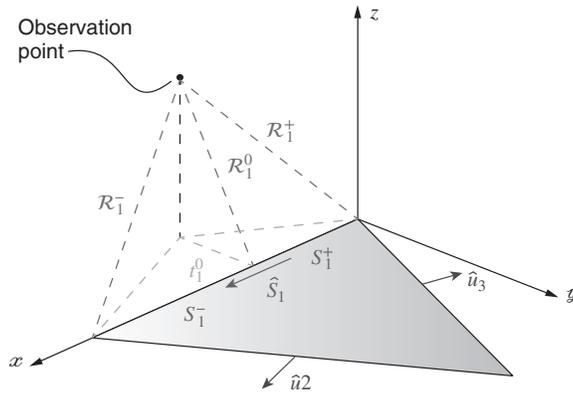
where  $|t_i^0|$  is the distance between the projection of the observation point on the  $x$ - $y$  plane and the  $i$ th edge. The sign of  $t_i^0$  is determined by the position of the projection of the observation point with respect to the edge.

- $s_i^+$  and  $s_i^-$  have magnitudes equal to the “+” and “–” segments of the  $i$ th edge. These segments are formed by the projection of the observation point on the edge. The signs of  $s_i^+$  and  $s_i^-$  are determined by the relative position of this projection compared to the “+” and “–” ends of the edge. In addition,  $\hat{s}_i$  is the unit vector in the direction of the edge using the right-hand rule.

In order to evaluate the integrals in the form of (1.103), one can use the identities

$$\nabla'_s \cdot \left( \frac{R}{P} \hat{P} \right) = \frac{1}{R} \quad (1.107)$$

$$\nabla'_s R = \frac{\mathbf{p}' - \mathbf{p}}{R} = \hat{x} \frac{(x' - x)}{R} + \hat{y} \frac{(y' - y)}{R}, \quad (1.108)$$



**Figure 1.13** Geometric variables introduced to express the results of the analytical integrals. Most of the variables are shown only for the first edge on the  $x$  axis.

where  $\nabla'_S$  represents the surface divergence or gradient in the primed coordinates and

$$P = \hat{P}P = \boldsymbol{\rho} - \boldsymbol{\rho}' = \hat{\mathbf{x}}(x - x') + \hat{\mathbf{y}}(y - y'). \quad (1.109)$$

Then, using the surface divergence theorem and evaluating the limits (as the infinitesimal surface in Figure 1.12 shrinks to zero), one can obtain

$$\int_{S_{nb}} d\mathbf{r}' \frac{1}{R} = \sum_{i=1}^3 t_i^0 f_{2,i} - |z| \sum_{i=1}^3 \beta_i \quad (1.110)$$

$$\int_{S_{nb}} d\mathbf{r}' \frac{(x' - x)}{R} = \frac{1}{2} \hat{\mathbf{x}} \cdot \sum_{i=1}^3 \hat{\mathbf{u}}_i f_{3,i} \quad (1.111)$$

$$\int_{S_{nb}} d\mathbf{r}' \frac{(y' - y)}{R} = \frac{1}{2} \hat{\mathbf{y}} \cdot \sum_{i=1}^3 \hat{\mathbf{u}}_i f_{3,i}, \quad (1.112)$$

where

$$\beta_i = \tan^{-1} \frac{t_i^0 s_i^+}{(R_i^0)^2 + |z| R_i^+} - \tan^{-1} \frac{t_i^0 s_i^-}{(R_i^0)^2 + |z| R_i^-} \quad (1.113)$$

$$f_{2,i} = \ln \left( \frac{R_i^+ + s_i^+}{R_i^- + s_i^-} \right) \quad (1.114)$$

$$f_{3,i} = s_i^+ R_i^+ - s_i^- R_i^- + (R_i^0)^2 f_{2,i}. \quad (1.115)$$

Similar procedures can be used to evaluate the integrals in the form of (1.104), leading to

$$\int_{S_{nb}} d\mathbf{r}' \nabla' \left( \frac{1}{R} \right) = \sum_{i=1}^3 \hat{\mathbf{u}}_i f_{2,i} + \hat{\mathbf{z}} \frac{z}{|z|} \sum_{i=1}^3 \beta_i \quad (1.116)$$

$$\int_{S_{nb}} d\mathbf{r}' (x' - x) \nabla' \left( \frac{1}{R} \right) = \hat{\mathbf{x}} |z| \sum_{i=1}^3 \beta_i - \sum_{i=1}^3 \hat{\mathbf{x}} \cdot \hat{\mathbf{s}}_i f_{1,i} - \hat{\mathbf{z}} z \hat{\mathbf{x}} \cdot \sum_{i=1}^3 \hat{\mathbf{u}}_i f_{2,i} \quad (1.117)$$

$$\int_{S_{nb}} d\mathbf{r}' (y' - y) \nabla' \left( \frac{1}{R} \right) = \hat{\mathbf{y}} |z| \sum_{i=1}^3 \beta_i - \sum_{i=1}^3 \hat{\mathbf{y}} \cdot \hat{\mathbf{s}}_i f_{1,i} - \hat{\mathbf{z}} z \hat{\mathbf{y}} \cdot \sum_{i=1}^3 \hat{\mathbf{u}}_i f_{2,i}, \quad (1.118)$$

where

$$f_{1,i} = \hat{\mathbf{s}}_i t_i^0 f_{2,i} - \hat{\mathbf{u}}_i (R_i^+ - R_i^-). \quad (1.119)$$

For the integrals in the form of (1.105),  $f(\mathbf{r}, \mathbf{r}')$  is usually unity and one can derive

$$\int_{S_{nb}} d\mathbf{r}' \nabla' R = \frac{1}{2} \sum_{i=1}^3 \hat{\mathbf{u}}_i f_{3,i} + \hat{\mathbf{z}} z |z| \sum_{i=1}^3 \beta_i - \hat{\mathbf{z}} z \sum_{i=1}^3 t_i^0 f_{2,i}. \quad (1.120)$$

It is essential to investigate the values of the integrals when the observation point approaches the source triangle. For example,

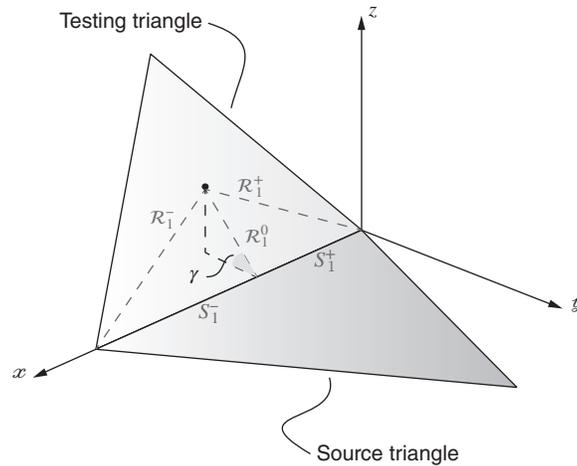
$$\lim_{z \rightarrow 0} \sum_{i=1}^3 \beta_i = \alpha(x, y), \tag{1.121}$$

where

$$\begin{cases} \alpha(x, y) = 0, & \mathbf{r} \notin S_{nb} \\ \alpha(x, y) = 2\pi, & \mathbf{r} \in S_{nb} \text{ and } \mathbf{r} \notin \partial S_{nb} \\ 0 < \alpha(x, y) < 2\pi, & \mathbf{r} \in \partial S_{nb} \end{cases} \tag{1.122}$$

and  $\partial S_{nb}$  represents the bounding contour around the triangle. Interestingly, the value of  $\alpha(x, y)$  at any edge depends on how the observation point approaches the edge. Specifically,  $\alpha \rightarrow 2\gamma$  as  $\mathbf{r} \rightarrow \partial S_{nb}$ , where  $\gamma$  is the angle between the approach path and the  $x$ - $y$  plane, as depicted in Figure 1.14. This angle-dependent factor vanishes in (1.110), (1.117), and (1.118), since  $z \rightarrow 0$  as the observation point approaches the edge. However, it does not vanish in (1.116). In fact, considering this integral, it can be shown that this factor is directly related to the solid-angle factor in the limit part of the  $\mathcal{K}$  operator [125] (see Appendix, Section A.1).

Another important parameter in the analytical integrals is  $f_{2,i}$  given in (1.114). The value of this parameter is infinite when the observation point is at one of the edges of the source triangle. In such a case, the value of the integral in (1.116) is also infinite due to the non-vanishing contribution of  $f_{2,i}$ . Since the singularity is logarithmic and it is mild, it does not cause a serious problem in the numerical integration over the testing triangle, as long as the integration (observation) points are taken inside the testing triangle. It is also possible to apply a singularity-extraction technique for the outer integrals in near-neighboring interactions [124].



**Figure 1.14** A limit case as the observation point approaches the edge of the basis triangle in a near-neighboring interaction.

As an example, consider the scenario in Figure 1.14 and an integral

$$\int_{S_{ma}} d\mathbf{r} f(x, y, z) f_{2,1} = \int_{S_{ma}} d\mathbf{r} f(x, y, z) \ln \left( \frac{R_1^+ + s_1^+}{R_1^- + s_1^-} \right) \quad (1.123)$$

over the testing triangle, where  $f(x, y, z)$  is a linear function. The logarithmic singularity due to the first edge of the basis triangle can be extracted as

$$\begin{aligned} \int_{S_{ma}} d\mathbf{r} f(x, y, z) \ln \left( \frac{R_1^+ + s_1^+}{R_1^- + s_1^-} \right) &= \int_{S_{ma}} d\mathbf{r} \left[ f(x, y, z) \ln \left( \frac{R_1^+ + s_1^+}{R_1^- + s_1^-} \right) + 2 \ln (R_1^0) \right] \\ &\quad - 2 \int_{S_{ma}} d\mathbf{r} f(x, y, z) \ln (R_1^0), \end{aligned} \quad (1.124)$$

where the second integral on the RHS can be evaluated analytically [124].

### 1.8.2 Gaussian Quadratures

Numerical integrations of nonsingular functions on triangles can be performed accurately and efficiently by using Gaussian quadratures [142]. A symmetrical quadrature rule on a triangle with a spatial support of  $S_m$  can be written as

$$\int_{S_m} d\mathbf{r} f(\mathbf{r}) \approx A_m \sum_{p=1}^P w_p f(\mathbf{r}_{pm}), \quad (1.125)$$

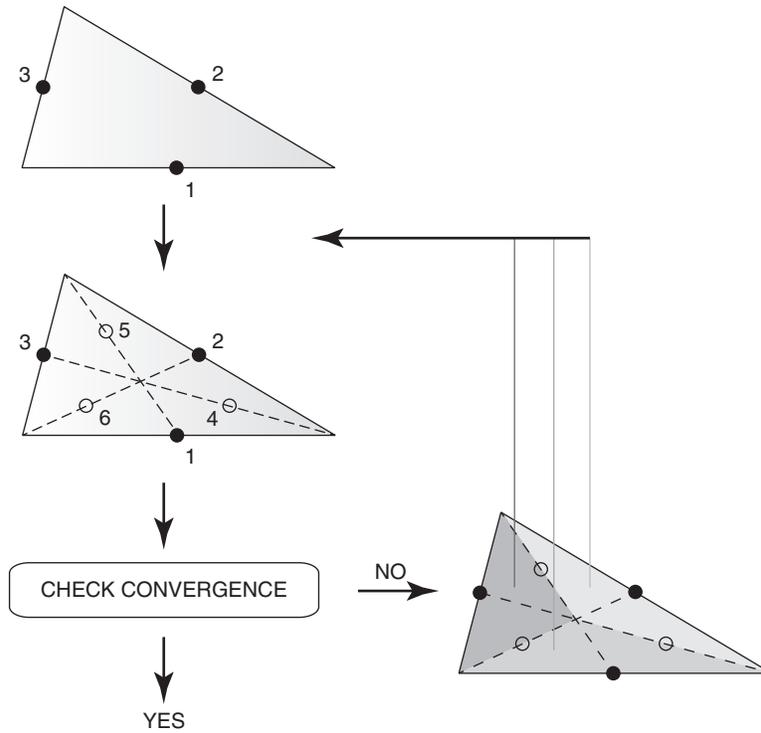
where  $A_m$  is the area of the triangle,  $P$  is the number of quadrature points, and  $w_p$  represents the integration weight for the  $p$ th sample point. Locations of the sample points are determined as

$$\mathbf{r}_{pm} = \zeta_{p1} \mathbf{r}_{m1} + \zeta_{p2} \mathbf{r}_{m2} + \zeta_{p3} \mathbf{r}_{m3}, \quad (1.126)$$

where  $\mathbf{r}_{m1}$ ,  $\mathbf{r}_{m2}$ , and  $\mathbf{r}_{m3}$  are node coordinates of the triangle, whereas  $\zeta_{p1}$ ,  $\zeta_{p2}$ , and  $\zeta_{p3}$  are the simplex coordinates satisfying  $\zeta_{p1} + \zeta_{p2} + \zeta_{p3} = 1$ .

### 1.8.3 Adaptive Integration

In order to reduce the integration error in electromagnetic interactions, the order of the Gaussian quadrature rule applied for the integrals can be increased. Alternatively, low-order Gaussian quadratures can be employed in an adaptive scheme, which can efficiently control the error by adjusting the sampling rate according to the integrand. Specifically, employing an adaptive integration method, the integration points can be selected economically by using more samples at critical locations, where the integrand changes rapidly. Figure 1.15 depicts a simple adaptive integration method based on a three-point Gaussian quadrature rule. The following steps are carried out for the integration of a function  $f$  on  $S_m$ :



**Figure 1.15** An adaptive integration method using a three-point Gaussian quadrature and three subtriangles for each triangle. (See color plate section for the color version of this figure)

- Given a triangle, three integration points are chosen at the middle of the edges. Using a simple Gaussian quadrature rule, the value of the integral is given as

$$\int_{S_m} drf(\mathbf{r}) \approx I_3 = \frac{A_m}{3} [f(\mathbf{r}_{p1}) + f(\mathbf{r}_{p2}) + f(\mathbf{r}_{p3})], \quad (1.127)$$

where  $\mathbf{r}_{p1}$ ,  $\mathbf{r}_{p2}$ , and  $\mathbf{r}_{p3}$  are the integration points.

- Three additional integration points are chosen on the medians. These points are located at 1/3 of the medians nearer to the vertices of the triangle. Using the three-point Gaussian quadrature rule in (1.127), the value of the integral using six points can be written as

$$\int_{S_m} drf(\mathbf{r}) \approx I_6 = \frac{I_3}{3} + \frac{2A_m}{9} [f(\mathbf{r}_{p4}) + f(\mathbf{r}_{p5}) + f(\mathbf{r}_{p6})], \quad (1.128)$$

where  $\mathbf{r}_{p4}$ ,  $\mathbf{r}_{p5}$ , and  $\mathbf{r}_{p6}$  are the new integration points on the medians.

- Three-point and six-point integration values are compared. If the error

$$\Delta_{3-6} = \frac{|I_6 - I_3|}{|I_6|} \quad (1.129)$$

is lower than a given threshold, then  $I_6$  is returned as the value of the integral over the triangle. Otherwise, the three subtriangles are considered separately. Note that in each subtriangle, the integration is carried out by sampling three more points (in addition to the three points that are already sampled) and comparing the three-point and six-point integration values.

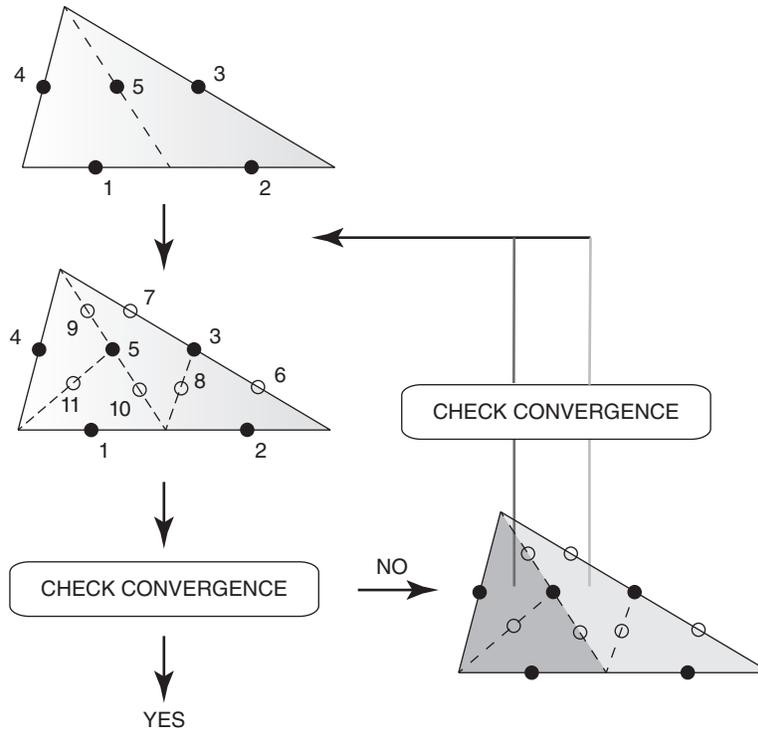
The steps listed above are performed recursively until the convergence is satisfied for all subtriangles. The final distribution of the samples on the main triangle can be nonuniform since the depth of the recursion varies depending on the integrand.

The adaptive integration method described above works well in many cases, but its efficiency may deteriorate for triangles with large aspect ratios. An alternative method is depicted in Figure 1.16, and can be described as follows:

- Given a triangle, five integration points are chosen. Using the three-point Gaussian quadrature rule in (1.127), the value of the integral over the triangle can be written as

$$\int_{S_m} df(\mathbf{r}) \approx I_5 = \frac{A_m}{6} [f(\mathbf{r}_{p1}) + f(\mathbf{r}_{p2}) + f(\mathbf{r}_{p3}) + f(\mathbf{r}_{p4}) + 2f(\mathbf{r}_{p5})]. \quad (1.130)$$

Note that two of the integration points are located at the longest edge of the triangle.



**Figure 1.16** An adaptive integration method using a three-point Gaussian quadrature and two subtriangles for each triangle. (See color plate section for the color version of this figure)

- Six additional integration points are chosen. In order to determine the location of these points, the medians are drawn to the longest edges of both subtriangles. The value of the integral using ten points can be written as

$$\int_{S_m} dr f(\mathbf{r}) \approx I_{10} = \frac{A_m}{12} [f(\mathbf{r}_{p1}) + f(\mathbf{r}_{p2}) + f(\mathbf{r}_{p4}) + f(\mathbf{r}_{p5}) + f(\mathbf{r}_{p6}) + f(\mathbf{r}_{p7}) + 2f(\mathbf{r}_{p8}) + f(\mathbf{r}_{p9}) + f(\mathbf{r}_{p10}) + 2f(\mathbf{r}_{p11})]. \quad (1.131)$$

- The five-point and ten-point integration values are compared. If the error

$$\Delta_{10-5} = \frac{|I_{10} - I_5|}{|I_{10}|} \quad (1.132)$$

is lower than a given threshold, then  $I_{10}$  is returned as the value of the integral over the triangle. Otherwise, the two subtriangles are considered separately.

- Before starting the next recursion cycle, the three-point and five-point integration values are compared for each subtriangle. Note that the required integration points were already sampled in the previous recursion step. The recursion continues for a subtriangle if the convergence is not satisfied.

The adaptive integration method shown in Figure 1.16 is more robust than the method in Figure 1.15. Although there is an additional cost of finding the longest edge of the triangle, this cost can be considered negligible compared to the gain from a more efficient distribution of the integration points. As an example, Figure 1.17 presents the integration of a function on a triangle with node coordinates  $(x_1, y_1) = (0.0, 0.0)$ ,  $(x_2, y_2) = (0.083, 0.0)$ , and  $(x_3, y_3) = (0.042, 0.048, 0.0)$ . The integrand is depicted in Figure 1.17(a) and the integral over the triangle is evaluated adaptively with 1% error. Figures 1.17(b) and 1.17(c) illustrate the integration points when the methods in Figures 1.15 and 1.16 are used, respectively. It can be observed that, using the first method (Figure 1.15), the integration points are not distributed efficiently and there is an accumulation near one of the edges of the triangle. In contrast, using the second method (Figure 1.16), the integration points are distributed efficiently and fewer points are used to compute the integral, but with the same level of accuracy.

## 1.9 Electromagnetic Excitation

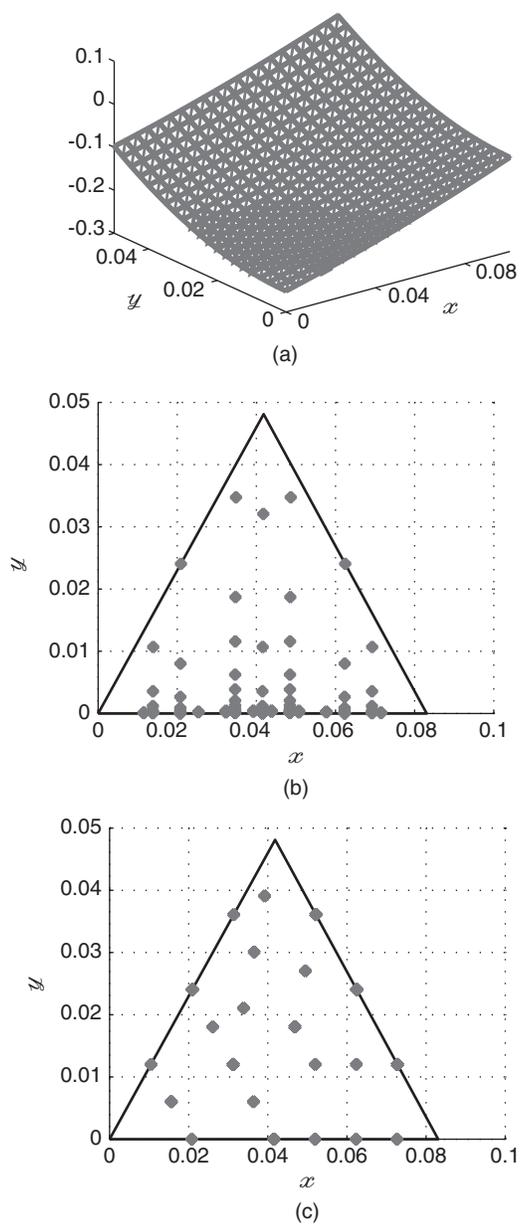
Let  $\bar{\mathbf{Z}} \cdot \mathbf{a} = \mathbf{w}$  be an  $N \times N$  matrix equation constructed using MOM. The excitation vector  $\mathbf{w}$  is obtained by testing the incident electric and magnetic fields, i.e., it contains integrals in the form of

$$\mathbf{v}[m] = \int_{S_m} d\mathbf{r} \left\{ \begin{array}{l} \mathbf{t}_m(\mathbf{r}) \cdot \\ \mathbf{t}_m(\mathbf{r}) \cdot \hat{\mathbf{n}} \times \end{array} \right\} \left\{ \begin{array}{l} \eta^{-1} \mathbf{E}^{inc}(\mathbf{r}) \\ \mathbf{H}^{inc}(\mathbf{r}) \end{array} \right\}, \quad (1.133)$$

where  $S_m$  is the spatial support of the testing function  $\mathbf{t}_m$  for  $m = 1, 2, \dots, N$ . Different types of excitations can be used, as described below.

### 1.9.1 Plane-Wave Excitation

A common excitation used in numerical simulations, e.g., RCS computations, is the plane-wave illumination. A plane wave propagating in the  $\hat{\mathbf{k}}$  direction with the electric field



**Figure 1.17** Numerical integration of a function on a triangle using the adaptive integration methods depicted in Figures 1.15 and 1.16. Figures show (a) the function and the integration points to evaluate the integral with 1% error using (b) the method in Figure 1.15 and (c) the method in Figure 1.16.

polarized in the  $\hat{e}$  direction ( $\hat{e} \perp \hat{k}$ ) can be written as

$$\mathbf{E}^{inc}(\mathbf{r}) = \hat{e} E_a \exp(ik\hat{k} \cdot \mathbf{r}) \quad (1.134)$$

$$\mathbf{H}^{inc}(\mathbf{r}) = \frac{1}{\eta} \hat{k} \times \mathbf{E}^{inc}(\mathbf{r}) = \hat{k} \times \hat{e} \frac{E_a}{\eta} \exp(ik\hat{k} \cdot \mathbf{r}), \quad (1.135)$$

where  $E_a$  is the amplitude of the plane wave.

### 1.9.2 Hertzian Dipole

A Hertzian (ideal) dipole can be used as a source in many simulations. The electric and magnetic fields of a Hertzian dipole with dipole moment  $\mathbf{I}_{DM}$  located at  $\mathbf{r}_d$  can be written as

$$\begin{aligned} \mathbf{E}^{inc}(\mathbf{r}) = i\omega\mu \frac{\exp(ik|\mathbf{r} - \mathbf{r}_d|)}{4\pi|\mathbf{r} - \mathbf{r}_d|} \left[ \mathbf{I}_{DM} \left( 1 + \frac{i}{k|\mathbf{r} - \mathbf{r}_d|} - \frac{1}{k^2|\mathbf{r} - \mathbf{r}_d|^2} \right) \right. \\ \left. - (\mathbf{r} - \mathbf{r}_d) \frac{\mathbf{I}_{DM} \cdot (\mathbf{r} - \mathbf{r}_d)}{|\mathbf{r} - \mathbf{r}_d|^2} \left( 1 + \frac{3i}{k|\mathbf{r} - \mathbf{r}_d|} - \frac{3}{k^2|\mathbf{r} - \mathbf{r}_d|^2} \right) \right] \end{aligned} \quad (1.136)$$

and

$$\mathbf{H}^{inc}(\mathbf{r}) = \mathbf{I}_{DM} \times (\mathbf{r} - \mathbf{r}_d) \frac{\exp(ik|\mathbf{r} - \mathbf{r}_d|)}{4\pi|\mathbf{r} - \mathbf{r}_d|^2} \left( \frac{1}{|\mathbf{r} - \mathbf{r}_d|} - ik \right), \quad (1.137)$$

respectively. Singularity extraction methods are required to evaluate the integrals in (1.133) when the dipole is close to the observation point and  $|\mathbf{r} - \mathbf{r}_d|$  is small.

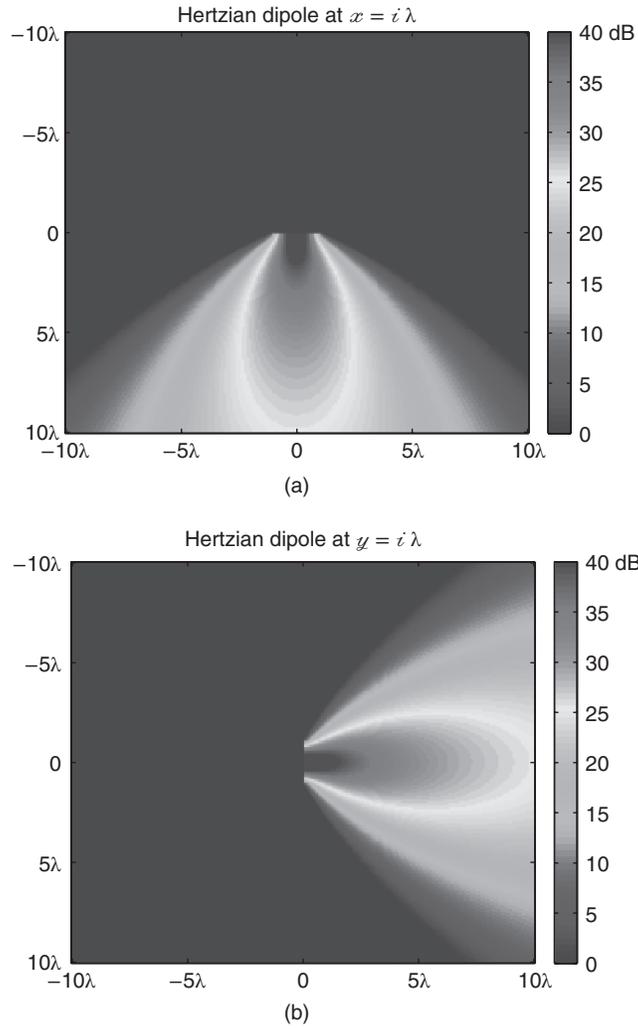
### 1.9.3 Complex-Source-Point Excitation

In order to obtain an electromagnetic beam, which satisfies Maxwell's equations exactly, an ideal dipole can be located at  $\mathbf{r}_d = \mathbf{r}_{d,R} + i\mathbf{r}_{d,I}$  in complex coordinates [145]. Then, the expressions in (1.136) and (1.137) can be used to compute the electric and magnetic fields at any point  $\mathbf{r}$ , where

$$|\mathbf{r} - \mathbf{r}_d| = \sqrt{(\mathbf{r} - \mathbf{r}_d) \cdot (\mathbf{r} - \mathbf{r}_d)^*}. \quad (1.138)$$

The real part of the source location ( $\mathbf{r}_{d,R}$ ) determines the position of the beam in real coordinates, whereas the imaginary part ( $\mathbf{r}_{d,I}$ ) determines the direction of the beam and the width of the beam waist.

Figures 1.18 and 1.19 present some experiments involving a Hertzian dipole with  $\mathbf{I}_{DM} = \hat{z}$  in complex coordinates. Figure 1.18 depicts the normalized electric field, i.e.,  $|\mathbf{E}(\mathbf{r})| \exp(-k|\mathbf{r}_{d,I}|)$ , on the  $x$ - $y$  plane, when  $\mathbf{r}_d = i\lambda\hat{x}$  and  $\mathbf{r}_d = i\lambda\hat{y}$ . Since  $\mathbf{r}_{d,R} = 0$ , both beams are located at the origin. However, the beam propagates in the positive  $x$  and  $y$  directions when  $\mathbf{r}_{d,I} = \lambda\hat{x}$  and  $\mathbf{r}_{d,I} = \lambda\hat{y}$ , respectively. Figure 1.19 presents two different beams in the  $x$  direction with  $\mathbf{r}_d = 2i\lambda\hat{x}$  and  $\mathbf{r}_d = 10i\lambda\hat{x}$ . Note that increasing the value of  $|\mathbf{r}_{d,I}|$  increases the width of the beam waist.

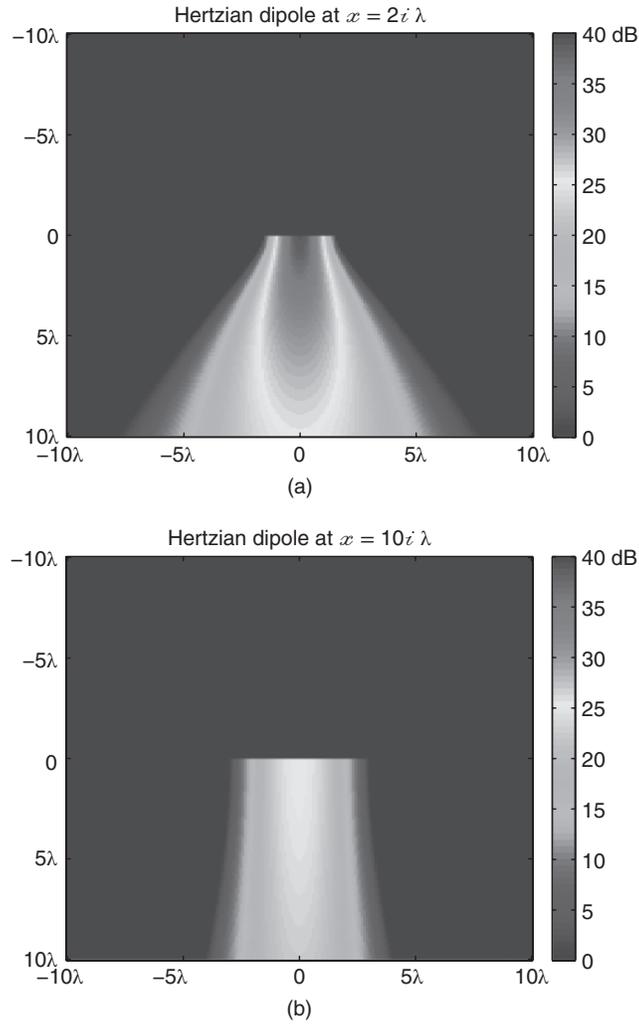


**Figure 1.18** The normalized electric field on the  $x$ - $y$  plane when a Hertzian dipole with  $\mathbf{I}_{DM} = \hat{\mathbf{z}}$  is located at (a)  $\mathbf{r}_d = i\lambda\hat{\mathbf{x}}$  and (b)  $\mathbf{r}_d = i\lambda\hat{\mathbf{y}}$ . (See color plate section for the color version of this figure)

#### 1.9.4 Delta-Gap Excitation

In some applications involving metallic objects, voltage sources are required as external excitations. A voltage source can be modeled as a delta-gap excitation, as depicted in Figure 1.20. A local electric field is defined inside an infinitely narrow opening between two coplanar triangles of the discretization, i.e.,

$$\mathbf{E}^{inc}(\mathbf{r}) = I_n \lim_{d \rightarrow 0} \hat{\mathbf{u}} \delta(\mathbf{r}, \mathbf{r}_n) / d, \quad (1.139)$$

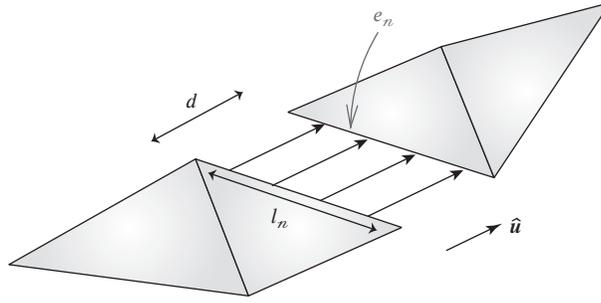


**Figure 1.19** The normalized electric field on the  $x$ - $y$  plane when a Hertzian dipole with  $\mathbf{I}_{DM} = \hat{\mathbf{z}}$  is located at (a)  $\mathbf{r}_d = 2i\lambda\hat{\mathbf{x}}$  and (b)  $\mathbf{r}_d = 10i\lambda\hat{\mathbf{x}}$ . (See color plate section for the color version of this figure)

where  $\mathbf{r}_n$  represents any point at the  $n$ th edge,  $\hat{\mathbf{u}}$  is the unit vector perpendicular to the edge in the plane of the triangles,  $I_n$  is a complex coefficient to represent the strength of the feed, and  $d$  is the width of the theoretical gap. In the limit  $d \rightarrow 0$ , the gap shrinks to the edge  $e_n$ . The Dirac delta function  $\delta(\mathbf{r}, \mathbf{r}_n)$  in (1.139) indicates that  $\mathbf{E}^{inc}(\mathbf{r})$  is zero outside the small gap.

As an example, consider T-EFIE and let a delta-gap excitation at the  $n$ th edge be tested by the RWG functions. Using (1.139), one can obtain

$$\mathbf{w}^{\text{T-EFIE}}[m] = -\eta^{-1} \mathbf{v}_E^T[m] = -\eta^{-1} \int_{S_m} d\mathbf{r} \mathbf{t}_m(\mathbf{r}) \cdot \mathbf{E}^{inc}(\mathbf{r}) = \pm \frac{I_n l_n}{\eta} \delta[m, n], \quad (1.140)$$



**Figure 1.20** A delta-gap source defined at the  $n$ th edge.

where  $l_n$  represents the length of the edge and

$$\delta[m, n] = \begin{cases} 1, & m = n \\ 0, & m \neq n. \end{cases} \quad (1.141)$$

When using a single delta-gap excitation, only one element ( $m = n$ ) of the excitation vector (1.133) is nonzero.

### 1.9.5 Current-Source Excitation

Similar to voltage sources, current sources are commonly used to excite metallic structures. A current source and a corresponding sink can be placed at two different edges, with an electrical connection established mathematically. As opposed to the delta-gap excitation, a current source can be placed at a physical edge that is not necessarily shared by two triangles.

As an example, consider again T-EFIE discretized with the RWG functions, and let a current source and a current sink be placed at two physical edges  $n_1$  and  $n_2$ , respectively. Then, one can define two half RWG functions at  $n_1$  and  $n_2$ . The resulting matrix equation can be written as

$$\sum_{n=1}^{N+2} \bar{\mathbf{Z}}^{\text{T-EFIE}}[m, n] \mathbf{a}[n] = \mathbf{w}^{\text{T-EFIE}}[m] = 0 \quad (m = 1, \dots, N), \quad (1.142)$$

where the dimension of the equation is increased to  $N \times (N + 2)$  due to the extra half functions. Also note that  $\mathbf{w}^{\text{T-EFIE}}[m] = 0$  since there is no incident-field excitation in this case and the integral for the RHS evaluates to zero. On the other hand, the coefficients corresponding to  $n_1$  and  $n_2$  are fixed, i.e.,

$$\mathbf{a}[n_1] = I_n \quad (1.143)$$

$$\mathbf{a}[n_2] = -I_n. \quad (1.144)$$

In other words, the expansion coefficients  $\mathbf{a}[n_1]$  and  $\mathbf{a}[n_2]$  are forced to be  $\pm I_n$  to simulate the source and sink. By setting the two coefficients as above, one can solve the system

$$\sum_{\substack{n=1 \\ n \neq n_1, n_2}}^{N+2} \bar{\mathbf{Z}}^{\text{T-EFIE}}[m, n] \mathbf{a}[n] = -I_n \bar{\mathbf{Z}}^{\text{T-EFIE}}[m, n_1] + I_n \bar{\mathbf{Z}}^{\text{T-EFIE}}[m, n_2] \quad (m = 1, \dots, N), \quad (1.145)$$

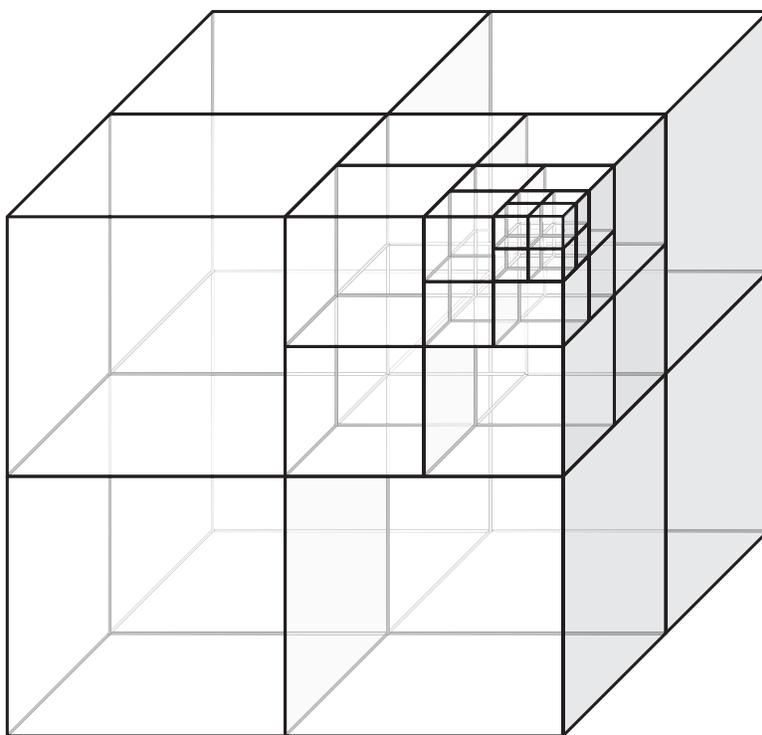
to determine the coefficients  $\mathbf{a}[n]$  for  $n \neq n_1, n_2$ .

### 1.10 Multilevel Fast Multipole Algorithm

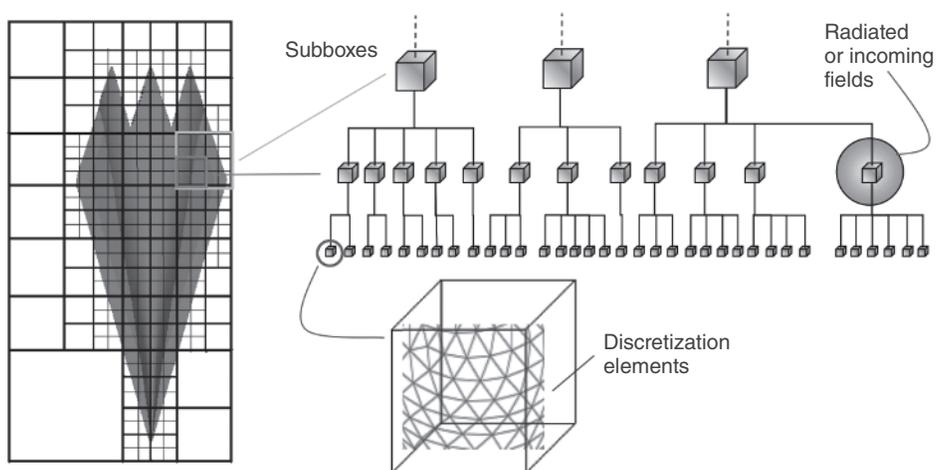
The fast multipole method (FMM) was developed for efficient solutions of radiation and scattering problems in electromagnetics [146],[147]. Discretizations of integral-equation formulations lead to dense matrix equations involving  $\mathcal{O}(N)$  unknowns that can be solved iteratively via a Krylov-subspace algorithm. Iterative solutions require MVMs, i.e.,  $\mathbf{y} = \bar{\mathbf{Z}} \cdot \mathbf{x}$ , which can be performed directly with  $\mathcal{O}(N^2)$  complexity. For large-scale problems, direct multiplications render the solution impossible with limited computational resources. Despite that, FMM provides an efficient way of performing the MVMs required by the iterative algorithms in  $\mathcal{O}(N^{3/2})$  time using  $\mathcal{O}(N^{3/2})$  memory. By reducing the computational complexity from  $\mathcal{O}(N^2)$  to  $\mathcal{O}(N^{3/2})$ , FMM enables the solution of large-scale problems on relatively inexpensive computing platforms [148],[149]. The clustering idea of FMM can be extended and applied in a recursive manner, leading to MLFMA [35], which enables the solution of even larger problems by reducing the complexity of MVMs to  $\mathcal{O}(N \log^2 N)$  [150] or  $\mathcal{O}(N \log N)$  [1], [68], [151]–[157].

In conventional implementations of MLFMA, a tree structure of  $L = \mathcal{O}(\log N)$  levels is constructed by placing the object in a cubic box and recursively dividing the computational domain into subdomains. This recursive clustering is illustrated in Figure 1.21, where a box is divided into eight identical boxes and one of those smaller boxes is further divided into eight boxes, etc. Practically, only nonempty boxes are considered and divided into subboxes, whereas empty boxes created at any stage are simply omitted. The recursion depth depends on the problem as well as the desired accuracy; but in general, the size of the smallest boxes at the lowest level is in the order of the wavelength. As depicted in Figure 1.22, nonempty boxes and the recursive relationship between them are used to construct a multilevel tree structure. In this structure, each box at the lowest level contains some of the discretization elements and each box at the higher levels contains maximum eight subboxes. In order to calculate the far-field interactions, radiated and incoming fields are defined and sampled on the unit sphere.

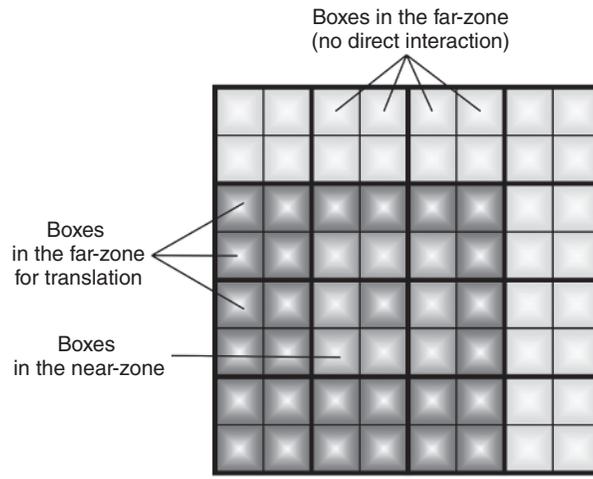
MLFMA calculates the interactions between the radiating (basis) and receiving (testing) elements that are far from each other in a group-by-group manner consisting of three stages, namely, aggregation, translation, and disaggregation. In each MVM, these stages are performed on the tree structure in a multilevel scheme. The group-by-group calculation of the interactions is based on the factorization and diagonalization of the Green's function, as detailed in Chapter 3, Section 3.1. At this point, it is essential to define the near and far zones for the boxes. A common approach is to use the one-box-buffer scheme illustrated in Figure 1.23. Considering the red box at an arbitrary level in this two-dimensional picture, the blue boxes (which are touching the red box) are assumed to be in the near-zone. In a three-dimensional clustering, there are 27 near-zone boxes, some of which may be empty; all



**Figure 1.21** Recursive clustering based on dividing each box into eight subboxes.



**Figure 1.22** Recursive clustering of an arbitrary object and the construction of the multilevel tree structure.



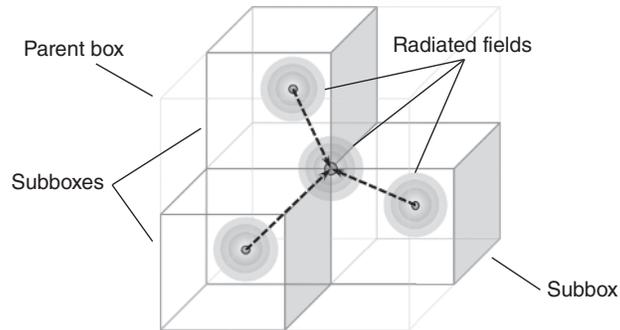
**Figure 1.23** Boxes in the near and far zones for a given box (red) according to the one-box-buffer scheme. (See color plate section for the color version of this figure)

other boxes are considered to be in the far zone. Among these far-zone boxes, only some of them (i.e., the magenta boxes, whose parents are in the near zone of the parent of the red box) are directly interacting with the red box through translations, while the remaining interactions are implicitly included at the higher levels.

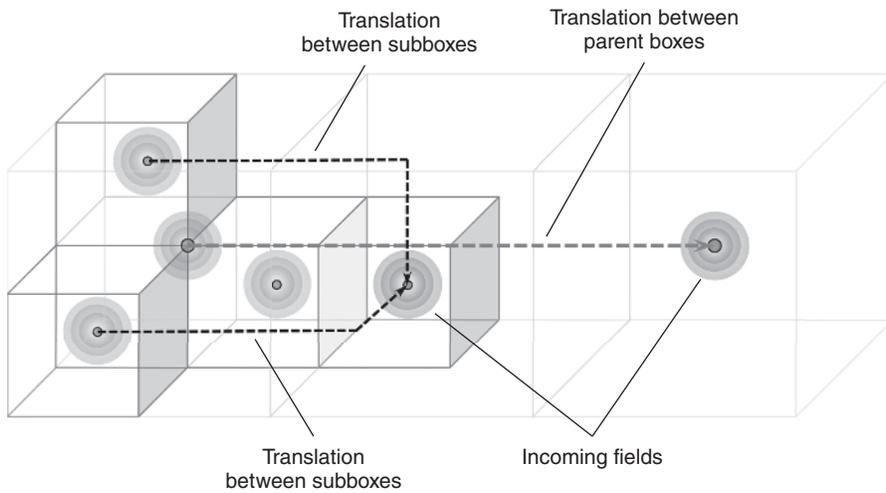
The major operations that are performed in the three stages of MLFMA are illustrated in Figures 1.24–1.26, and can be summarized as follows.

- Aggregation (Figure 1.24): Radiated fields of boxes are calculated from the lowest level of the tree structure to the highest level. At the lowest level (bottom), the radiated field of a box is obtained by combining radiation patterns of the basis functions inside the box. At the higher levels, the radiated field of a box is the combination of the radiated fields of its subboxes.
- Translation (Figure 1.25): Radiated fields computed during the aggregation stage are translated into incoming fields. For each box at any level, there are  $\mathcal{O}(1)$  boxes to translate the radiated field to. Using identical cubic boxes and the one-box-buffer scheme, there is a maximum of  $7^3 - 3^3 = 316$  different translations at each level, independent of the number of boxes [158]. Examples of translations performed at two consecutive levels are depicted.
- Disaggregation (Figure 1.26): Total incoming fields at the box centers are calculated from the top of the tree structure to the bottom. As depicted, the total incoming field for a box is obtained by combining the incoming fields due to translations and the incoming field from the center of its parent box, if it exists. At the lowest level, incoming fields are received by testing functions.

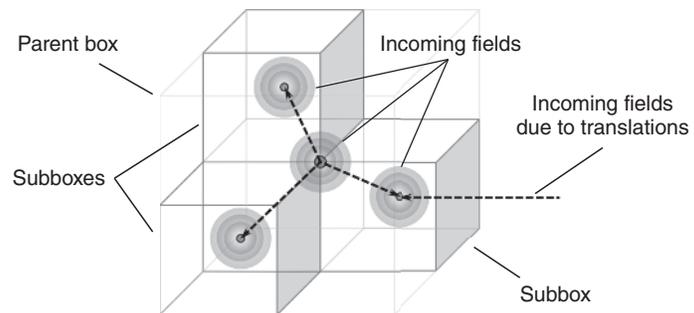
In MLFMA, radiated and incoming fields are sampled on the unit sphere with respect to spherical coordinates  $\theta$  and  $\phi$ . The number of samples required for each box is proportional to the size of the box, as measured by the wavelength [159]. Therefore, to match different sampling rates of consecutive levels, interpolation and transpose interpolation (anterpolation)



**Figure 1.24** Operations in an aggregation stage for a parent box involving three subboxes.



**Figure 1.25** Operations in a translation stage for some boxes in two consecutive levels. (See color plate section for the color version of this figure)



**Figure 1.26** Operations in a disaggregation stage for a parent box involving three subboxes.

[160] are required during the aggregation and disaggregation stages, respectively. There are two major ways of implementing interpolations (and antepolations), namely, through global and local interpolation methods. Global interpolations are usually based on FFT along the  $\phi$  direction and the Legendre transform along the  $\theta$  direction, performed via one-dimensional FMM [150],[161]. Using a uniform sampling, FFT can also be used along the  $\theta$  direction [162]. Such an MLFMA implementation has  $\mathcal{O}(N \log^2 N)$  time complexity, while interpolations/antepolations are performed without error, provided that the Nyquist criterion is applied for the sampling rate. On the other hand, local interpolation methods introduce controllable errors [163]–[165], and such methods lead to more efficient MLFMA implementations with  $\mathcal{O}(N \log N)$  complexity.

To be employed repeatedly, translation operators must usually be calculated and stored in memory before iterations. A direct calculation of these operators requires  $\mathcal{O}(N^{3/2})$  operations, and the processing time for their setup may become substantial as the problem size grows. As a remedy, a two-step computation can be used based on the interpolation of the translation operator [166],[167]. First, the translation operator is expressed as a band-limited function and it is sampled at maximum  $\mathcal{O}(N)$  points. Second, the operator is evaluated at the required points by interpolation from the previous samples. With an efficient interpolation algorithm, the processing time for the calculation of the translation operators can be reduced to  $\mathcal{O}(N)$ , as detailed in Chapter 3, Section 3.3.4.

In MLFMA, there are also  $\mathcal{O}(N)$  near-field interactions, which are calculated directly in the setup stage of the program and stored in memory to be used multiple times during iterations. These interactions are between the basis and testing functions that are near to each other, i.e., in the same box or in two touching boxes at the lowest level using the one box-buffer scheme. Near-field interactions can be calculated accurately by using singularity extraction techniques [57],[143],[168],[169], adaptive integration methods, and Gaussian quadratures [142] (see this chapter, above, Section 1.8).

In order to achieve increased efficiency and reduced complexity, MLFMA employs several ways of gradual elimination of redundancies in the computations. Due to their gradual nature, such techniques may be interpreted as controllable error sources [1],[159],[166],[170]–[173]. For example, truncation of an infinite series and the angular integration over the unit sphere are two error sources of controllable error, which are also common to FMM. Interpolation and antepolation operations constitute the third error source introduced by MLFMA. In addition, other errors arise due to MOM, such as the modeling of the geometry, discretization of integral equations, and numerical integrations on spatial basis and testing functions.

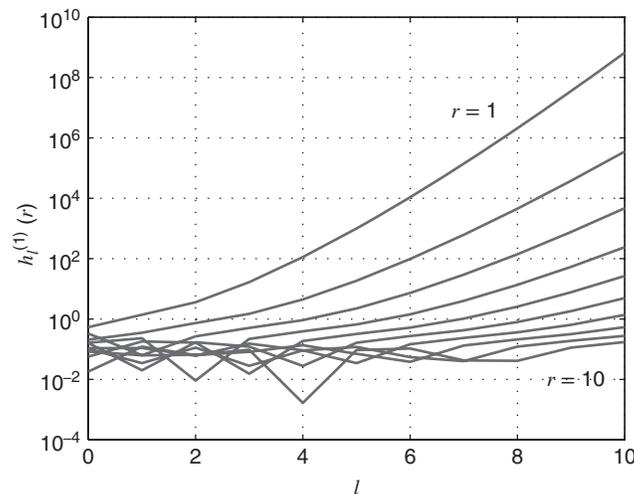
### 1.11 Low-Frequency Breakdown of MLFMA

MLFMA is an efficient algorithm to solve large-scale electromagnetics problems discretized with large numbers of unknowns. However, it suffers from the low-frequency breakdown, i.e., MLFMA boxes that are small compared to the wavelength cause errors. This is because MLFMA is based on the plane wave expansion, which becomes invalid for short distances. Therefore, MLFMA becomes inefficient for problems involving small objects discretized with large numbers of unknowns. Specifically, when ordinary MLFMA is applied to those low-frequency problems, the lowest-level boxes may involve many discretization elements. This significantly increases the processing time and memory required for the near-field interactions, which must be calculated directly. Even the complexity of the MLFMA

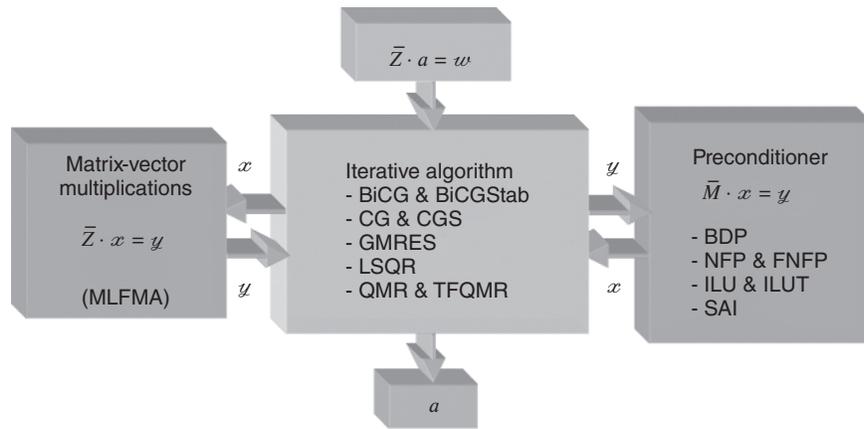
implementation can be more than  $\mathcal{O}(N \log N)$  due to excessively large numbers of near-field interactions.

In the literature, there are two common approaches to solve low-frequency problems efficiently via (modified versions of) MLFMA. The first approach is based on a spectral representation of the Green's function, where radiated fields are divided into propagating and evanescent parts [174]–[177]. Evanescent waves are employed to compute the interactions between small boxes separated by short distances. The second approach is based on the multipole representation of radiated fields [178]–[183]. As also detailed in Chapter 3, Section 3.7, the Green's function is factorized in a series of multipoles, but the multipoles are not converted into plane waves. In both approaches, box size is not restricted, and one can divide the object into boxes that can be much smaller than the wavelength. There are also new studies using alternative expansion methods for stable implementations at low frequencies [184],[185].

Obviously, diagonalization is the major source for the low-frequency breakdown of MLFMA. As a heuristic explanation, plane waves introduced by diagonalization cannot accurately represent electromagnetic fields of localized sources in small volumes (e.g., boxes). For a rigorous analysis, one should consider the factorized and diagonalized forms of the Green's function. As detailed in Chapter 3, Section 3.1, the factorization of the homogeneous-space Green's function is represented by an infinite summation involving the spherical Bessel and Hankel functions of the first kind, i.e.,  $j_l(r)$  and  $h_l^{(1)}(r)$ . In the factorized form,  $j_l(r)$  and  $h_l^{(1)}(r)$  balance each other, even for localized sources. Further stabilization is also possible by scaling  $j_l(r)$  and  $h_l^{(1)}(r)$  with the size of the source [183]. Applying the diagonalization, however, spherical waves are converted into plane waves and the Green's function is represented by an angular integration involving diagonal shifts and a diagonal translation operator. Unfortunately, in this form, the translation operator becomes unstable for small distances. Specifically, this operator is a summation of the spherical Hankel functions multiplied with the Legendre polynomials. As depicted in Figure 1.27, the spherical Hankel



**Figure 1.27** Magnitude of the imaginary part of the spherical Hankel function of the first kind  $h_l^{(1)}(r)$  with respect to  $l$  for different values of  $r$  from 1 to 10.



**Figure 1.28** Iterative solutions of dense matrix equations derived from integral equations.

functions grow rapidly for small arguments. Hence, the Green's function is to be obtained by adding and subtracting large numbers, which can be prone to numerical problems, depending on the precision used. As a trivial but rigorous solution, avoiding diagonalization for small boxes in MLFMA effectively prevents the low-frequency breakdown, as also detailed in Chapter 3, Section 3.7.

## 1.12 Iterative Algorithms

Figure 1.28 depicts how MLFMA is used to solve large-scale electromagnetics problems. Specifically, an iterative algorithm performs iterative solutions by employing MLFMA, which provides an efficient means of executing the required MVMs. Note that the iterative algorithm also interacts with a preconditioner, as detailed in this chapter, below, Section 1.13. In the literature, various Krylov-subspace methods are available to solve electromagnetics problems formulated with integral equations [186]–[196]. Some of these methods are listed below:

- BiCG: Biconjugate gradient
- BiCGStab: Biconjugate gradient stabilized [188]
- CG: Conjugate gradient
- CGS: Conjugate gradient squared [189]
- GMRES: Generalized minimal residual [190]
- LSQR: Least-squares QR [191]–[193]
- MINRES: Minimal residual
- QMR: Quasi-minimal residual [194]
- TFQMR: Transpose-free quasi-minimal residual [195]

In all algorithms, the solution is (implicitly) expanded in a Krylov subspace, i.e.,

$$a \in \text{Span}\{w, \bar{Z} \cdot w, \bar{Z}^2 \cdot w, \dots, \bar{Z}^j \cdot w\}. \quad (1.146)$$

For a given problem, the performance of each algorithm (the number of iterations  $j$ ) may vary significantly, depending on the shape of the geometry, discretization, and the type of formulation; there is no one algorithm that performs best in all cases. Since reducing the number of iterations is extremely important to obtain efficient solutions, one should investigate and compare the iterative solutions provided by various algorithms [197],[198]. Sequential and parallel implementations of iterative algorithms can be found in [186] and [199]. Chapter 3, Section 3.6 includes many examples and comparisons of iterative methods for solving electromagnetics problems.

The Krylov-subspace methods listed above can be investigated in the context of matrix diagonalization.

### 1.12.1 Symmetric Lanczos Process

Consider a matrix equation

$$\bar{\mathbf{Z}} \cdot \mathbf{a} = \mathbf{w}, \quad (1.147)$$

where  $\bar{\mathbf{Z}}$  is an  $N \times N$  symmetric matrix. One can define

$$\beta^{(1)} \mathbf{v}^{(1)} = \mathbf{w} \quad (1.148)$$

and construct

$$\beta^{(k+1)} \mathbf{v}^{(k+1)} = \bar{\mathbf{Z}} \cdot \mathbf{v}^{(k)} - \alpha^{(k)} \mathbf{v}^{(k)} - \beta^{(k)} \mathbf{v}^{(k-1)}, \quad (1.149)$$

where

$$\alpha^{(k)} = [\mathbf{v}^{(k)}]^H \cdot \bar{\mathbf{Z}} \cdot \mathbf{v}^{(k)} - [\mathbf{v}^{(k)}]^H \cdot \beta^{(k)} \mathbf{v}^{(k-1)} \quad (1.150)$$

and  $\beta^{(k)} \geq 0$  are chosen appropriately such that  $\|\mathbf{v}^{(k)}\|_2 = 1$ . In (1.150), the superscript ' $H$ ' represents the Hermitian operation. The recursion can be rewritten as

$$\begin{aligned} [\bar{\mathbf{Z}}]_{N \times N} \cdot [\bar{\mathbf{V}}]_{N \times k} &= [\bar{\mathbf{V}}]_{N \times k} \cdot [\bar{\mathbf{\Delta}}]_{k \times k} \\ &+ \beta^{(k+1)} [0 \quad 0 \quad \dots \quad 0 \quad \mathbf{v}^{(k+1)}]_{N \times k}, \end{aligned} \quad (1.151)$$

where

$$\bar{\mathbf{V}} = [\mathbf{v}^{(1)} \quad \mathbf{v}^{(2)} \quad \dots \quad \mathbf{v}^{(k)}]_{N \times k} \quad (1.152)$$

is a unitary matrix, i.e.,

$$\bar{\mathbf{V}}^H \cdot \bar{\mathbf{V}} = \bar{\mathbf{I}}. \quad (1.153)$$





BiCG and QMR methods, as well as their transpose-free variants, i.e., BiCGStab, CGS, and TFQMR, are based on the nonsymmetric Lanczos process. For example, in BiCG, a subproblem similar to (1.155) is solved via an LU factorization in iterative methods factorization, i.e.,

$$\bar{\mathbf{\Delta}}_{k \times k} = \bar{\mathbf{L}}_{k \times k} \cdot \bar{\mathbf{D}}_{k \times k} \cdot \bar{\mathbf{R}}_{k \times k}. \quad (1.169)$$

The QMR method is similar to MINRES, and the norm of

$$\begin{bmatrix} \bar{\mathbf{\Delta}} & \\ 0 & \dots & 0 & \delta^{(k)} \end{bmatrix}_{(k+1) \times k} \cdot \mathbf{y}^{(k)} - \begin{bmatrix} \delta^{(1)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(k+1) \times 1} \quad (1.170)$$

is minimized by using a QR factorization.

### 1.12.3 Arnoldi Process

The Arnoldi process involves a recursion in the form of

$$\begin{aligned} [\bar{\mathbf{Z}}]_{N \times N} \cdot [\bar{\mathbf{V}}]_{N \times k} &= [\bar{\mathbf{V}}]_{N \times k} \cdot [\bar{\mathbf{H}}]_{k \times k} \\ &+ \beta^{(k+1)} [0 \ 0 \ \dots \ 0 \ \mathbf{v}^{(k+1)}]_{N \times k}, \end{aligned} \quad (1.171)$$

$$\begin{aligned} [\bar{\mathbf{Z}}]_{N \times N} \cdot [\bar{\mathbf{V}}]_{N \times k} &= [\bar{\mathbf{V}}]_{N \times k} \cdot [\bar{\mathbf{H}}]_{k \times k} \\ &+ \beta^{(k+1)} [0 \ 0 \ \dots \ 0 \ \mathbf{v}^{(k+1)}]_{N \times k}, \end{aligned} \quad (1.172)$$

where  $\bar{\mathbf{V}} = [\mathbf{v}_n^{(1)} \ \mathbf{v}_n^{(2)} \ \dots \ \mathbf{v}_n^{(k)}]$  is again a unitary matrix. As opposed to the matrix obtained by a tridiagonalization,  $\bar{\mathbf{H}}$  is a Hessenberg matrix. In general, a subproblem similar to (1.155) is considered for the solution. The efficient GMRES method is based on the Arnoldi process.

### 1.12.4 Golub-Kahan Process

For a general  $N \times N$  matrix equation in the form of (1.147), the Golub-Kahan process starts with

$$\beta^{(1)} \mathbf{x}^{(1)} = \mathbf{w} \quad (1.173)$$

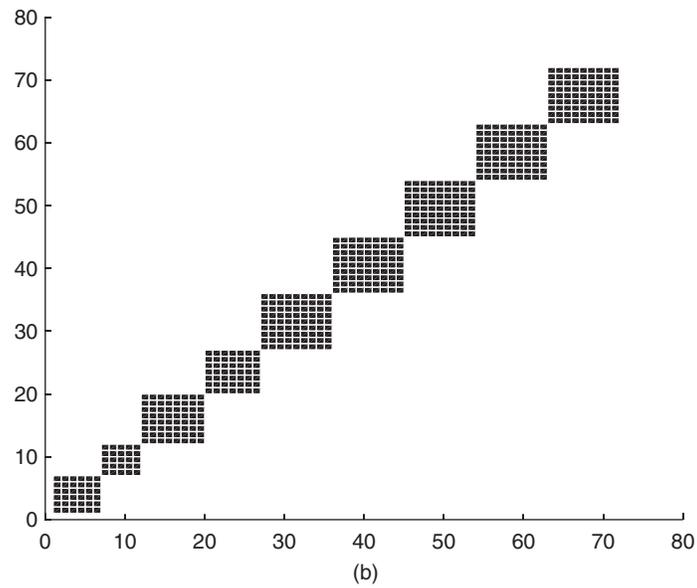
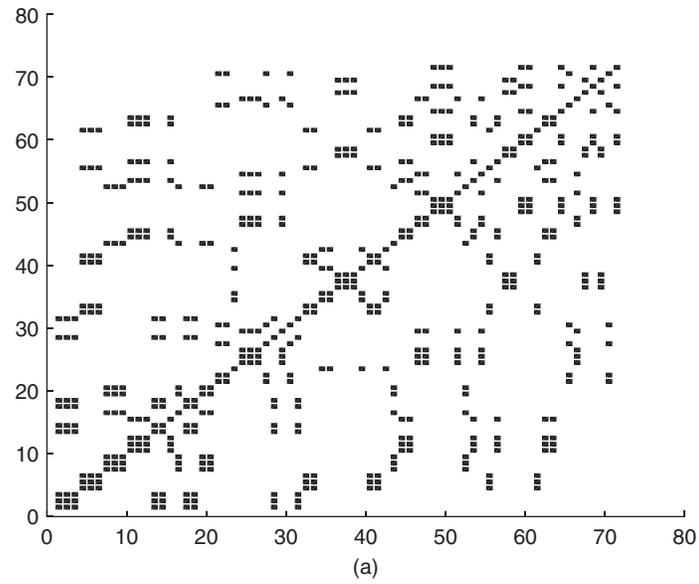
$$\alpha^{(1)} \mathbf{v}^{(1)} = \bar{\mathbf{Z}}^H \cdot \mathbf{x}^{(1)} \quad (1.174)$$

and constructs a recursion as

$$\beta^{(k+1)} \mathbf{x}^{(k+1)} = \bar{\mathbf{Z}} \cdot \mathbf{v}^{(k)} - \alpha^{(k)} \mathbf{x}^{(k)} \quad (1.175)$$

$$\alpha^{(k+1)} \mathbf{v}^{(k+1)} = \bar{\mathbf{Z}}^H \cdot \mathbf{x}^{(k+1)} - \beta^{(k+1)} \mathbf{v}^{(k)}, \quad (1.176)$$





**Figure 1.29** Distribution of the matrix elements in the self interactions of the boxes for a 72-unknown problem (a) before reordering and (b) after reordering based on the box indices.

With the direct factorization in preconditioners factorization of the sparse near-field matrix of all near-field interactions, one can obtain the near-field preconditioner (NFP). However, in most cases, near-field interactions are distributed irregularly in the full matrix and cannot be collected in a bandlimited diagonal region, even when using reordering algorithms [201]. Hence, the factorization performed once before iterations and the solution of the preconditioner system applied in each iteration can be expensive in terms of processing time; thus the preconditioner becomes a bottleneck as the problem size grows. As a remedy, a block-diagonal preconditioner (BDP), which is much more efficient, can be obtained by extracting only the self-interactions of the lowest-level boxes [35] (e.g., Figure 1.29). Both the factorization and application of BDP are  $\mathcal{O}(N)$ ; this is appropriate for MLFMA since its complexity is  $\mathcal{O}(N \log N)$  and the complexity of the preconditioner should not be higher than that. Due to its simplicity and favorable computing cost, BDP is commonly used for MLFMA solutions of normal and mixed formulations, such as T-N-CFIE. Unfortunately, BDP usually decelerates iterative solutions of tangential formulations, such as T-EFIE, except in a few special cases [202]. In addition, for large problems involving complicated objects, acceleration provided by BDP may not be sufficient, even when normal and mixed formulations are used.

Recently, there have been many efforts to improve iterative solutions of surface integral equations, particularly for metallic objects, using preconditioners [203]–[211]. In the context of MLFMA, one of the most common preconditioning techniques is the incomplete LU (ILU) method [200],[203]–[206]. This is a forward-type preconditioning technique, where the preconditioner matrix  $\bar{\mathbf{P}}$  approximates the system matrix, and one can solve for

$$\bar{\mathbf{P}}^{-1} \cdot \bar{\mathbf{Z}} \cdot \mathbf{a} = \bar{\mathbf{P}}^{-1} \cdot \mathbf{w} \quad (1.181)$$

or

$$(\bar{\mathbf{Z}} \cdot \bar{\mathbf{P}}^{-1}) \cdot (\bar{\mathbf{P}} \cdot \mathbf{a}) = \mathbf{w}, \quad (1.182)$$

instead of the original matrix equation. In (1.181) and (1.182), which correspond to left and right preconditioning, respectively, the solution of  $\bar{\mathbf{P}} \cdot \mathbf{x} = \mathbf{y}$  for a given  $\mathbf{y}$  should be cheaper than the solution of the original matrix equation. During the factorization of the preconditioner matrix, the ILU method sacrifices some of the fill-ins and provides an approximation to the near-field matrix, i.e.,

$$\bar{\mathbf{P}} = \bar{\mathbf{L}} \cdot \bar{\mathbf{U}} \approx \bar{\mathbf{Z}}_{NF}. \quad (1.183)$$

It is shown that the ILU preconditioner without a threshold provides inexpensive and good approximations to the near-field matrices for T-N-CFIE, hence it reduces the iteration counts and solution times substantially [206]. For ill-conditioned T-EFIE matrices, however, the ILUT (i.e., threshold-based ILU) preconditioner with pivoting [205] is required to prevent the potential instability. Other successful adoptions of the ILU-type preconditioners are presented in [204].

Despite the remarkable success of the ILU-type preconditioners, they are limited to sequential implementations due to difficulties in parallelizing their factorization algorithms and forward-backward solutions. Hence, the sparse-approximate-inverse (SAI) preconditioner, which is well-suited for parallel implementations, has been more preferable for the solution of large-scale electromagnetics problems [207]–[210]. The SAI preconditioner is based on a

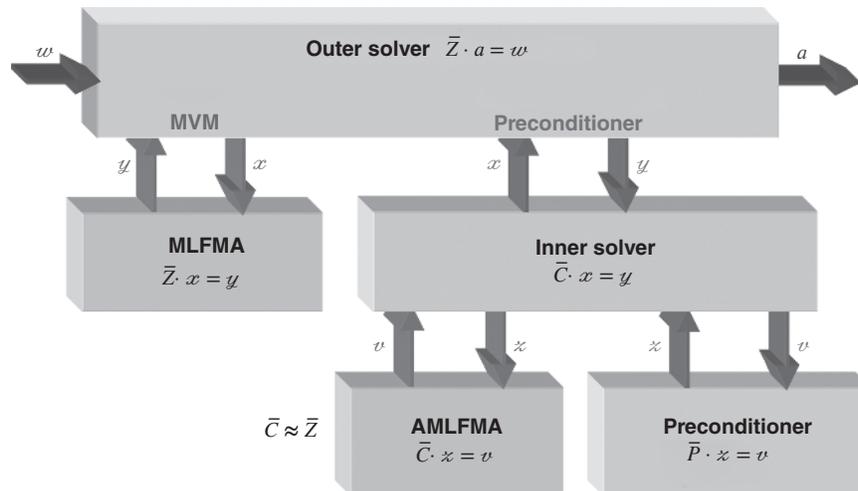
backward-type scheme, where the inverse of the system matrix is directly approximated, i.e.,  $\bar{P} \approx \bar{Z}^{-1}$ . In MLFMA, only the near-field matrix is considered, and one can minimize

$$\|\bar{I} - \bar{P} \cdot \bar{Z}_{NF}\|_F, \tag{1.184}$$

where  $\|\cdot\|_F$  represents the Frobenius norm. Using the pattern of the near-field matrix for the nonzero pattern of  $\bar{P}$  provides some advantages by decreasing the number of QR factorizations required during the minimization in (1.184) [208]. In parallel implementations, one can use row-wise partitioning to distribute the near-field interactions among processors. Therefore, left-preconditioning must be used to accelerate the iterative solutions with the SAI preconditioner. However, right-preconditioning can be used for the symmetric matrix equations, e.g., those derived from T-EFIE [210].

Preconditioners that are based on near-field interactions can be insufficient to accelerate iterative solutions of large-scale problems, especially those formulated with T-EFIE. For more efficient solutions, it is possible to use far-field interactions in addition to near-field interactions and construct more effective preconditioners. As depicted in Figure 1.30, this can be achieved by using flexible solvers and employing approximate and ordinary versions of MLFMA in an inner-outer scheme [211]. Using a reasonable approximation of the inner solutions, the number of outer iterations can be reduced substantially. There are two ways to describe the advantages of this strategy:

- MVMs performed by an ordinary MLFMA are replaced with more efficient MVMs performed by an approximate MLFMA (AMLFMA). Different from the relaxation strategies, however, only a single specific implementation of AMLFMA is sufficient to construct an inner-outer scheme. In addition, reasonable accuracy (without strict limits) is sufficient for the approximation.



**Figure 1.30** An inner-outer scheme involving a flexible iterative solver using an inner solver as a preconditioner.

- Iterative solutions by an ordinary MLFMA are preconditioned with a very strong preconditioner that is constructed by approximating the full matrix instead of the sparse part of the matrix.

Strategies for building a less-accurate MLFMA are discussed in Chapter 3, Section 3.5.

In addition to more efficient solutions, the inner-outer scheme prevents numerical errors that arise because of deviations in the computed residual from the true residual by significantly decreasing the number of outer iterations. This is because the residual gap, i.e., the difference between the true and computed residuals, increases with the number of iterations [212]. Another benefit of the reduction in iteration counts appears when the iterative solutions are performed with the GMRES algorithm, which is usually an optimal method for T-EFIE in terms of the processing time [206],[208]. Even though other variants of GMRES, e.g., the flexible GMRES (FGMRES) [205], require the storage of two vectors per iteration instead of one, nested solutions require significantly less memory than ordinary GMRES solutions because they dramatically reduce the iteration counts.

#### 1.14 Parallelization of MLFMA

For a dense matrix equation involving  $\mathcal{O}(N)$  unknowns, MLFMA reduces the complexity of MVMs from  $\mathcal{O}(N^2)$  to  $\mathcal{O}(N \log N)$ , allowing for the solution of large problems with limited computational resources. On the other hand, accurate solutions of many real-life problems require discretizations with millions of elements, leading to matrix equations involving millions of unknowns, which cannot easily be solved with sequential implementations of MLFMA running on a single processor. To solve such large problems, it is helpful to increase computational resources by assembling parallel computing platforms and, at the same time, by parallelizing MLFMA.

Of the various parallelization schemes for MLFMA, the most popular use distributed-memory architectures by constructing clusters of computers with local memories connected via fast networks [158],[214]–[232]. Parallelization tools, such as the message passing interface (MPI) [233], organizing parallel solutions via communication protocols are available. However, the parallelization of MLFMA is not trivial due to the complicated structure of this algorithm [217]. Although other algorithms with higher complexities, such as FMM, have been parallelized very successfully with simple parallelization techniques [216],[226], this is not the case for MLFMA. This is due to the unavoidable occurrences of communications among processors, poor load balancing, and duplications of computations over multiple processors in MLFMA. Consequently, there have been many efforts to improve the parallelization of MLFMA by using novel partitioning strategies, load-balancing algorithms, and optimized communications [217]–[232].

In a typical tree structure of MLFMA, the lowest level involves  $\mathcal{O}(N)$  boxes with dimensions in the order of wavelength, and the number of boxes decreases by a factor of four from one level to the next higher level. On the other hand, the number of samples required for the factorization and diagonalization of the Green's function depends on the size of boxes as measured by the wavelength. Consequently, the number of samples increases by a factor of four from one level to the next higher level. It is remarkable that the number of boxes and the number of samples balance each other, and all levels in MLFMA have approximately an equal cost with

$\mathcal{O}(N)$  complexity. Therefore, an efficient parallelization of MLFMA should consider the best possible partitioning at each level. With such a strategy, it would be possible to solve problems with hundreds of millions of unknowns on relatively inexpensive computing platforms with distributed-memory architectures [225]–[232].

In a simple parallelization of MLFMA, boxes are distributed among processors at all levels. Such a technique provides efficient parallelization for small numbers of processors, usually fewer than 16. For more processors, however, the efficiency drops drastically due to poor load balancing. This is because small numbers of boxes in the higher levels of MLFMA cannot be distributed evenly among processors. In addition, dense communications among processors are required. Note that low parallelization efficiency not only increases processing time but also restricts the size of problems that can be solved. As a remedy, a hybrid technique has been proposed to improve the parallelization of MLFMA [217],[221]. This technique is based on using an appropriate partitioning scheme for the higher levels of MLFMA by distributing samples instead of boxes. Due to the improved load balancing in the higher levels, the hybrid technique significantly increases the parallelization efficiency, compared to the simple technique. Nevertheless, the hybrid technique fails to provide efficient solutions when the number of processors is larger than 32 [219],[225].

Recently, a hierarchical parallelization technique has been developed for efficient parallelization of MLFMA [219],[223],[225],[230]–[232]. In this technique, the boxes and their samples are partitioned among processors by employing load-balancing algorithms. Although changing the partitioning between levels bears an additional cost, the hierarchical technique offers a higher parallelization efficiency than previous parallelization techniques for MLFMA. Specifically, this hierarchical partitioning strategy provides two important advantages. First, simultaneous partitioning of the boxes and their samples leads to improved load balancing among processors at each level. Second, communications between processors are reduced, i.e., average package size is enlarged, the number of communication events is reduced, and the communication time is significantly shortened. Details of the hierarchical partitioning strategy and comparisons with previous parallelization techniques can be found in Chapter 4, Section 4.6.

