

Numerical Modeling in Antenna Engineering

W. C. Chew^{a*}, L. J. Jiang^b, S. Sun^b, W. E. I. Sha^b, Q. I. Dai^a, M. Fallahpour^a and Y. M. Wu^c

^aDepartment of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, China

^bDepartment of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China

^cKey Laboratory for Information Science of Electromagnetic Waves (MoE), School of Information Science and Technology, Fudan University, Shanghai, China

Abstract

The principal computational electromagnetics techniques for solving antenna problems are reviewed. An introduction is given on a historical review of how antenna problems were solved in the past. The call for precise solutions calls for the use of numerical methods as found in computational electromagnetics. A brief introduction on differential equation solutions and integral solutions is given. The Green's function concept is introduced to facilitate the formulation of integral equations. Numerical methods and fast algorithms to solve these equations are discussed.

Then an overview of how electromagnetic theory relates to circuit theory is presented. Then the concept of partial element equivalence circuit is introduced to facilitate solutions to more complex problems. In antenna technology, one invariably has to have a good combined understanding of the wave theory and circuit theory.

Next, the discussion on the computation of electromagnetic solutions in the "twilight zone" where circuit theory meets wave theory was presented. Solutions valid for the wave physics regime often become unstable facing low-frequency catastrophe when the frequency is low.

Due to advances in nanofabrication technology, antennas can be made in the optical frequency regime. But their full understanding requires the full solutions of Maxwell's equations. Also, many models, such as perfect electric conductors, which are valid at microwave frequency, are not valid at optical frequency. Hence, many antenna concepts need rethinking in the optical regime.

Next, an emerging area of the use of eigenanalysis methods for antenna design is discussed. This can be the characteristic mode analysis or the natural mode analysis. These analysis methods offer new physical insight not possible by conventional numerical methods.

Then the discussion on the use of the domain decomposition method to solve highly complex and multi-scale antenna structures is given. Antennas, due to the need to interface with the circuit theory, often have structures ranging from a fraction of a wavelength to a tiny fraction of a wavelength. This poses a new computational challenge that can be overcome by the domain decomposition method.

Many antenna designs in the high-frequency regime or the ray optics regime are guided by ray physics and the adjoining mathematics. These mathematical techniques are often highly complex due to the rich physics that come with ray optics. The discussion on the use of these new mathematical techniques to reduce computational workload and offering new physical insight is given.

A conclusion section is given to summarize this chapter and allude to future directions.

*Email: w-chew@uiuc.edu

Keywords

Computational electromagnetics; Antenna modeling; Fast algorithms; Circuit theory; Wave theory; PEEC; Low-frequency breakdown; Calderon preconditioning; Nanoantennas; Design rules; Gain; Directivity; Characteristic mode analysis; Normal mode expansion; Domain decomposition; Equivalence principle algorithm; High-frequency asymptotics; Numerical steepest descent path

Introduction

Maxwell's equations, completed by James Clerk Maxwell in 1864, have been guiding electromagnetic theory for over 150 years now (Maxwell 1865). Solutions to Maxwell's equations have offered physical insight and guidance for myriads of electromagnetics technologies including antenna design. In the beginning, it was the use of analytic solutions. For instance, the solution of a radiating Hertzian dipole offers insight on the physics of its radiation. It provides physical insight on its radiation pattern, the near field and far field of the dipole, its radiation resistance, its directivity, and its gain. These are all figures of merit for an antenna designer.

Due to the complex shapes of antennas, closed-form solutions are quickly exhausted. Then approximate solutions are sought, and they have been used to gain physical insight into the antenna design. For instance, the half-wave dipole has no closed-form solution, but physical intuition tells that the current distribution is close to being sinusoidal. Hence, many figures of merit of a half-wave dipole can be calculated based on approximate models.

Another area of antenna design where approximation methods reign superior is in high-frequency antennas or reflector antennas. Due to the large radius of curvature of the antennas, quasi-optical techniques such as the physical optics approximation can be used to ascertain the radiation field of the antenna. Ray optics can be used to gain physical insight. Edge diffraction from the edge of the parabolic reflector can be estimated by using the canonical solution of the Sommerfeld half-plane (Sommerfeld 1896). Uniform asymptotic theory and uniform theory of diffraction are both used to refine the solutions (Keller 1962; Kouyoumjian and Pathak 1974; Lee and Deschamps 1976; Hansen 1981). The idea of approximate methods to gain physical insight prevails up to the late 1970s and early 1980s even for microstrip antenna designs (Lo et al. 1979; Chew and Kong 1981).

Even though modeling and simulation are important, physical insight is equally important. The most famous of this is the design of the Yagi-Uda antenna, proposed by Professor Yagi and his student, Uda, in 1926 (Yagi and Uda 2014). They had the insight that an array of dipoles can act like a waveguide. The simple dipole, made of a length of straight wire, is also called a linear antenna. When this linear antenna is slightly less than a half wavelength, it is capacitive, and an array of them can act like a waveguide. Hence, the dipole array, put in front of a half-wave dipole, can help guide the energy of the radiation field in the forward direction, enhancing its radiation in that direction and, hence, its directivity. When the linear antenna is slightly larger than a half wavelength, it becomes inductive and ceases to be a waveguide. Such a linear antenna is used as a reflector in the Yagi-Uda array.

When numerical methods became popular, integral equation methods were designed to solve linear antennas more accurately. Examples of such integral equation are the Hallén integral equation and Pocklington integral equation (Balanis 2012b). Approximate numerical method together with semi-analytic methods was used to solve these integral equations. Semi-analytic methods were popular because of the slowness of computers back then (King 1956).

Computers became powerful very quickly with the advancement of the electronic computers. Differential equation solvers as well as integral equation solvers were developed to solve Maxwell's equations.

In differential equation solvers, the unknowns are the field. Hence, the unknown counts are spread over volumetric space and, therefore, scale as the volume of the space of L^3 where L is the typical length of the simulation region. Moreover, in order to simulate an infinite region, absorbing boundary conditions are needed at the boundary of the simulation domain.

In integral equation solvers, the unknowns are the sources or the currents on the antenna. In many cases, these unknowns reside on the surface of the antenna structure, and thus, they live in a 2D manifold. In the cases of antennas that can be modeled by wire structures, the unknowns reside on a 1D manifold. Consequently, integral equation solvers became popular because the unknown count can be smaller. Very soon, integral equations were solved with the method of moments and applied to linear antennas. Linear antennas were popular because the unknowns were functions of a one-dimensional variable, and henceforth, they live in a one-dimensional space. Usually, many practical problems can be solved with a small number of unknowns and therefore needing the solution of small matrix systems.

As the antenna structure becomes more complex, more unknowns are needed. Early complex antenna structures were modeled with wire structures as in the numerical electromagnetic code (NEC). Very soon, surface patch models were developed and popularized (Burke and Poggio 1981; Rao et al. 1982).

A drawback of the integral equation solver is the memory usage. The matrix system is often dense and hence consumes $O(N^2)$ memory usage. For antennas that can be modeled as surface patches, $N \sim O((kL)^2)$ where L is the typical dimension of the antenna. Henceforth, memory usage scales as $(kL)^4$ for them. When used to solve linear antennas, $N \sim O(kL)$ and therefore the memory usage scales as $(kL)^2$.

On the other hand, differential equation solvers are easier to implement even though unknown counts are larger. Moreover, the innate matrix system that is related to a differential equation is inherently sparse. Therefore, when combined with iterative solvers, the matrix system can be efficiently stored with $O(N)$ memory where N is the dimension of the matrix and, hence, the number of unknowns. Furthermore, each matrix-vector product can be effected in $O(N)$ operations. Consequently, the iterative solver can be effected with $O(N)$ operations per iteration. Typically, the memory usage and CPU time usage scale as $(kL)^3$ when grid dispersion error is ignored. Hence, it could be more efficient than integral equation solvers.

A powerful differential equation solver is the finite-difference time-domain method invented by Yee (1966) and popularized by Taflov (1995). In this method, the construction of the matrix system is not needed, and a matrix-vector product is effected by some rather simple operations. Since only the unknown fields at different time steps need to be stored, the memory and CPU time usage are both $O(N)$. Even though FDTD is solved with time stepping, it is equivalent to iteration in iterative solvers. Iterative solvers are generally applied to frequency-domain solutions.

Due to advancements of differential equation solvers, their efficiency, and their simplicity, they have become rather popular. They can be used to solve problems with tens of thousands of unknowns easily and, hence, many practical real-world problems.

In recent years, however, fast integral equation solvers have been developed (Rokhlin 1990; Coifman et al. 1993; Chew et al. 2001). In these solvers, iterative methods are used to solve integral equations without the need to construct the matrix system. Thus, the matrix is never stored and only the action of the matrix on a vector needs to be stored. Henceforth, memory requirements of the fast integral equation solvers are reduced to $O(N \log N)$ for dynamic problems while to $O(N)$ to static or low-frequency problems. Moreover, the matrix-vector product can be effected in $O(N \log N)$ operations. So, the memory usage for an antenna that can be modeled by surface currents scales as $(kL)^2$ and the operations per iteration scale as $(kL)^2 \log kL$, greatly reducing the memory and CPU usage compared to differential equation solvers.

Differential Equations-Maxwell's Equations

Maxwell's equations establish that the theory of electricity and magnetism and theory of optics are actually one and the same theory. Both of them have waves propagating at the velocity of light. The velocity of light, and astronomical number, was established by Roemer via astronomical observation long ago (in year 1676).

Maxwell's equations in the time domain are completely described by the four equations:

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (1)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (2)$$

$$\nabla \cdot \mathbf{D} = \rho \quad (3)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (4)$$

When Maxwell's equations were first derived by Maxwell, they were not as elegant and succinct as the above. It was the work of Oliver Heaviside who casts them into the above form (Heaviside 1888).

For time-varying problems, the third and the fourth equations are derivable from the first two by taking their divergence. Hence, only the first two equations can be considered as independent, and only them need to be solved for time-varying problems.

Since there are four unknowns \mathbf{E} , \mathbf{H} , \mathbf{D} , and \mathbf{B} with two independent equations, the constitutive relations are the addition equations in order to obtain four equations and four unknowns, viz., for isotropic media,

$$\mathbf{D} = \epsilon \mathbf{E} \quad (5)$$

$$\mathbf{B} = \mu \mathbf{H} \quad (6)$$

The above constitutive relations are sufficient to describe most materials at microwave frequencies. For some materials such as ferrites, the above will have to be replaced with the constitutive relations for anisotropic media.

The Yee algorithm will solve the first two of the above first-order partial differential equations directly. But many algorithms will first convert the above into one second-order partial differential equation. By assuming $\exp(-i\omega t)$ dependence, $\partial/\partial t$ can be replaced by $-i\omega$ in the above equations. Then by dividing the first equation by ϵ , taking its curl, and then replacing the resultant right-hand side with the second Maxwell's equations, one arrives at

$$\nabla \times \epsilon^{-1} \nabla \times \mathbf{H} - \omega^2 \mu \mathbf{H} = \nabla \times \epsilon^{-1} \mathbf{J} \quad (7)$$

Similarly, one has

$$\nabla \times \mu^{-1} \nabla \times \mathbf{E} - \omega^2 \epsilon \mathbf{E} = i\omega \mathbf{J} \quad (8)$$

One needs only solve one of the two equations above since they are derivable from each other. When the above equation is solved by a numerical method, the field \mathbf{H} or \mathbf{E} is treated as the unknown, and the source is assumed known.

Scalar and Dyadic Green's Function

The Green's function is useful because it expresses the field in terms of the source of a system. It is the point source response of the partial differential equation or the field generated due to a point source. When a medium is homogeneous, such a response or Green's function can be found easily. It is well known that when the Helmholtz equation is driven by a point source, the response is the scalar Green's function, namely,

$$\nabla^2 g(\mathbf{r}, \mathbf{r}') + k^2 g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}') \quad (9)$$

and $g(\mathbf{r}, \mathbf{r}') = \exp(ik|\mathbf{r} - \mathbf{r}'|)/(4\pi|\mathbf{r} - \mathbf{r}'|)$ and $k = \omega\sqrt{\mu\epsilon} = \omega/c$. Hence, when the medium is homogeneous, the above vector wave equation can be written as

$$\nabla\nabla \cdot \mathbf{E} - \nabla \cdot \nabla \mathbf{E} - k^2 \mathbf{E} = i\omega\mu\mathbf{J} \quad (10)$$

where the identity has been used that $\nabla \times \nabla \times \mathbf{E} = \nabla\nabla \cdot \mathbf{E} - \nabla \cdot \nabla \mathbf{E}$. Furthermore, by noting that $\nabla \cdot \mathbf{E} = \rho/\epsilon$, the above becomes

$$\nabla^2 \mathbf{E} + k^2 \mathbf{E} = -i\omega\mu\mathbf{J} + \nabla\rho/\epsilon \quad (11)$$

By using the principle of linear superposition and the solution to Eq. 9, the solution to the above can be written as

$$\mathbf{E}(\mathbf{r}) = i\omega\mu \int_V d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \mathbf{J}(\mathbf{r}') - \int_V d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \nabla' \rho(\mathbf{r}')/\epsilon \quad (12)$$

$$= i\omega\mu \int_V d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \mathbf{J}(\mathbf{r}') - \frac{1}{i\omega\epsilon} \nabla \int_V d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') \nabla' \cdot \mathbf{J}(\mathbf{r}') \quad (13)$$

The above is reminiscent of that

$$\mathbf{E} = i\omega\mathbf{A} - \nabla\phi \quad (14)$$

One can also symbolically write that

$$\mathbf{E}(\mathbf{r}) = i\omega\mu \int_V d\mathbf{r}' \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}') \quad (15)$$

where the action of the dyadic Green's function on the current \mathbf{J} plus the integral is to effect the integral in Eq. 12 above.

Integral Equation of Excitation

An antenna is usually made of metal and excited by a source. The source generates an incident field on the antenna. At microwave frequencies, metallic structures can be approximated by a perfect electric conductor (PEC) where the tangential component of the electric field is zero. Otherwise, there would be infinite surface current flowing on the surface of the PEC structure.

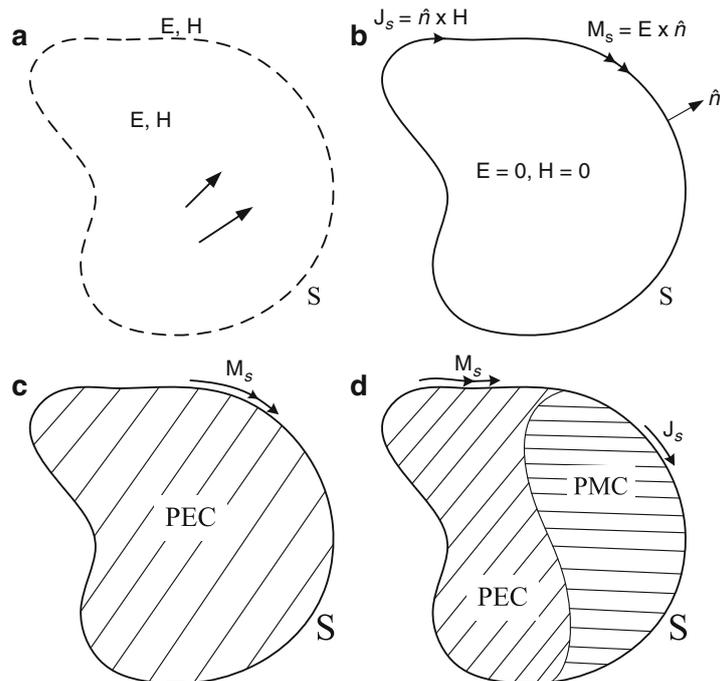


Fig. 1 By the equivalence principle, the four cases (a–c) and (d) generate equivalent fields outside the surface S

As a result, a surface current is induced on the antenna so as to produce a field that cancels the incident field on the surface of the antenna structure. This physical concept can be represented by the following integral equation:

$$-\hat{t} \cdot \mathbf{E}_{inc}(\mathbf{r}) = i\omega\mu\hat{t} \cdot \int_S dS' \overline{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}(\mathbf{r}'), \quad \mathbf{r} \in S \quad (16)$$

Source Excitation Methods

Reflector antennas are usually driven by a feed which is usually a horn antenna. The modeling of the feed can be done by the equivalence principle. For smaller antennas at lower frequencies, oftentimes, an antenna is driven by a current source or voltage source at its terminal. The equivalence principle can be used to obtain the equivalence problem that needs to be solved before approximations are made.

Equivalence Principle

The equivalence principle can be derived mathematically by invoking Green's theorem for scalar waves or vector Green's theorem for electromagnetic waves. But it can be argued physically or by the Gedankenexperiment. Imaging a set of sources generates \mathbf{E} and \mathbf{H} fields in free space. One defines a surface S on which one will impress the surface electric current $\mathbf{J}_s = \hat{n} \times \mathbf{H}$ and surface magnetic current $\mathbf{M}_s = \mathbf{E} \times \hat{n}$. One assumes that the fields $\mathbf{E} = \mathbf{H} = 0$ inside S and that the original fields are unchanged outside S . These fields definitely satisfy the jump conditions induced by \mathbf{J}_s and \mathbf{M}_s at the interface. Furthermore, the fields outside satisfy the boundary conditions (including the radiation condition at infinity) on the surface S , and hence, by the uniqueness theorem, they are the only solution (Figs. 1, 2, 3, 4, and 5).

Since the fields are zero inside S , one can insert a PEC inside S without affecting the fields outside. When the PEC fills the whole of the volume inside S , then the surface current \mathbf{J}_s cannot radiate, as it will induce an image current that is negative of itself. Consequently, the magnetic current \mathbf{M}_s alone, when

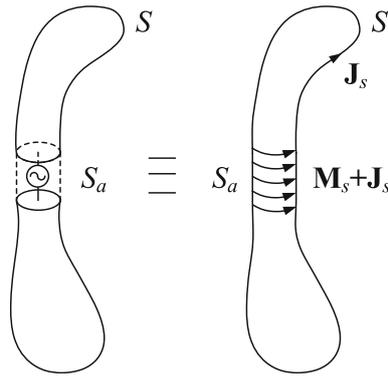


Fig. 2 For a voltage source driving an antenna at the gap, the equivalence principle can be invoked to make the case on the right to be equivalent to the case on the left. When the gap is filled with PEC, the electric current can be removed, with only the magnetic current impressed on it. When the gap is filled with PMC, the magnetic current can be removed with the electric current impressed on the PMC surface

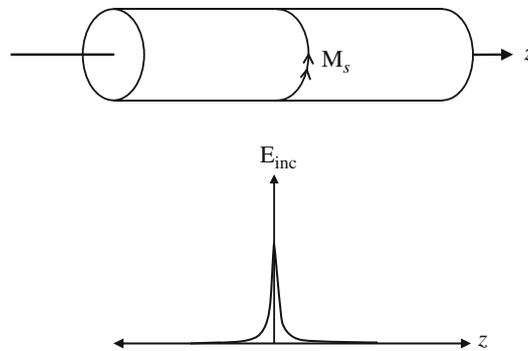


Fig. 3 An antenna driven by a very thin magnetic current ribbon. The incident field generated by the magnetic current ribbon can be approximated by a delta function

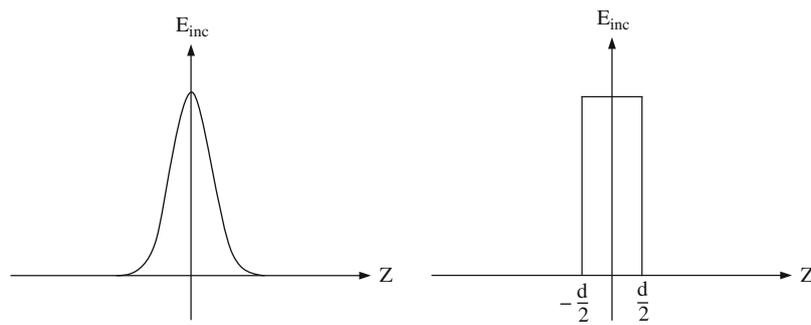


Fig. 4 The plot of the incident field generated by a finite-width magnetic ribbon. The incident field generated by the finite-width magnetic current ribbon can be approximated by a pulse function

impressed on the PEC surface, will radiate the same fields outside. A similar argument can be made when the volume is filled with PMC or a mixture thereof.

With the above equivalence principle established, one can use that to derive an equivalence model for an antenna driven by a source at its gap. First, by wrapping a surface S on the antenna snugly, one can represent the fields outside the antenna with impressed currents \mathbf{J}_s and \mathbf{M}_s per the above discussion. On the PEC surface of the antenna, $\mathbf{M}_s = 0$ since $\hat{n} \times \mathbf{E} = 0$. The only nonzero surface currents are \mathbf{J}_s and \mathbf{M}_s

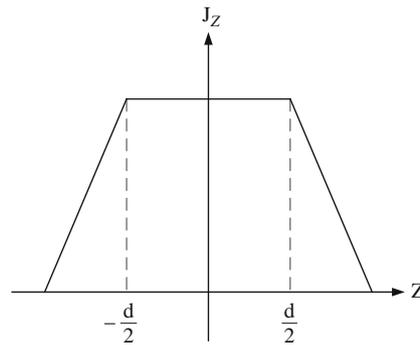


Fig. 5 The antenna can also be excited by a current source. The current source can be represented by one basis function with known and fixed amplitude

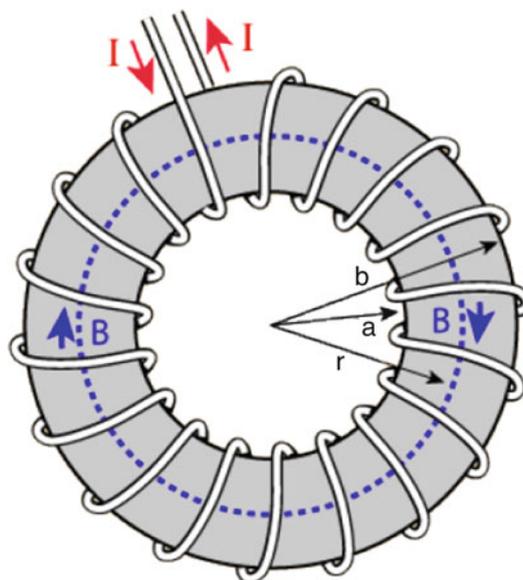


Fig. 6 A magnetic ribbon source can be approximated by a small toroid. The figure is from hyperphysics.phy-astr.gsu.edu/hbase/magnetic/toroid.html

at the gap region. Now, one can fill the gap with PEC, and again, \mathbf{J}_s does not radiate. Hence, only \mathbf{M}_s impressed on the surface associated with the gap region is needed to produce the equivalent fields outside the antenna.

The impressed magnetic current can be modeled by a magnetic ribbon current. This magnetic current can be used to generate the incident field on the antenna, which will in turn induce a current on it. What is to be noted is that the magnitude of the magnetic current is related to the voltage across the gap. When the magnetic current ribbon is very thin, it produces a field that is sharply peaked at the location just beneath the ribbon. This sharply peaked function can be approximated by a delta function. Therefore, one can approximate the incident field just as a delta function. This is the well-known delta-gap model.

When the magnetic ribbon has a finite width, the incident field it generates will have a Gaussian-like pulse shape. This can be used as the incident field on the antenna. This Gaussian pulse will induce a voltage across the gap. But for simplicity, one can approximate the Gaussian pulse incident field with a square pulse with the same area. It will induce the same voltage drop across the gap (Lo et al. 2013).

To realize a magnetic ribbon source in the real world, one can use a small toroid as shown in Fig. 6. As is well known, a small loop antenna is like a magnetic dipole, and the toroid can be thought of as a string of

magnetic dipoles forming a circle emulating a magnetic current loop. The current induced on the antenna wire is a transformer action, and the antenna wire is the secondary winding of the transformer. Such toroidal antennas have been used in well-logging exploration tools (Clark 1985).

Alternatively, one can fill the gap region with a perfect magnetic conductor (PMC) and impress only the electric current \mathbf{J}_s as the equivalent current on the gap surface. One can solve this excitation problem to obtain $\hat{n} \times \mathbf{E}$ on the surface of the PMC gap and hence the voltage. With this voltage, one can calculate the input impedance of the antenna.

But a gap filled with a PMC material is an idealization of an open circuit where no current can flow. As an approximation, one can replace the gap with one filled with free space again and drive the antenna with a current source at the gap. The resultant voltage at the gap can then be calculated to obtain the input impedance. This is the current source driven antenna excitation model.

Subspace Projection Method

Integral equations and differential equations previously derived represent linear operator equations. One can describe a general linear problem with a linear operator equation as

$$\mathcal{L}f = h \quad (17)$$

where \mathcal{L} represents a linear operator. This linear operator represents a differential operator in the case of differential equations, while it represents an integral operator for integral equations. In mathematical notation, one often writes $\mathcal{L} : V \rightarrow W$, meaning that it is a linear operator that maps elements of the vector space V to the elements of the vector space W , where V is known as the domain space and W the range space of the operator \mathcal{L} .

One illustrates the subspace projection method with scalar equations as the mathematical concept does not differ greatly when vector equations are used. For the differential operator, an explicit form may be

$$(\nabla^2 + k^2)f(\mathbf{r}) = h(\mathbf{r}) \quad (18)$$

which is the Helmholtz wave equation or the frequency-domain version of the wave equation. A boundary condition together with the radiation condition or a loss condition has to be stipulated in order to make the solution to the above equation unique (Chew 1990).

For the integral equation, one can consider a scalar integral equation and an explicit form is given:

$$\int d\mathbf{r}' g(\mathbf{r}, \mathbf{r}') f(\mathbf{r}') = h(\mathbf{r}) \quad (19)$$

where $g(\mathbf{r}, \mathbf{r}') = \exp(ik|\mathbf{r} - \mathbf{r}'|)/(4\pi|\mathbf{r} - \mathbf{r}'|)$ is the Green's function which is also called the kernel.

These equations are not amenable to computation. They are equations in the infinite-dimensional Hilbert space with uncountably infinite indices. To make them computable, one can find a subspace in which an approximate solution of the above equations can be found and then project the solution into this subspace. This process can be methodically described as follows.

One chooses a basis set with N basis functions f_n , $n = 1, \dots, N$ that spans the subspace that can approximate the domain space of the operator. One expands the unknown f in terms of the basis functions in this subspace, namely,

$$f(\mathbf{r}) \doteq \sum_{n=1}^N a_n f_n(\mathbf{r}) \quad (20)$$

The linear operator equation can then be written as

$$\sum_{n=1}^N a_n \mathcal{L}f_n \doteq h \quad (21)$$

The above equation can only be approximately satisfied since one has picked a finite number of basis functions. Moreover, the solution to Eq. 21 is still untenable. To make it more easily solvable, one converts Eq. 21 into a matrix equation by weighting or testing the equation with $w_m(\mathbf{r})$, $m = 1, \dots, N$. Consequently, one has

$$\sum_{n=1}^N a_n \langle w_m, \mathcal{L}f_n \rangle = \langle w_m, h \rangle, \quad m = 1, \dots, N \quad (22)$$

where the inner product is defined as

$$\langle u, v \rangle = \int u(\mathbf{r})v(\mathbf{r})d\mathbf{r} \quad (23)$$

The integral is over the support of the functions u and v . When they are defined over a volume (surface), one has a volume (surface) integral.

The above Eq. 22 implies that it can be exactly satisfied if one projects the Eq. 21 onto a subspace spanned by $w_m(\mathbf{r})$, $m = 1, \dots, N$.

Recent research has shown that if the testing functions w_m are chosen from the dual space of the range space, the above can be a good approximation or matrix representation of the original operator equation (Cools et al. 2009). The dual space can be larger or smaller than the range space.

Equation 22 now constitutes a matrix equation:

$$\bar{\mathbf{L}} \cdot \mathbf{a} = \mathbf{h} \quad (24)$$

$$(\bar{\mathbf{L}})_{mn} = \langle w_m, \mathcal{L}f_n \rangle \quad (25)$$

$$(\mathbf{a})_n = a_n \quad (26)$$

$$(\mathbf{h})_m = \langle w_m, h \rangle \quad (27)$$

$$(\mathbf{a}) = (\bar{\mathbf{L}})^{-1} \cdot \mathbf{h} \quad (28)$$

The matrix equation can be solved by various means. In the above, the choice of the testing function is determined by the auxiliary equation (Chew 1990). The testing function should be chosen so that the auxiliary equation is solved accurately, whereas the choice of the expansion or basis function is to solve the primary equation accurately. The auxiliary equation is the adjoint or the transpose of the original equation where the range space and the domain space are swapped. It can be shown that the left domain space of an operator is the dual space of the range space. Hence, if the testing functions solve the auxiliary equation well, they approximate the left domain space of the original operator, which is the dual space of the range space of the original equation, well. This is in agreement with the aforementioned recent findings.

The above procedure of converting the operator equation into a matrix equation is the underpinning method behind the finite-element method or the method of moments. They are variously known as Galerkin's method, Petrov–Galerkin's method, method of weighted residuals, collocation method, and point matching method (Chew 1990, Chap. 5). But they all can be regarded as subspace projection methods.

When \mathcal{L} is a differential operator, the matrix $\bar{\mathbf{L}}$ is sparse because a differential operator is a local operator. However, when \mathcal{L} is an integral operator, the matrix $\bar{\mathbf{L}}$ is dense. In differential equations, the unknown is the field that permeates all of space. Hence, the unknown count is usually large in differential equation solvers. But they come with sparse matrix systems that are cheaper to solve and store.

For integral equations, the unknown is the induced current on the antenna structure. Hence, the unknown count is smaller since the current resides only on or in a structure with finite support. But the ensuing matrix system is dense and is hard to store. Moreover, the matrix assembly of forming the matrix elements

$$\langle w_m, \mathcal{L}f_n \rangle \quad (29)$$

is tedious since the operator involves singularity that has to be evaluated with care (Graglia 1993).

Iterative Methods

When the unknown count is large, the matrix $\bar{\mathbf{L}}$ is never generated. Instead, iterative, matrix-free methods in numerical linear algebra are used to solve the matrix equation instead. Iterative methods can be made matrix-free because one needs only to write a code to produce the result

$$\bar{\mathbf{L}} \cdot \mathbf{a} \quad (30)$$

namely, the result of the action of a matrix on a vector, but not the matrix itself. For sparse matrices, this action can be effected with $O(N)$ operations. For dense matrices, it costs $O(N^2)$ to effect this action. However, various fast solvers have been developed that allow this action to be effected in $O(N)$ or $O(N \log N)$ operations (Chew et al. 2001, ref. therein).

When it comes to solving a matrix equation iteratively, numerical linear algebra plays an important role (Trefethen et al. 1997; Chew et al. 2009). The condition number and the distribution of the eigenvalues of the matrix system determine the convergence rate of iterative methods. The Krylov subspace method is a popular way to understand the convergence rate of iterative methods. In this method, one finds the solution to the matrix equation by finding the best-fit solution in a subspace called the Krylov subspace

$$\mathcal{K}^K(\bar{\mathbf{L}}, \mathbf{r}_0) = \left\{ \mathbf{r}_0, \bar{\mathbf{A}} \cdot \mathbf{r}_0, \bar{\mathbf{A}}^2 \cdot \mathbf{r}_0, \dots, \bar{\mathbf{A}}^{K-1} \cdot \mathbf{r}_0 \right\} \quad (31)$$

where $\mathbf{r}_0 = \mathbf{h} - \bar{\mathbf{L}} \cdot \mathbf{a}_0$ where \mathbf{a}_0 is the initial guess to the solution. The above Krylov subspace can be generated by performing $K - 1$ matrix-vector product with the matrix $\bar{\mathbf{L}}$ on the vector \mathbf{r}_0 . The method finds the optimal solution at the K -th iteration, \mathbf{a}_K , by letting it be $\mathbf{a}_K = \mathbf{a}_0 + \mathbf{z}_K$, such that $\mathbf{z}_K \in \mathcal{K}^K(\bar{\mathbf{L}}, \mathbf{r}_0)$. By so doing, the residual at the K -th iteration is

$$\begin{aligned} \mathbf{r}_K &= \mathbf{h} - \bar{\mathbf{L}} \cdot \mathbf{a}_K \\ &= \mathbf{r}_0 - \bar{\mathbf{L}} \cdot \mathbf{z}_K \in \left\{ \mathbf{r}_0, \bar{\mathbf{L}} \cdot \mathbf{r}_0, \bar{\mathbf{L}}^2 \cdot \mathbf{r}_0, \dots, \bar{\mathbf{L}}^K \cdot \mathbf{r}_0 \right\} \\ &\in \mathcal{K}^{K+1}(\bar{\mathbf{L}}, \mathbf{r}_0) \end{aligned} \quad (32)$$

In other words, the residual error that is to be minimized can be written as

$$\begin{aligned} \mathbf{r}_K &= \mathbf{r}_0 + \alpha_1 \bar{\mathbf{L}} \cdot \mathbf{r}_0 + \alpha_2 \bar{\mathbf{L}}^2 \cdot \mathbf{r}_0 + \cdots + \alpha_K \bar{\mathbf{L}}^K \cdot \mathbf{r}_0 \\ &= \sum_{k=0}^K \alpha_k \bar{\mathbf{L}}^k \cdot \mathbf{r}_0 = P_K^o(\bar{\mathbf{L}}) \cdot \mathbf{r}_0 \end{aligned} \quad (33)$$

where

$$P_K^o(x) = 1 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_K x^K$$

is an optimal polynomial with $\alpha_0 = 1$ and the coefficients $\alpha_1, \dots, \alpha_K$ are chosen to minimize \mathbf{r}_K .

To understand the convergence of the above matrix polynomial, one expands \mathbf{r}_0 in terms of the left and right eigenvectors of $\bar{\mathbf{L}}$ or (Chew et al. 2009, ref. therein)

$$\mathbf{r}_0 = \sum_{n=1}^N \mathbf{v}_n (\mathbf{w}_n^t \cdot \mathbf{r}_0) \quad (34)$$

where \mathbf{v}_n is the right eigenvector, while \mathbf{w}_n is the left eigenvector. They can be shown to share the same set of eigenvalues and are mutually orthogonal. That is, $\mathbf{w}_n \cdot \mathbf{v}_{n'} = \delta_{nn'}$. Therefore,

$$\bar{\mathbf{L}} \cdot \mathbf{v}_n = \lambda_n \mathbf{v}_n \quad (35)$$

$$\mathbf{w}_n^t \cdot \bar{\mathbf{L}} = \lambda_n \mathbf{w}_n^t \quad (36)$$

where the eigenvalue λ_n can be complex. One can easily show that

$$\bar{\mathbf{L}}^k \cdot \mathbf{r}_0 = \sum_{n=1}^N \lambda_n^k \mathbf{v}_n (\mathbf{w}_n^t \cdot \mathbf{r}_0) \quad (37)$$

By substituting the above into Eq. 33, one has

$$\mathbf{r}_K = \sum_{n=1}^N \sum_{k=0}^K \alpha_k \lambda_n^k \mathbf{v}_n (\mathbf{w}_n^t \cdot \mathbf{r}_0) = \sum_{n=1}^N P_K^o(\lambda_n) \mathbf{v}_n (\mathbf{w}_n^t \cdot \mathbf{r}_0) \quad (38)$$

$$\mathbf{r}_K = \sum_{k=0}^K \bar{\mathbf{L}}^k \cdot \mathbf{r}_0 = \sum_{n=1}^N P_K^o(\lambda_n) \mathbf{v}_n (\mathbf{w}_n^t \cdot \mathbf{r}_0) \quad (39)$$

$$\mathbf{r}_0 = \mathbf{g} - \bar{\mathbf{L}} \cdot \mathbf{a}_0 \quad (40)$$

$$\mathbf{r}_1 = \mathbf{r}_0 + \alpha'_1 \bar{\mathbf{L}} \cdot \mathbf{r}_0 \quad (41)$$

It is seen that the residual error in the K -th iteration is proportional to the value of the optimal polynomial $P_K^o(\lambda_n)$ at the N eigenvalues λ_n . If $K = N$, a polynomial can be found such that $P_K^o(\lambda_n)$ is exactly zero at all these eigenvalues, meaning that the residual error is zero. When K is less than N , if the

eigenvalues are clustered together on the complex plane, the residual error can still be made small. Since the polynomial has value such that $P_K^o(0) = 1$, if there are many eigenvalues near to the origin, it is difficult to fit the polynomial so that it is small close to the origin. Hence, having eigenvalues close to the origin is bad for convergence, since it is hard to find the fitting polynomial. By the same token, if the eigenvalues are spread widely over the complex plane, it is hard to find a fitting polynomial as well. All these cases give rise to a bad condition number of the matrix system which is defined as $|\lambda_{\max}|/|\lambda_{\min}|$, the ratio of the largest eigenvalue to the smallest eigenvalue. Therefore, an ill-conditioned matrix has bad convergence when solved with iterative solvers.

Fast Algorithm for Integral Equations

Due to the increased workload of numerical methods when the antenna structure becomes electrically large (large compared to wavelength), there has been much interest in fast methods to solve the ensuing matrix equations derived from Maxwell's equations.

Differential equation solvers naturally give rise to sparse matrices. The solution process can be made matrix-free easily. The downside is the existence of the grid dispersion error whose deleterious effect increases with the size of the problem (Lee and Cangellaris 1992). In this case, the effort has been in reducing the unknown counts as the unknowns are fields that permeate all of space pervasively. Also, accuracy improvement is necessary to mitigate grid dispersion errors. A higher-order method is used to reduce unknown counts but with the peril of reducing the sparsity of the matrix. Also, the unknown count in differential equation solvers grows with the volume of the simulation region and, hence, suffers from the cruelty of dimensionality.

Methods to invert the finite-element matrix directly have been studied extensively. Because of the sparsity of the FEM matrix for differential equations, it can be directly inverted in $N^{1.5}$ in 2D and N^2 in 3D by the nested dissection ordering method (Axelsson and Barker 1984). Moreover, for domains where the shape is oblong, the frontal method proves popular in inverting the finite-element matrix.

Another hot area of research in differential equation solvers is the design of absorbing boundary conditions (ABCs) (Chew 1990, Chap. 4). Berenger's perfectly matched layers (PMLs) (Berenger 1994) and coordinate-stretching PML (Chew and Weedon 1994) have become highly popular among numericists in this area. Coordinate-stretching PML draws inspirations from the area of matched asymptotics where coordinate stretching is used to emphasize a certain physics of the problem. Furthermore, an anisotropic-medium PML has been developed (Sacks et al. 1995). This topic has spurred the interest of many researchers.

Most ABCs are not perfect; a rigorous, perfect ABC is actually a boundary integral equation truncation of a differential equation solution domain solved by FEM or FDM (Jin 2002; Volakis et al. 1998; Zhu and Cangellaris 2006). This has been avoided in the past because boundary integral equations give rise to dense matrix systems that are expensive to solve and store. But the advent of fast algorithms has changed the landscape (Chew et al. 2001; Peng et al. 2011). Boundary integral equations accelerated by fast algorithms have been used to reduce the domain size of finite-element methods since they act as rigorous absorbing boundary conditions.

Cruelty of Computational Complexity

Integral equations were difficult and expensive to solve in the past. They usually give rise to a dense matrix system requiring $O(N^3)$ computer time to solve and $O(N^2)$ memory requirements. These computational complexities are just too unwieldy for large problems. But advances in fast methods have eliminated this bottleneck by reducing the size of α in the exponent in N^α in these computational complexity scalings.

For iterative solvers, a sleuth of method can reduce the computational time complexity to $O(N \log N)$. Moreover, many of these methods can be made matrix-free so that only the unknown vector and the diagonal part of the matrix corresponding to near-neighbor interactions need to be stored. Hence, the memory requirements can be reduced to $O(N)$. There are essentially three popular methods to speed up the solutions of integral equations:

- **Fast Fourier transform-based techniques:** In this technique, the integral operator is cast into a form that resembles a convolutional integral as much as possible. Then FFT is used to expedite the convolution, enabling its evaluation of matrix-vector products in $O(N \log N)$ operations. For surface structures or sparsely packed structures where a lot of zero padding in the FFT is needed, the complexity is worse and is not optimal (Borup and Gandhi 1984).
- **Matrix compression-based techniques:** The MOM matrix that follows from solving integral equations is low rank. The reasons for low rank are two: over-discretization of mesh density and far interactions between currents on the object. The Nyquist sampling theorem necessitates the discretization of at least two points per wavelength in order to capture the oscillatory nature of the currents on an object. But oftentimes, discretization far above the Nyquist sampling rate is used. In this case, redundancies are created in the unknown counts, and the rank of the ensuing matrix system is lower rank than the unknown counts. Such matrix systems can be easily compressed using matrix compression techniques, such as wavelets (Wagner et al. 1993, ref. therein), adaptive cross approximations, simple fast multipole, etc. (Zhao et al. 2004, ref. therein). Interaction matrices with low ranks due to far interactions are harder to compress. They cannot be compressed beyond the Nyquist barrier (Wagner et al. 1993), even though ray physics-based methods have been used to compress them further beyond the Nyquist barrier. When wave physics is involved, the only viable way to compress the matrices for far interaction efficiently is the multipole-based methods (Chew et al. 2001).
- **Multipole-based methods:** The simple fast multipole method can expedite the solution involving circuit physics or Laplace's solutions very easily. Extension of such an algorithm for wave physics cases has been proposed, but its verbatim use for wave physics cases does not work. The only viable method of expediting the wave physics case is the multilevel fast multipole algorithm (Chew et al. 2001), where antepolation and interpolation between levels are added. The algorithm is a tree-based algorithm. The matrices for far interactions are analytically diagonalized on paper rather than by algebraic or numerical means. The number of diagonal elements needed is proportional to the rank of the matrices. The reason is that the ranks of the matrices increase with the group size compared to wavelength. Near the upper levels of the inverted tree, the ranks of the matrices are higher, while near the lower levels, their ranks are lower.

To be of the varying ranks of these matrices, their dimensions are different for different levels even if they are diagonalized. Hence, antepolation and interpolation are needed between levels due to their different ranks, very much in the spirit of the multilevel multigrid schemes. With this augmentation, the multilevel fast multipole algorithm provides optimal complexity of performing a matrix-vector product with $O(N \log N)$ complexity. It is with this algorithm that wave physics problems with tens of millions to hundreds of millions and over three billion unknowns have been solved (Michiels et al. 2014).

The key to the multilevel fast multipole algorithm is the factorization of the matrix element L_{ij} when elements i and j are far apart. Then one can express (Chew et al. 2001).

$$L_{ij} = \tilde{\mathbf{V}}_{f,i,i_1}^t \cdot \tilde{\mathbf{I}}_1 \cdot \tilde{\beta}_{i_1,i_2} \cdot \tilde{\mathbf{I}}_2 \cdots \tilde{\beta}_{i_N,L} \cdot \tilde{\alpha}_{LL'} \cdot \tilde{\beta}_{L',j_N} \cdots \tilde{\mathbf{I}}_2 \cdot \tilde{\beta}_{j_2,j_1} \cdot \tilde{\mathbf{I}}_1 \cdot \tilde{\mathbf{V}}_{s,j_1,j} \quad (42)$$

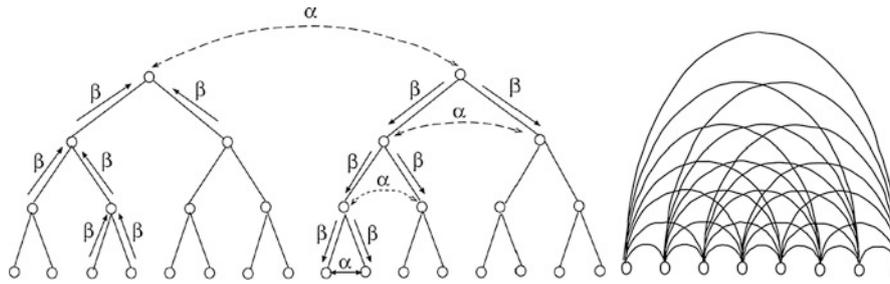


Fig. 7 The factorization of the matrix L_{ij} in Eq. 42 allows the matrix-vector product to emulate a telephone network connection (left) (Chew et al. 2001). The direct line connections between telephones (right) requires a large number of lines

The factorization allows a matrix-vector product to emulate a multilevel telephone network connection as shown in Fig. 7, where the number of telephone lines can be greatly reduced compared to a direct telephone connection. In the above, the $\tilde{\alpha}$ and $\tilde{\beta}$ matrices are diagonal, while the $\tilde{\mathbf{I}}$ matrices are quasi-diagonal interpolation and ant interpolation matrices that are not square. Hence, the storage requirements of the factorized matrices are small, and they can be reused in a tree algorithm. There are $\log N$ levels in this multilevel algorithm. Each level requires $O(N)$ in workload as well as memory requirements. Hence, the CPU and memory complexity are both $O(N \log N)$. Other factorizations do not lead to $O(N \log N)$ complexity.

The Circuit Models of Electromagnetic Structures

When electromagnetic waves are guided between electronic components through interconnects such as coaxial lines, strip lines, or other type of waveguides, a circuit is formed (Pozar 2011; Celik et al. 2002). The electronic circuit system is the static or quasi-static approximation of electrodynamics. Both fundamental circuit principles, Kirchhoff's current law (KCL) and Kirchhoff's voltage law (KVL), can be derived from Maxwell's equations (Pillegi et al. 1995; Balanis 2012a). Antennas can be considered as terminating devices in the circuit system for matching the impedance of the circuit to that of the propagation channel within a certain frequency band. However, unlike most quasi-static or static circuit components, the antenna's working principle has to depend on the dynamic field. Due to the displacement current, the electric field and the magnetic field couple to each other, which enables wave propagation (Chew 1990). It means that the reliable antenna modeling methods have to be based on the full-wave analysis. However, the circuit model of radiating structures can help to integrate their designs with other circuit subsystems to characterize and optimize the system performance. Hence, parameters such as the input impedance and return loss are popularly used. In practice, because a lot of antennas behave as the low Q resonating tank, many of their designs also rely on circuit interpretations for the performance optimization. Hence, circuit modeling methodologies for distributive electromagnetic structures will be the focus in this part.

Basic Circuit Laws from Maxwell's Equations

Resistance

For a current filament in a segment of the metallic wire with the finite conductivity σ , if the internal electric field \mathbf{E} is uniform, the conducted current due to \mathbf{E} is

$$\mathbf{J} = \sigma \mathbf{E} \quad (43)$$

Both current density and field are vectors in the same direction for isotropic media. For the current component flowing along the x direction in the Cartesian coordinate system, it is only related to the x component of \mathbf{E}

$$J_x = \sigma E_x \quad (44)$$

For the static field, the electric field is equal to the negative gradient of the potential. If the conductor cross-section area that is orthogonal to the x direction is ΔS and J_x is assumed to be constant over a very small length ΔL_x , one can derive the current through the conductor along the x direction from Eq. 44

$$I_x = \frac{\Delta V_x}{R_x} \quad (45)$$

where ΔV_x is the potential (voltage) drop along the x direction and R_x is

$$R_x = \frac{\Delta L_x}{\sigma \Delta S} = \frac{\rho \Delta L_x}{\Delta S} \quad (46)$$

where ρ is the conventional resistivity. Hence, under the static approximation, Eq. 44 is the circuit Ohm's law in the x direction while Eq. 43 is the circuit Ohm's law in all directions. For electrodynamics, the electric field is not a simple gradient of the scalar potential. The current distribution could change rapidly as a function of space. Hence, Ohm's circuit law can only be applied for electrically small parts. For electrically large circuit systems, the full circuit model truly based on electrodynamics will be very complicated.

Inductance

Inductance characterizes the ability of a conductor in creating a voltage in either itself or other conductors through the current flowing through it. Based on Faraday's law, this definition is usually connected with looped conductors (Pozar 2011; Balanis 2012a). However, with the help from auxiliary potentials, one can define the partial inductance that is also consistent with the regular inductance definition. The relationship between the induced voltage v , current i , and inductance L is defined as

$$v = L \frac{\partial i}{\partial t} \quad (47)$$

If the time-dependent voltage v is measured on the current loop itself, L is called the self-inductance. If it is measured on another loop, L is called the mutual inductance.

It is also convenient to define the inductance through the magnetic flux when the dimension of the circuit system is electrically small. If ψ is the total magnetic flux through the loop generated by a current i , one has

$$\psi = Li \quad (48)$$

Assume that a loop is made of the perfect electric conductor and its cross section is zero. If the operating frequency is so low that the loop is electrically small, then the current on the loop has almost zero phase change. However, for the high-frequency case, the current changes rapidly. The conventional loop-based inductance will become improper for the physical interpretation and modeling.

Capacitance

The capacitance is the measure of the charge capacity of a conductor. It usually needs to be relative to a certain reference voltage or potential. If it is relative to infinity, it is the self-capacitance. If it is relative to another conductor, it is the mutual capacitance with that conductor. A classic definition of the capacitance is

$$C = \frac{Q}{V} \quad (49)$$

where Q is the net charge on the conductor while V is the bias DC voltage on the conductor relative to its reference (the potential of its reference is usually considered as 0). This definition is in a dual relationship with Eq. 48. Similar to Eq. 47, another definition can be written as (Balanis 2012a)

$$i = C \frac{\partial v}{\partial t} \quad (50)$$

Hence, the capacitance causes the branch currents between the conductor and its reference. Under the dynamic situation, the voltage variation at different locations of the conductor could be highly different and the potential itself is a limited term in describing the dynamic field. Hence, the practical capacitance definition for the full-wave case is more complicated than the static case.

Generally, the capacitance is less sensitive to frequencies since most conductors being used are of high conductivities. Hence, the static capacitance approximation based on Eq. 49 works very well in many practical modeling methods.

KVL

Based on the integral form of Faraday's law, the overall voltage drop along a closed loop can be written as

$$\sum_c v_c = \oint_C d\mathbf{l} \cdot \mathbf{E} = -\frac{\partial \psi}{\partial t} \quad (51)$$

where the total magnetic flux is $\psi = \int_S d\mathbf{s} \cdot \mathbf{B}$, the subscript c could be R , L , or C , v_R means the voltage drop due to resistors in the loop, v_L means the voltage drop due to inductors in the loop, and v_C means the voltage drop due to capacitors in the loop. Using Eq. 48, one has

$$\sum_c v_c = -L_s \frac{\partial i}{\partial t} \quad (52)$$

or

$$\sum_c v_c + L_s \frac{\partial i}{\partial t} = 0 \quad (53)$$

Here, L_s is the parasitic loop self-inductance when all lumped components are shorted. Sometimes it is also called the strayed inductance (Balanis 2012a). When the circuit dimension is electrically small, L_s is a very small value. For example, a circular loop with a 1 cm radius and 1 mm wire cross-section radius has a self-inductance of 2.99×10^{-8} H. And it is already a very big circuit loop. Hence, if one ignores the

parasitic voltage introduced by L_s under the low-frequency assumption, Eq. 53 becomes the standard Kirchhoff's voltage law (KVL):

$$\sum_c v_c = 0 \quad (54)$$

But if the frequency is very high, the voltage introduced by the self-inductance cannot be ignored. Conventional KVL will not be accurate anymore. Equation 53 has to be used to consider the effect of L_s .

KCL

Based on the continuity equation, a circuit node enclosed by a small closed volume V satisfies

$$\oint_S d\mathbf{S} \cdot \mathbf{J} + \frac{\partial}{\partial t} \int_V dr \rho = 0 \quad (55)$$

where ρ is the volumetric charge density while \mathbf{J} is the current density. Under the low-frequency approximation, currents only flow through metal wires. The closed surface integral of the current density becomes the summation of branch currents. Further, by Eqs. 49 and 50, one has

$$\sum_b I_b + C_n \frac{\partial v_n}{\partial t} = 0 \quad (56)$$

This equation is also for low frequencies (Balanis 2012a). The subscript b refers to all branches. Here, v_n is the node voltage while C_n is the total parasitic self-capacitance of all branches connected to the node relative to the zero potential ground at infinity. It can be computed by excluding all lumped components inside V and computing the total self-capacitance of all conductors. At low frequencies, C_n is so small that it could be approximated as zero. Then the standard Kirchhoff's current law (KCL) is achieved:

$$\sum_b I_b = 0 \quad (57)$$

At high frequencies, the parasitic current introduced by the self-capacitance cannot be ignored. The conventional KCL based on the pure circuit theory will not be accurate anymore.

Partial Element Equivalent Circuit Method

The partial element equivalent circuit (PEEC) method was developed by Albert Ruehli to solve parasitic coupling problems by building the equivalence circuit based on the integral equation (Ruehli 1972, 1974; Ruehli and Brennan 1973). It was initiated for static or quasi-static applications with no retardation (delay) considered. Later it was extended to the full-wave case where the retardation is included. PEEC is a very convenient bridging method between electromagnetics and circuit theories. It is broadly used by many electronic design automation (EDA) softwares. It also greatly helps the modeling of resonating structures such as antennas. Even though the PEEC method solves the problem in the circuit solver, its formulation actually starts from the integral equation (Chew 1990). The modified nodal analysis (MNA)-based circuit solvers (Ho et al. 1975) basically consider the electric field and magnetic field decoupling effect happening at the low-frequency regime. Hence, it involves both KCL and KVL in the formulation. Some low-frequency full-wave integral equation simulation methods have also been developed based on

the same physical insight, such as the augmented electric field integral equation method (A-EFIE) (Qian and Chew 2009).

Partial Inductance

Assume that there are two loops i and j with currents I_i and I_j (Ruehli 1972). For low frequencies, both I_i and I_j are approximated to be constant over each loop. Then the mutual inductance L_{ij} between two loops is

$$L_{ij} = \frac{\psi_{ij}}{I_j}, \quad (58)$$

where ψ_{ij} is the magnetic flux through loop i due to the current in loop j . It can be further written as

$$\psi_{ij} = \int_{S_i} d\mathbf{S}_i \cdot \mathbf{B}_{ij} = \frac{1}{a_i} \int_{a_i} da_i \int_i d\mathbf{l}_i \cdot \mathbf{A}_{ij}, \quad (59)$$

where S_i is the loop surface of loop i , \mathbf{B}_{ij} is the magnetic flux density, \mathbf{A}_{ij} is the magnetic vector potential at loop i due to the current in loop j , and a_i is the cross-section area of the loop i wire. The vector potential \mathbf{A}_{ij} is formulated through an integral using the homogeneous medium Green's function as its kernel:

$$\mathbf{A}_{ij} = \oint_{\text{loop}_j} d\mathbf{r}' \frac{e^{-jk|\mathbf{r}_i - \mathbf{r}'|}}{4\pi|\mathbf{r}_i - \mathbf{r}'|} \mu \mathbf{J}_j(\mathbf{r}'). \quad (60)$$

Assuming that the working frequency is low and cross sections of wires are always small, then the current is almost uniform along the loop and at the cross section. Hence, one has

$$\mathbf{A}_{ij} = \frac{\mu}{4\pi} \frac{I_j}{a_j} \int_{a_j} da_j \oint_j d\mathbf{l}_j \left[\frac{1}{r_{ij}} \right], \quad (61)$$

where I_j is the loop j 's current from the common circuit point of view. Using Eqs. 58 and 59, one has

$$L_{ij} = \frac{1}{a_i a_j} \frac{\mu}{4\pi} \oint_i \oint_j \int_{a_i} \int_{a_j} \left[d\mathbf{l}_i \cdot d\mathbf{l}_j da_i da_j \frac{1}{r_{ij}} \right]. \quad (62)$$

If a_i and a_j are small (current filaments), it can be further simplified to be independent from the cross section

$$L_{ij} = \frac{\mu}{4\pi} \oint_i \oint_j \left[d\mathbf{l}_i \cdot d\mathbf{l}_j \frac{1}{r_{ij}} \right]. \quad (63)$$

If the wire cross section is not small, the wire has to be split along the longitudinal direction into many current filaments. Each current filament is assumed to have a constant current at its cross section. Then Eq. 63 depicts the result of a group of inductors connected in a parallel fashion.

With the thin- and short-wire assumption, if one divides each loop into small segments, the mutual inductance is the summation of partial inductive contributions. For example, if loop i is partitioned into K pieces while loop j is divided into M pieces, then

$$L_{ij} = \sum_{k=1}^K \sum_{m=1}^M \frac{\mu}{4\pi} \frac{1}{a_{ik}a_{jm}} \int_{l_{ik}} \int_{a_{ik}} \int_{l_{jm}} \int_{a_{jm}} da_{ik} da_{jm} \frac{d\mathbf{l}_{ik} \cdot d\mathbf{l}_{jm}}{r_{km}}, \quad (64)$$

where subscript ik means the k -th current filament on loop i , jm means the m -th current filament on loop j , and p means *partial*. Here, a_{ik} is the cross-section area of the k -th current filament on loop i and l_{ik} is the k -th current filament of loop i . Other symbols are defined in a similar way. If one defines the kernel of the double summations in Eq. 64 to be the partial inductance $L_{p_{ij}}^{km}$, one has

$$L_{p_{ij}}^{km} = \frac{\mu}{4\pi} \frac{1}{a_{ik}a_{jm}} \int_{l_{ik}} \int_{a_{ik}} \int_{l_{jm}} \int_{a_{jm}} da_{ik} da_{jm} \frac{|d\mathbf{l}_{ik} \cdot d\mathbf{l}_{jm}|}{r_{km}}. \quad (65)$$

Then the total inductance is the summation of partial inductances

$$L_{ij} = \sum_k \sum_m S_{ij}^{km} L_{p_{ij}}^{km}, \quad (66)$$

where S_{ij}^{km} is the sign of the dot project of $d\mathbf{l}_{ik} \cdot d\mathbf{l}_{jm}$. Hence, it is either +1 or -1. If one ignores the integration over cross sections, the partial inductance is simplified to

$$L_{p_{ij}}^{km} = \frac{\mu}{4\pi} \int_{l_{ik}} \int_{l_{jm}} \frac{|d\mathbf{l}_{ik} \cdot d\mathbf{l}_{jm}|}{r_{km}} \quad (67)$$

The physical meaning of $L_{p_{ij}}^{km}$ is very intriguing. If one places both current filaments l_{ik} and l_{jm} on the same plane, as shown in Fig. 8, the projection lines from ending points of current ik to jm will form an infinite loop with current ik and the infinity. Since the integration kernel (static Green's function) is zero at infinity, l_{jm} is orthogonal to both projection lines, and $\int_{l_{ik}}$ can be written as a loop integral in the partial inductance definition (Eq. 62) (if one ignores the absolute value operation $|\cdot|$ in the kernel). Hence,

$$L_{p_{ij}}^{km} = \frac{\mu}{4\pi} \oint_{l_{ik}} d\mathbf{l}_{ik} \cdot \int_{l_{jm}} \frac{d\mathbf{l}_{jm}}{r_{km}} = \oint_{l_{ik}} d\mathbf{l}_{ik} \cdot \frac{\mathbf{A}_{i,km}}{I_{jm}} = \frac{1}{I_{jm}} \int_{S_{ik}} d\mathbf{S} \cdot \mathbf{B}_{i,km}. \quad (68)$$

where $\mathbf{A}_{i,km}$ is the magnetic vector potential defined in Eq. 60 generated inside the infinite loop by the current filament l_{jm} . Hence, even though the formulation of the partial inductance does not have a complete loop, it actually uses the loop formed with infinity through current project lines,

$$L_{p_{ij}}^{km} = \frac{\psi_{ik,jm}}{I_{jm}}. \quad (69)$$

The above definitions and properties are also applicable to cases when $i = j$ and $k = m$. It means that they are valid for self-inductances too. Using the mutual partial inductance as the reference, it can be understood that the self-partial inductance corresponds to the magnetic flux through a loop formed by the infinity, two orthogonal projection lines, and the current filament itself.

If two current filaments are both on the same rectangular loop and parallel to each other, they will have opposite current flow directions, as shown in Fig. 9. Then in Eq. 66 $S_{ii}^{km} = -1$ is for the mutual partial inductance $L_{p_{ii}}^{km}$ and $S_{ii}^{mm} = +1$ is for the self-partial inductance $L_{p_{ii}}^{km}$. Hence, the magnetic flux area from l_{ik}

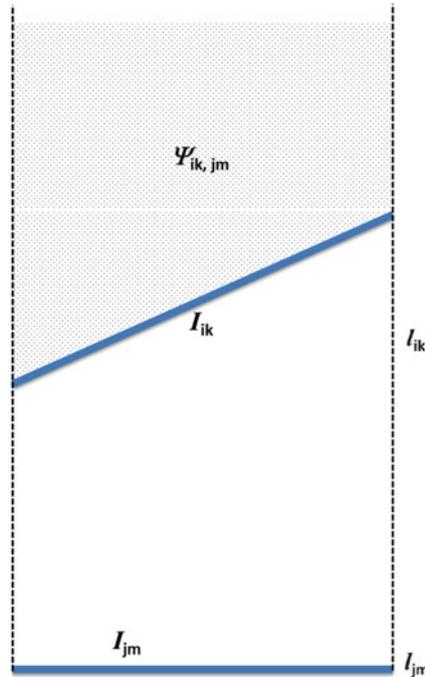


Fig. 8 The magnetic flux loop for the partial inductance

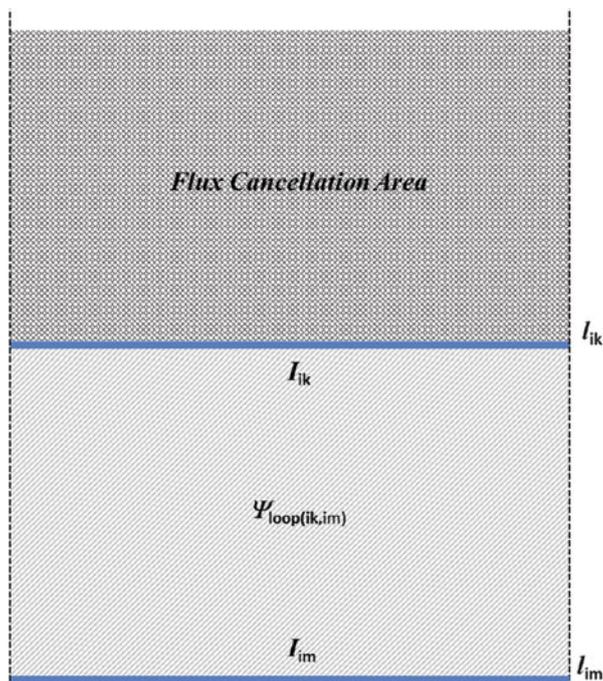


Fig. 9 The magnetic flux loop for the loop based on partial inductances

to infinity is completely canceled from the area from l_{im} . Hence, only the flux area in between the two current filaments is kept. When this area is closed by another pair of vertical current filaments, it forms a complete loop and the loop area will directly determine the total inductance, which is consistent with the conventional circuit concept.

PEEC Formulation

Assume that circuit conducting interconnects are partitioned into N_L current filaments (Ruehli et al. 2003, 2013; Rong and Cangellaris 2001). If the conductivity is finite, its time-domain electric field for one wire segment is

$$\mathbf{E}(\mathbf{r}, t) = \frac{\mathbf{J}(\mathbf{r}, t)}{\sigma} + \frac{\partial \mathbf{A}(\mathbf{r}, t)}{\partial t} + \nabla \phi(\mathbf{r}, t) \quad (70)$$

The vector potential \mathbf{A} is contributed from all current filaments including itself. Hence, it can be written as

$$\mathbf{A}(\mathbf{r}, t) = \sum_{n_l=1}^{N_L} \frac{\mu}{4\pi} \int_{V_{n_l}} d\mathbf{r}' G(\mathbf{r}, \mathbf{r}') \mathbf{J}_{n_l}(\mathbf{r}', t') \quad (71)$$

where \mathbf{J}_{n_l} is the current density vector on the n_l -th current filament and $G(\mathbf{r}, \mathbf{r}')$ is the Green's function

$$G(\mathbf{r}, \mathbf{r}') = \frac{1}{|\mathbf{r} - \mathbf{r}'|}. \quad (72)$$

t' is the source time with the retardation

$$t' = t - \frac{|\mathbf{r} - \mathbf{r}'|}{c/\sqrt{\mu_r \epsilon_r}}. \quad (73)$$

The scalar potential ϕ is contributed from the charge distribution. Since good conductors are usually used in the design, there is almost no charge inside the conductor. Hence, all charges are on the conductor surface. Assuming that all surfaces are partitioned into N_C nonoverlapping patches and the charge density on each patch is generally defined as ρ , the scalar potential is written as

$$\phi(\mathbf{r}, t) = \sum_{n_c=1}^{N_C} \frac{1}{4\pi\epsilon} \int_{S_{n_c}} d\mathbf{r}' G(\mathbf{r}, \mathbf{r}') \rho_{n_c}(\mathbf{r}', t') \quad (74)$$

Hence, Eq. 70 can be written as

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \frac{\mathbf{J}(\mathbf{r}, t)}{\sigma} + \sum_{n_l=1}^{N_L} \frac{\mu}{4\pi} \int_{V_{n_l}} d\mathbf{r}' G(\mathbf{r}, \mathbf{r}') \frac{\partial \mathbf{J}_{n_l}(\mathbf{r}', t')}{\partial t} \\ &\quad + \sum_{n_c=1}^{N_C} \frac{1}{4\pi\epsilon} \nabla \int_{S_{n_c}} dS' G(\mathbf{r}, \mathbf{r}') \rho_{n_c}(\mathbf{r}', t') \end{aligned} \quad (75)$$

For the m -th current filament, it is either inside or on the surface of the conductor. Hence, the electric field along its filament direction is always zero if one does Galerkin's testing along the current filament. Hence, one can have

$$\begin{aligned} & \frac{1}{\sigma} \int_{V_m} d\mathbf{r}_m J_m(\mathbf{r}, t) \\ & + \sum_{n_l=1}^{N_L} \frac{\mu}{4\pi} \int_{V_m} \int_{V_{n_l}} d\mathbf{r}_m d\mathbf{r}'_{n_l} G(\mathbf{r}, \mathbf{r}') \frac{\partial J_{n_l}(t'_m)}{\partial t} S_{m, n_l} \\ & + \sum_{n_c=1}^{N_C} \frac{1}{4\pi\epsilon} \int_{V_m} d\mathbf{r}_m \mathbf{u}_m \cdot \nabla \int_{S_{n_c}} ds'_{n_c} G(\mathbf{r}, \mathbf{r}') \rho_{n_c}(\mathbf{r}', t') = 0. \end{aligned} \quad (76)$$

Its first term corresponds to the resistive voltage drop

$$V_{R_m} = \frac{1}{\sigma} \int_{V_m} d\mathbf{r}_m J_m(\mathbf{r}, t) = \frac{I_m}{\sigma} \int_{l_m} dl_m = \frac{l_m}{\sigma} I_m = R_m I_m. \quad (77)$$

where l_m is the longitudinal length of the m -th current filament. Hence, the first term shows the ohmic voltage drop due to the lossy conductor.

The second term corresponds to the inductance effect. Here, S_{m, n_l} is the vector dot between the testing current direction and the source current direction. If one assumes all filaments are along the Cartesian coordinate axes, $S_{m, n_l} = +1, -1$, or 0 depending on their relative directions. Using the partial inductance concept introduced in the previous subsection, one has

$$V_{L_m} = \sum_{n_l=1}^{N_L} \frac{\mu}{4\pi} \frac{1}{a_m a_{n_l}} \int_{V_m} \int_{V_{n_l}} d\mathbf{r}_m d\mathbf{r}'_{n_l} G(\mathbf{r}, \mathbf{r}') \frac{\partial J_{n_l}(t'_m)}{\partial t} S_{m, n_l} \quad (78)$$

$$= \sum_{n_l=1}^{N_L} L_p^{m n_l} \frac{d}{dt} I_{n_l}(t'_m). \quad (79)$$

Hence, the second term is the voltage drop due to partial inductances. It shall be noted that there are two types of inductances contributed to V_{L_m} : one is the self-partial inductance and another one is the mutual partial inductance. Because there are many current filaments in reality, mutual partial inductances are much more complicated than the self-partial inductance.

The third term corresponds to the capacitive coupling. It could be represented by potentials or capacitances. It uses the central difference to approximate the gradient where \mathbf{u}_m is the unit vector direction of the m th current flow. Assume that u_m represents x, y or z direction, then

$$\begin{aligned} V_{C_m} &= \sum_{n_c=1}^{N_C} \frac{a_m}{4\pi\epsilon} \left[\int_{S_{n_c}} dS' G(\mathbf{r}_{m+}, \mathbf{r}') \rho_{n_c}(\mathbf{r}', t') - \int_{S_{n_c}} dS' G(\mathbf{r}_{m-}, \mathbf{r}') \rho_{n_c}(\mathbf{r}', t') \right] \\ &= \sum_{n_c=1}^{N_C} \frac{1}{4\pi\epsilon} \rho_{n_c} \cdot \left[\int_{S_{n_c}} dS' G(\mathbf{r}_{m+}, \mathbf{r}') - \int_{S_{n_c}} dS' G(\mathbf{r}_{m-}, \mathbf{r}') \right] \\ &= \sum_{n_c=1}^{N_C} Q_{n_c} \left[pp_{i(m, n_c)}^+ - pp_{i(m, n_c)}^- \right]. \end{aligned} \quad (80)$$

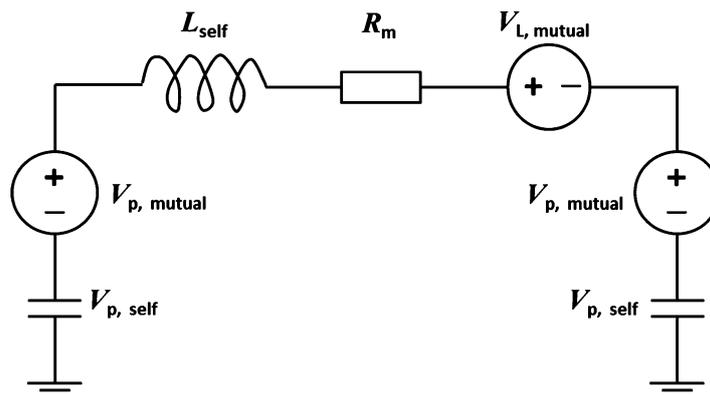


Fig. 10 The equivalent circuit of the PEEC model

where

$$pp_{i(m, n_c)}^{\pm} = \frac{1}{4\pi\epsilon} \frac{1}{a_{n_c}} \int_{S_{n_c}} dS' G(\mathbf{r}_{m\pm}, \mathbf{r}'). \quad (81)$$

Here, the position \mathbf{r}_{m+} and \mathbf{r}_{m-} are the two end points of the m -th current filament. The current filaments' terminating nodes are located at the centers of piecewise constant charge density patches. This is very similar to the cell structure in the FDTD algorithm (Chew 1990). The third term therefore is interpreted as the charge-controlled voltage source (CCVC) in the equivalent circuit. By considering all three terms and their relationships with other parts of the system, an equivalent circuit can be easily built, as shown in Fig. 10.

The L_{self} is the self-partial inductance, R_m is the ohmic loss due to the finite conductivity, and $V_{i,mutual}$ is the voltage generated by other currents from other places through partial mutual inductances. Here, $V_{p,mutual}$ is the voltage generated by other branch currents from other places through partial mutual potentials, and $V_{p,self}$ is the voltage generated by the self-capacitive coupling or the self-partial potential coefficient.

Through this approach, the distributive structure can be modeled using a circuit. Since each electrically small piece could contribute to the mutual partial inductance and capacitively coupled potentials, the resultant circuit is a densely coupled one. If this circuit is solved by the circuit solver SPICE (Nagel and Pederson 1973; Warwick 2009), it results in a dense numerical matrix. Its solving process will be very expensive. Many methods could be employed to further solve this problem. It shall be noted that conventional SPICE solvers do not handle retardation. It means that the true full-wave circuit models cannot be solved by SPICE directly. Hence, certain modifications to SPICE solvers are necessary for handling circuit models based on the PEEC method.

Low-Frequency Computational Electromagnetics

An electromagnetic field can be obtained by solving the wave equations with enforced boundary condition. Because of the linearity of problems, the solution can be obtained in the frequency domain by Fourier transforming the equations in time, which theoretically covers the frequency range from static to optical regime. Generally speaking, the problems in solving the wave equations in the frequency domain can be classified into three regimes, which are defined as low-frequency regime, mid-frequency regime, and high-frequency regime (Chew et al. 2008). For low frequency, the typical electrical size (L) is much less than wavelength (λ), i.e., $L \ll \lambda$. At low frequency, the electromagnetic field physics is close to

that of the static field, where the magnetic field and the electric field are decoupled from each other. Hence, the electromagnetic wave equations reduce to the equation of electrostatics and the equation of magnetostatics that are completely decoupled from each other (Zhao and Chew 2000). Furthermore, the electrostatics equation or Poisson's equation governs the capacitive circuit physics, while the magnetostatics equation or Laplace's equation governs the inductive circuit physics. In the mid-frequency regime, the wavelength is comparable with the object size, $L \approx \lambda$. In this regime, the electric and magnetic fields are tightly coupled together. Thus, the electromagnetic fields become waves and are oscillatory as is typical of wave physics. In the high-frequency regime, the wavelength is much less than the object size, $L \gg \lambda$. The plane waves of the ray physics thus become dominant. In this section, our discussion will be focused on the low-frequency computation electromagnetics, which becomes more and more important in recent interdisciplinary research.

Understanding of Low-Frequency Region

Low-frequency computational electromagnetics has been studied for many years (Chew et al. 2008; Zhao and Chew 2000). Theoretically, the current or charge at static limit does not radiate but generates the field which is inversely proportional to the square of the distance from the source. For the on-chip-level antenna design, the typical size of the interconnects and devices is usually so much smaller than the wavelength (nanometer to micrometer levels in the standard CMOS process).

It implies that the surrounding field is primarily quasi-static; that is why circuit designers can utilize the lumped-circuit model instead of using the distributed-circuit model for the passive components, such as the interconnects, the inductors, and the capacitors. Hence, in order to capture the circuit physics with a universal solver, the well-developed full-wave solvers have to be revisited and corrected at low-frequency limit.

When $\omega \rightarrow 0$, the four Maxwell equations can be reduced into two parts: the magnetostatic part:

$$\nabla \times \mathbf{H} = \mathbf{J} \quad (82)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (83)$$

and the electrostatic part:

$$\nabla \times \mathbf{E} = 0 \quad (84)$$

$$\nabla \cdot \mathbf{D} = \rho = \lim \frac{\nabla \cdot \mathbf{J}}{i\omega} \quad (85)$$

Note that the current \mathbf{J} in the electrostatic part must be of the order of ω in order to keep the right-hand side bounded. According to the Helmholtz decomposition (Collins 1966), the total current \mathbf{J} naturally decomposes into a solenoidal (divergence-free) part and an irrotational (curl-free) part as

$$\mathbf{J} = \mathbf{J}_{sol} + \mathbf{J}_{irr} \quad (86)$$

where the solenoidal part \mathbf{J}_{sol} represents eddy currents that produce primarily the magnetic field and the irrotational part \mathbf{J}_{irr} represents charge currents that produce primarily the electric field at low frequencies. Both of them are equally important in capturing the inductive and capacitive circuit physics. When $\omega \rightarrow 0$, \mathbf{J}_{irr} becomes much smaller than \mathbf{J}_{sol} ; hence, the former usually cannot be captured correctly in the numerical calculation, due to the limit of computer precision. This is also known as the low-frequency inaccuracy problem (Chew et al. 2008). The problem in the electric field integral equation (EFIE) and its

solutions will be discussed in the following sections. On the other hand, the similar low-frequency problem also happened in the time-domain EFIE (TD-EFIE) (Andriulli et al. 2009), while the low-frequency inaccuracy problem happened in the magnetic field integral equation (MFIE) (Zhang et al. 2003).

Low-Frequency Problems in EFIE Formulation

Due to the decoupling between the electric field and magnetic field, there is a low-frequency breakdown problem in the EFIE formulation. The reasons for this breakdown can be classified into two catalogs: the frequency breakdown and the mesh density breakdown. The frequency breakdown occurs because the contribution of the vector potential is swamped by that of the scalar potential at low frequencies. Moreover, it will evoke the null space of the divergence operator in the scalar potential term and make the MoM matrix become extremely ill conditioned. Basically, the traditional EFIE operator can be written in its mixed potential form as

$$\mathcal{T}(\mathbf{J}) = \mathcal{T}_s(\mathbf{J}) + \mathcal{T}_h(\mathbf{J}) \quad (87)$$

in which the smoothing (\mathcal{T}_s) and hypersingular (\mathcal{T}_h) terms are defined as

$$\mathcal{T}_s(\mathbf{J}) = i\omega\mu\hat{\mathbf{n}}_r \times \int_{S'} dS' g(\mathbf{r}, \mathbf{r}') \mathbf{J} \quad (88)$$

$$\mathcal{T}_h(\mathbf{J}) = -\frac{1}{i\omega\varepsilon} \hat{\mathbf{n}}_r \times \nabla \int_{S'} dS' g(\mathbf{r}, \mathbf{r}') \nabla' \cdot \mathbf{J} \quad (89)$$

where

$$g(\mathbf{r}, \mathbf{r}') = \frac{e^{ik_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \quad (90)$$

is the free-space Green's function, k_0 is the wave number in the free space, ε and μ are the relative permeability and permittivity, and \mathbf{J} is the surface current on an arbitrarily shaped PEC surface S' whose outward pointing unit normal at \mathbf{r} is denoted by $\hat{\mathbf{n}}_r$. The domain of both \mathcal{T}_s and \mathcal{T}_h is the surface S' . Here, \mathcal{T}_s is referred to as a smoothing operator because the range space of the operator is smoothed by one order relative to its domain, while the hypersingular operator \mathcal{T}_h lowers the smoothness of the domain space.

When $\omega \rightarrow 0$, the hypersingular term \mathcal{T}_h which is $O(\omega^{-1})$ dominates over the smoothing term \mathcal{T}_s which is $O(\omega^1)$. As mentioned above, because of the existence of the divergence operator in Eq. 89 and $\nabla' \cdot \mathbf{J}_{sol} = 0$, \mathcal{T}_h has a null space. Thus, the ill-conditioned \mathcal{T} behaves like a first-kind operator between the solenoidal and irrotational subspaces (Adams 2004). This makes the impedance matrix nearly singular and unsolvable at low frequencies (Zhao and Chew 2000). This is the so-called low-frequency breakdown problem for the EFIE operator.

Low-Frequency Remedies

In past decades, many research efforts have been carried out to avoid the imbalance inherent in the traditional EFIE operator. The most popular one is the loop-tree or loop-star decomposition, which separates the electrostatic and magnetostatic physics at low frequencies (Zhao and Chew 2000; Vecchi 1999; Lee et al. 2003). By introducing the charge in the unknown list and enforcing the current continuity constraint, the augmented version of EFIE was proposed, which is known as A-EFIE (Qian and Chew

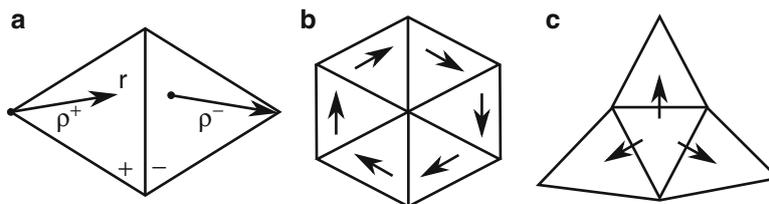


Fig. 11 Representations of (a) RWG, (b) loop, and (c) star basis functions

2008). The frequency scaling can be normalized in a balanced manner, thus remedying the low-frequency breakdown problem. In addition, the success of the developed Calderón multiplicative preconditioned EFIE (CMP-EFIE) method (Hsiao and Kleinman 1997; Adams 2004) has also made the original EFIE into a second-kind Fredholm integral equation operator, which is the so-called self-regularizing property of the \mathcal{T} operator and immune to mesh density breakdown. Unfortunately, although stable at low frequencies, A-EFIE and CMP-EFIE all suffer from the low-frequency inaccuracy problem for the solved currents. As a remedy, the perturbation method has been applied to capture the accurate currents at different frequency orders (Qian and Chew 2010; Sun et al. 2013b).

Loop-Tree Decomposition

Instead of using the popular Rao–Wilton–Glisson (RWG) basis function, one can decompose the RWG basis function into the loop basis function whose members have zero divergence and the tree basis (or star basis) function whose members have nonzero divergence, as shown in Fig. 11. Here, $\rho^\pm(\mathbf{r})$ is the vector from the point \mathbf{r} to the apex of the respective triangles, and \pm is the support of the respective triangles. This is known as the loop-tree or loop-star decomposition, which is a quasi-Helmholtz decomposition because the tree or the star expansion functions are not curl-free. In some work based on the finite-element method, this is also known as the tree-cotree decomposition (Manges and Cendes 1995). Here, one shall call the space spanned by the RWG basis function as the RWG space, the subspace spanned by the tree (or star) basis function as the tree (or star) space, and that spanned by the loop basis function as the loop space. Therefore, the total surface current can be expanded as

$$\mathbf{J}(r') = \sum_{n=1}^{N_L} I_{Ln} \mathbf{J}_{Ln}(r') + \sum_{n=1}^{N_C} I_{Cn} \mathbf{J}_{Cn}(r') \quad (91)$$

where $\mathbf{J}_{Ln}(r')$ is the divergence-free loop basis function such that $\nabla \cdot \mathbf{J}_{Ln}(r') = 0$ and $\mathbf{J}_{Cn}(r')$ is the nondivergence-free tree (or a star) basis function such that $\nabla \cdot \mathbf{J}_{Cn}(r') \neq 0$, and it is used to model the charge in the system. The above can be written in the matrix form as

$$\mathbf{J}(r') = \mathbf{J}_L^t(r') \cdot \mathbf{I}_L + \mathbf{J}_C^t(r') \cdot \mathbf{I}_C \quad (92)$$

where the four column vectors in the right-hand side contain the expansion and coefficient elements in Eq. 91. By substituting the above into the EFIE and testing the result with the same set of the functions as in Galerkin's method, one has a matrix equation as

$$\begin{bmatrix} \bar{\mathbf{Z}}_{LL} & \bar{\mathbf{Z}}_{LC} \\ \bar{\mathbf{Z}}_{CL} & \bar{\mathbf{Z}}_{CC} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I}_L \\ \mathbf{I}_C \end{bmatrix} = \begin{bmatrix} \mathbf{V}_L \\ \mathbf{V}_C \end{bmatrix} \quad (93)$$

where

$$\mathbf{V}_L = -\langle \mathbf{J}_L(r), \mathbf{E}^{inc}(r) \rangle \quad (94)$$

$$\mathbf{V}_C = -\langle \mathbf{J}_C(r), \mathbf{E}^{inc}(r) \rangle \quad (95)$$

$$\bar{\mathbf{Z}}_{LL} = i\omega\mu \langle \mathbf{J}_L(r), g(r, r'), \mathbf{J}_L^t(r') \rangle \quad (96)$$

$$\bar{\mathbf{Z}}_{LC} = i\omega\mu \langle \mathbf{J}_L(r), g(r, r'), \mathbf{J}_C^t(r') \rangle \quad (97)$$

$$\bar{\mathbf{Z}}_{CL} = i\omega\mu \langle \mathbf{J}_C(r), g(r, r'), \mathbf{J}_L^t(r') \rangle = \bar{\mathbf{Z}}_{LC}^t \quad (98)$$

$$\begin{aligned} \bar{\mathbf{Z}}_{CC} &= i\omega\mu \langle \mathbf{J}_C(r), g(r, r'), \mathbf{J}_C^t(r') \rangle \\ &\quad - \frac{i}{\omega\epsilon} \langle \nabla \cdot \mathbf{J}_C(r), g(r, r'), \nabla' \cdot \mathbf{J}_C^t(r') \rangle \end{aligned} \quad (99)$$

In order to solve the system equation when $\omega \rightarrow 0$, the frequency normalization should be implemented as detailed in (Zhao and Chew 2000). It implies that even after the loop-tree decomposition, the matrix system is still ill conditioned because of the imbalance of the matrix elements, which is the so-called frequency-scaling behavior. Recently, the application of the loop-star/tree decomposition with the magnetic field and the Calderón preconditioned electric field integral operators is also studied (Andriulli 2012).

A-EFIE

Another remedy to the low-frequency breakdown problem is the A-EFIE, which avoids the imbalance inherent in the conventional EFIE by introducing the charge as the additional unknown. The same notations as those used in (Qian and Chew 2008) are used here. By introducing the current continuity condition, one can obtain the A-EFIE system as

$$\begin{bmatrix} \bar{\mathbf{V}} & \bar{\mathbf{D}}^T \cdot \bar{\mathbf{P}} \\ \bar{\mathbf{D}} & k_0^2 \bar{\mathbf{I}} \end{bmatrix} \cdot \begin{bmatrix} ik_0 \mathbf{j} \\ c_0 \rho \end{bmatrix} = \begin{bmatrix} \eta_0^{-1} \mathbf{b} \\ 0 \end{bmatrix} \quad (100)$$

where $\bar{\mathbf{V}}$ and $\bar{\mathbf{P}}$ are the vector potential and scalar charge matrices, respectively, $\bar{\mathbf{I}} \in \mathbf{R}^{p \times e}$ is an identity matrix, $\mathbf{j} \in \mathbf{C}^{e \times 1}$ represents the vector of current coefficient, $\rho \in \mathbf{C}^{p \times 1}$ represents the vector of charge coefficient, and $b \in \mathbf{C}^{e \times 1}$ denotes the excitation. The superscripts p and e indicate the number of patches and edges, respectively. Mathematically, it has a generalized saddle point form and can be reduced to a standard saddle point form in the low-frequency regime. Physically, this new equation system can also be understood as a complete Kirchhoff circuit law, which is solvable and stable with low-frequency circuit physics. It means that the first equation in Eq. 100 represents the Kirchhoff voltage law, while the second equation represents the Kirchhoff current law.

At very low frequencies, charge neutrality still causes rank deficiency of the A-EFIE system in Eq. 100. A simple method has been proposed to remedy this issue by dropping one charge unknown for each spanning tree, which results the A-EFIE system as

$$\begin{bmatrix} \bar{\mathbf{V}} & \bar{\mathbf{D}}^T \cdot \bar{\mathbf{P}} \cdot \bar{\mathbf{B}} \\ \bar{\mathbf{F}} \cdot \bar{\mathbf{D}} & k_0^2 \bar{\mathbf{I}}_r \end{bmatrix} \cdot \begin{bmatrix} ik_0 \mathbf{j} \\ c_0 \rho_r \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix} \quad (101)$$

where the backward matrix $\bar{\mathbf{B}}$ and forward matrix $\bar{\mathbf{F}}$ are defined to map the charge unknowns back and forth, respectively, and ρ_r denotes the charge of unknown vectors. Then, one has

$$\rho_r = \bar{\mathbf{F}} \cdot \rho, \quad \rho = \bar{\mathbf{B}} \cdot \rho_r \quad (102)$$

However, the saddle point problems always suffer from the poor spectral property due to the zero diagonal block. Hence, the constraint preconditioner is usually required in order to regularize the saddle point matrix for easier solution. Since this preconditioning matrix should be inverted first, a highly sparse matrix is preferred and is given by

$$\bar{\mathbf{M}} = \begin{bmatrix} \bar{\mathbf{V}}_d & \bar{\mathbf{D}}_r^T \cdot \bar{\mathbf{P}}_{rd} \\ \bar{\mathbf{D}}_r & k_0^2 \bar{\mathbf{I}}_r \end{bmatrix} \quad (103)$$

which can be solved by a fast direct solver like the multifrontal method. It implies that it will incur more computational cost in addition to the more unknowns in comparison with the traditional EFIE system. More details about this constraint preconditioning technique can be found in Benzi et al. (2005). Alternatively, the Calderón multiplicative preconditioner can be further implemented on A-EFIE, which does not need inverse preconditioning and avoids the saddle point problems as well (Sun et al. 2013a).

Calderon Preconditioning

Another popular remedy is to utilize the Calderón identity, which can be expressed as (Adams 2004).

$$\mathcal{T}^2(\mathbf{J}) = -\frac{\mathbf{J}}{4} + \mathcal{K}^2(\mathbf{J}) \quad (104)$$

It can be used to conduct a new operator by projecting the operator \mathcal{T} of Eq. (87) into itself. In the above, \mathcal{K} is the magnetic field integral equation (MFIE) operator (Chew et al. 2008)

$$\mathcal{K}(\mathbf{J}) = \hat{\mathbf{n}}_r \times \nabla \times \int_{S'} dS' g(\mathbf{r}, \mathbf{r}') \mathbf{J}. \quad (105)$$

It is important to notice from Eq. 104 that the composite operator \mathcal{T}^2 is actually a second-kind integral operator with its spectrum accumulating at -0.25 . In other words, the \mathcal{T} operator can be utilized to precondition itself, the so-called “self-regularizing property”.

A multiplicative form of the Calderón multiplicative preconditioner (CMP) (Andriulli et al. 2008) was developed based on the div- and quasi-curl-conforming basis function, called the Buffa–Christiansen (BC) basis function (Buffa and Christiansen 2007). This preconditioning is straightforward to implement and easily integrated into existing MoM codes based on RWG basis functions. For the discretization of composite operator \mathcal{T}^2 , the inner \mathcal{T} operator is discretized by using div-conforming RWG basis function (source) and $\hat{\mathbf{n}}_r \times$ RWG basis function, while the outer \mathcal{T} operator is discretized by using div- and quasi-curl-conforming BC basis function (source) and $\hat{\mathbf{n}}_r \times$ BC basis function. Therefore, the resulting Gram matrix is a mixed curl- and quasi-curl-conforming matrix, which is highly sparse and invertible. Subsequently, the CMP is applied on the combined field integral equation (CFIE) formulation for PEC objects (Bağcı et al. 2009) and also for the single-source integral equations (Valdés et al. 2011).

It should also be noted that the BC basis function represents a subset of the functions proposed by Chen and Wilton in 1990 (Chen and Wilton 1990), which was named the dual basis. In other words, the idea of

these two basis functions is identical. Both of these two basis functions are a linear combination of RWG basis functions on the barycentrically refined triangles within a polygon pair and have the same dual basis property, i.e., approximately orthogonal to the original RWG basis function (Tong et al. 2009). The reason why the BC basis function has received more attention in the electromagnetics community is because of its success in the CMP (Andriulli et al. 2008), where the well-conditioned nature of the Gram matrices linking the BC basis functions to $\hat{n}_r \times$ RWG basis functions is ensured. Hence, one should rightfully call them the Chen–Wilton–Buffa–Christiansen (CWBC) basis function.

According to the decomposition in Eq. 87, the composite operator \mathcal{T}^2 can be further expanded as

$$\mathcal{T}^2 = \mathcal{T}_s^2 + \mathcal{T}_s \mathcal{T}_h + \mathcal{T}_h \mathcal{T}_s + \mathcal{T}_h^2 \quad (106)$$

where

$$\mathcal{T}_s^2(\mathbf{J}) = -\omega^2 \mu^2 \hat{\mathbf{n}}_r \times \int_{S'} dS' g(\mathbf{r}, \mathbf{r}') \left[\hat{\mathbf{n}}_r' \times \int_{S''} dS'' g(\mathbf{r}', \mathbf{r}'') \mathbf{J} \right] \quad (107)$$

$$\mathcal{T}_s \mathcal{T}_h(\mathbf{J}) = -\hat{\mathbf{n}}_r \times \int_{S'} dS' g(\mathbf{r}, \mathbf{r}') \left[\hat{\mathbf{n}}_r' \times \nabla' \int_{S''} dS'' g(\mathbf{r}', \mathbf{r}'') \nabla'' \cdot \mathbf{J} \right] \quad (108)$$

$$\mathcal{T}_h \mathcal{T}_s(\mathbf{J}) = -\hat{\mathbf{n}}_r \times \nabla \int_{S'} dS' g(\mathbf{r}, \mathbf{r}') \nabla' \cdot \left[\hat{\mathbf{n}}_r' \times \int_{S''} dS'' g(\mathbf{r}', \mathbf{r}'') \mathbf{J} \right] \quad (109)$$

$$\mathcal{T}_h^2(\mathbf{J}) = -\frac{1}{\omega^2 \epsilon^2} \hat{\mathbf{n}}_r \times \nabla \int_{S'} dS' g(\mathbf{r}, \mathbf{r}') \nabla' \cdot \left[\hat{\mathbf{n}}_r' \times \nabla' \int_{S''} dS'' g(\mathbf{r}', \mathbf{r}'') \nabla'' \cdot \mathbf{J} \right]. \quad (110)$$

Recalling the surface Helmholtz decomposition of the surface current \mathbf{J} (Hsiao and Kleinman 1997),

$$\mathbf{J} = \nabla_S \phi + \hat{\mathbf{n}}_r \times \nabla_S \psi \quad (111)$$

where ϕ and ψ are scalar functions defined on S . The first term of Eq. 111 is purely irrotational (curl-free) while the second term is purely solenoidal (divergence-free). It is easy to show that (Adams 2004)

$$\nabla_S \cdot (\mathcal{T}_h(\mathbf{J})) = \nabla_S \cdot (\hat{\mathbf{n}}_r \times \nabla \Phi) = 0. \quad (112)$$

where $\nabla_S \cdot$ is the surface divergence operator. As a result, the square of the hypersingular term (\mathcal{T}_h^2) in Eq. 110 is identically zero. Consequently, the decomposed operator \mathcal{T}^2 in Eq. 106 can be rewritten as

$$\mathcal{T}^2 = \mathcal{T}_s^2 + \mathcal{T}_h \mathcal{T}_s + \mathcal{T}_s \mathcal{T}_h. \quad (113)$$

Notice that from Eqs. 107 to 109, the $\mathcal{T}_s^2 \sim O(\omega^2)$, while the $\mathcal{T}_s \mathcal{T}_h + \mathcal{T}_h \mathcal{T}_s \sim O(\omega^0)$, when $\omega \rightarrow 0$. In other words, the $\mathcal{T}_s \mathcal{T}_h$ and $\mathcal{T}_h \mathcal{T}_s$ are frequency invariant, so that these two terms behave like the identity operator. Meanwhile, \mathcal{T}_s^2 behaves like a compact operator and approaches zero when $\omega \rightarrow 0$. Therefore, the total operator \mathcal{T}^2 can be considered as an identity operator plus a compact operator, making it a well-conditioned second-kind Fredholm integral operator immune to low-frequency breakdown. Meanwhile, $\mathcal{T}_h^2 \sim O(\omega^{-2})$ implies that the square of the hypersingular term has to be set to zero; otherwise it will swamp the contributions from the other three terms at low frequencies. However, if \mathbf{J} is of the order

smaller than ω^0 and the higher-order current is also important for certain problems, the decomposed \mathcal{T}^2 (without \mathcal{T}_h^2) still decreases with frequency. It does not cause the breakdown of the MoM but causes the inaccuracy of the electric current at low frequencies.

Perturbation Method

The low-frequency inaccuracy problem happens in many formulations such as EFIE, MFIE, and CMP-EFIE. As detailed in (Zhang et al. 2003; Qian and Chew 2010), although the formulations are stable at extreme low frequencies, the current unknowns cannot be computed accurately. To remedy this inaccuracy problem, a perturbation method introduced in this section will be discussed based on the expanded CMP-EFIE formulation in Eq. 113 (Sun et al. 2013b).

By expanding the Green's function with Taylor series as

$$g(\mathbf{r}, \mathbf{r}') \approx \frac{1}{4\pi R} \left[1 + ik_0 R + \frac{1}{2} (ik_0 R)^2 \right] \quad (114)$$

the sub-matrices can be expanded with respect to a small parameter $\delta = ik_0$. Then, one has

$$\left[\bar{\mathbf{Z}}_p^q \right] = \left[\bar{\mathbf{Z}}_p^{q(0)} \right] + \delta \left[\bar{\mathbf{Z}}_p^{q(1)} \right] + \delta^2 \left[\bar{\mathbf{Z}}_p^{q(2)} \right] + O(\delta^3) \quad (115)$$

where the impedance matrices for the smoothing term are given by

$$\left[\bar{\mathbf{Z}}_p^{s(0)} \right]_{mn} = \frac{\mu}{4\pi} \int_{S_m} dS \mathbf{f}_{pm}(\mathbf{r}) \cdot \int_{S_n} dS' \frac{1}{|\mathbf{r} - \mathbf{r}'|} \mathbf{f}_{pn}(\mathbf{r}') \quad (116)$$

$$\left[\bar{\mathbf{Z}}_p^{s(1)} \right]_{mn} = \frac{\mu}{4\pi} \int_{S_m} dS \mathbf{f}_{pm}(\mathbf{r}) \cdot \int_{S_n} dS' \mathbf{f}_{pn}(\mathbf{r}') \quad (117)$$

$$\left[\bar{\mathbf{Z}}_p^{s(2)} \right]_{mn} = \frac{\mu}{8\pi} \int_{S_m} dS \mathbf{f}_{pm}(\mathbf{r}) \cdot \int_{S_n} dS' |\mathbf{r} - \mathbf{r}'| \mathbf{f}_{pn}(\mathbf{r}') \quad (118)$$

and the impedance matrices for the hypersingular term are given by

$$\left[\bar{\mathbf{Z}}_p^{h(0)} \right]_{mn} = \frac{1}{4\pi\epsilon} \nabla \int_{S_m} dS \mathbf{f}_{pm}(\mathbf{r}) \cdot \int_{S_n} dS' \frac{1}{|\mathbf{r} - \mathbf{r}'|} \nabla' \mathbf{f}_{pn}(\mathbf{r}') \quad (119)$$

$$\left[\bar{\mathbf{Z}}_p^{h(1)} \right]_{mn} = \frac{1}{4\pi\epsilon} \nabla \int_{S_m} dS \mathbf{f}_{pm}(\mathbf{r}) \cdot \int_{S_n} dS' \nabla' \cdot \mathbf{f}_{pn}(\mathbf{r}') \quad (120)$$

$$\left[\bar{\mathbf{Z}}_p^{h(2)} \right]_{mn} = \frac{1}{8\pi\epsilon} \nabla \int_{S_m} dS \mathbf{f}_{pm}(\mathbf{r}) \cdot \int_{S_n} dS' |\mathbf{r} - \mathbf{r}'| \nabla' \cdot \mathbf{f}_{pn}(\mathbf{r}'). \quad (121)$$

At low frequencies, $\bar{\mathbf{Z}}_p^q \approx \bar{\mathbf{Z}}_p^{q(0)}$, because the zeroth-order of the Green's function in Eq. 114 is equal to its original form of Eq. 90.

For the current and excitation vectors, the same notations as those in Qian and Chew (2010) are used as

$$ik_0\mathbf{j} = \mathbf{j}^{(0)} + \delta\mathbf{j}^{(1)} + \delta^2\mathbf{j}^{(2)} + O(\delta^3) \quad (122)$$

$$\mathbf{b} = \mathbf{b}^{(0)} + \delta\mathbf{b}^{(1)} + \delta^2\mathbf{b}^{(2)} + O(\delta^3) \quad (123)$$

where

$$\left[\mathbf{b}^{(0)}\right]_m = -\eta^{-1} \int_{S_m} dS \mathbf{f}_{RWGm}(\mathbf{r}) \cdot \mathbf{E}_0 \quad (124)$$

$$\left[\mathbf{b}^{(1)}\right]_m = -\eta^{-1} \int_{S_m} dS \mathbf{f}_{RWGm}(\mathbf{r}) \cdot \mathbf{E}_0 \left(\hat{\mathbf{k}}_{inc} \cdot \tilde{\mathbf{r}}\right) \quad (125)$$

$$\left[\mathbf{b}^{(2)}\right]_m = -\frac{1}{2}\eta^{-1} \int_{S_m} dS \mathbf{f}_{RWGm}(\mathbf{r}) \cdot \mathbf{E}_0 \left(\hat{\mathbf{k}}_{inc} \cdot \tilde{\mathbf{r}}\right)^2 \quad (126)$$

and $\hat{\mathbf{k}}_{inc}$ is the unit vector of the incident direction and the tilde above \mathbf{r} indicates the normalization by l , where l is a typical length scale.

Substituting them into Eq. 113 and matching the coefficients of like powers of δ , a recurrent system of equations for the current functions \mathbf{j} is obtained. Firstly, matching the zeroth order of δ gives the lowest-order equation as

$$\left[\bar{\mathbf{Z}}_{CWBC}^{s(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{h(0)} + \bar{\mathbf{Z}}_{CWBC}^{h(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{s(0)}\right] \cdot \mathbf{j}^{(0)} = \bar{\mathbf{Z}}_{CWBC}^{h(0)} \bar{\mathbf{G}}_m^{-1} \mathbf{b}^{(0)} \quad (127)$$

It can be noticed that the impedance matrix is equal to the three-term CMP-EFIE without \mathcal{T}_h^2 in Eq. 113 in static regime, where the contributions from the \mathcal{T}_h^2 of $O(\omega^2)$ disappear. That means the resultant matrix is similar to the form of an identity operator as the aforementioned analysis. Hence, it has a good spectral property at low frequencies and its iterative solution converges well. For the inductive circuit problem like an inductor, the loop current is inversely proportional to frequency, so that the leading term of current is ω^{-1} . According to Eq. 122, the inductive current $ik_0\mathbf{j}$ would be of this zeroth-order solution.

For the plane-wave scattering problem, the leading term of the current \mathbf{j} is of the order of ω^0 . Hence, $\mathbf{j}^{(0)} = 0$ and one can obtain the first- and second-order equations by matching the first and second order of δ as

$$\left[\bar{\mathbf{Z}}_{CWBC}^{s(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{h(0)} + \bar{\mathbf{Z}}_{CWBC}^{h(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{s(0)}\right] \cdot \mathbf{j}^{(1)} = \bar{\mathbf{Z}}_{CWBC}^{h(1)} \bar{\mathbf{G}}_m^{-1} \mathbf{b}^{(0)} + \bar{\mathbf{Z}}_{CWBC}^{h(0)} \bar{\mathbf{G}}_m^{-1} \mathbf{b}^{(1)} \quad (128)$$

and

$$\begin{aligned} & \left[\bar{\mathbf{Z}}_{CWBC}^{s(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{h(0)} + \bar{\mathbf{Z}}_{CWBC}^{h(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{s(0)}\right] \cdot \mathbf{j}^{(2)} \\ &= \bar{\mathbf{Z}}_{CWBC}^{s(0)} \bar{\mathbf{G}}_m^{-1} \mathbf{b}^{(0)} + \bar{\mathbf{Z}}_{CWBC}^{h(1)} \bar{\mathbf{G}}_m^{-1} \mathbf{b}^{(1)} + \bar{\mathbf{Z}}_{CWBC}^{h(0)} \bar{\mathbf{G}}_m^{-1} \mathbf{b}^{(2)} \\ & \quad - \left[\bar{\mathbf{Z}}_{CWBC}^{s(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{h(1)} + \bar{\mathbf{Z}}_{CWBC}^{h(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{s(1)}\right] \cdot \mathbf{j}^{(1)} \\ & \quad - \left[\bar{\mathbf{Z}}_{CWBC}^{s(1)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{h(0)} + \bar{\mathbf{Z}}_{CWBC}^{h(1)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RWG}^{s(0)}\right] \cdot \mathbf{j}^{(1)} \end{aligned} \quad (129)$$

Once the currents in the first three orders are obtained by solving (127, 128, and 129), the far-field results at very low frequencies can be obtained accurately.

It is important to notice that the impedance matrices on the left-hand side of Eqs. 127, 128, and 129 have both contributions from the smoothing and hypersingular terms, thus avoiding the null-space problem and breakdown at low frequencies. Similar to the augmented EFIE formulations, the decomposed CMP-EFIE operator is balanced at low frequencies after setting $\mathcal{T}_h^2 = 0$.

In particular, it can be observed that the leading term of the current \mathbf{j} is of the order of ω^1 for the capacitive or differential circuit problems. It implies that the leading term of $ik_0\mathbf{j}$ is of the order of ω^2 . Hence, the zeroth-order current $\mathbf{j}^{(0)}$ in Eq. 122, which is of the order of ω^0 , and the first-order current $\mathbf{j}^{(1)}$ should be zero. For the delta-gap voltage source excitation, there is only the zeroth order $\mathbf{b}^{(0)}$ left in the right-hand side of the system. By utilizing this natural characteristic, the solution complexity of the matrix systems in Eqs. 127, 128, and 129 can be reduced significantly as

$$\left[\bar{\mathbf{Z}}_{CWB C}^{s(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RW G}^{h(0)} + \bar{\mathbf{Z}}_{CWB C}^{h(0)} \bar{\mathbf{G}}_m^{-1} \bar{\mathbf{Z}}_{RW G}^{s(0)} \right] \cdot \mathbf{j}^{(2)} = \bar{\mathbf{Z}}_{CWB C}^{s(0)} \bar{\mathbf{G}}_m^{-1} \mathbf{b}^{(0)} \quad (130)$$

where the second term in the left-hand side is usually much smaller than the first term, thus leading to a simpler formulation for the capacitive problems (Liu et al. 2014).

Optical Nanoantennas

A microwave or radiowave antenna is an electrical device which converts electric power into radiowaves, and vice versa. As analogues of microwave and radiowave antennas, optical nanoantennas (see Fig. 12) convert freely propagating optical radiation into localized electromagnetic energy, and vice versa (Novotny and Hulst 2011). They are indispensable building blocks for manipulating and controlling light–matter interaction at subwavelength scales. Nanoantennas have broad applications (see Fig. 13) in photodetection, solar energy, light emission, sensing, microscopy, and (surface-enhanced Raman) spectroscopy (Bharadwaj et al. 2009; Giannini et al. 2011; Biagioni et al. 2012; Krasnok et al. 2013). Currently, the tools of nanoscience and nanotechnology, such as focused ion beam milling (Muhlschlegel et al. 2005), electron-beam lithography (Kinkhabwala et al. 2009), and self-assembly schemes (Kalkbrenner et al. 2005), enable fabrication of nanoantennas down to nanoscales.

Theoretical Model

Computational electromagnetics (Chew et al. 2000; Jin 2014; Taflove and Hagness 2005), which is used for modeling the interaction of electromagnetic fields with physical objects and surrounding environment, plays an important role in characterizing and optimizing the optical design of nanoantennas. A rigorous, fast, and efficient solution to Maxwell’s equations facilitates understanding underlying working principles, reducing experimental costs, and accelerating research and development periods. With the aid of state-of-the-art methods, critical physical parameters in the design of nanoantennas can be illustrated for observation and analyzed for optimization. It is highly desirable to know the strengths and weaknesses of various theoretical methods in modeling nanoantennas. Different from classical antennas, rigorous full-wave simulation of nanoantennas faces the following new challenges: (1) accurate near-field calculation; (2) computing numerical dyadic Green’s functions in inhomogeneous electromagnetic environment; (3) complex multilayered, periodic, random, or composite structures; (4) highly dispersive and lossy materials (not perfect electric conductors); (5) strong evanescent wave couplings and mode hybridizations; and (6) multi-scale, nonlinear, and multiphysics effects.

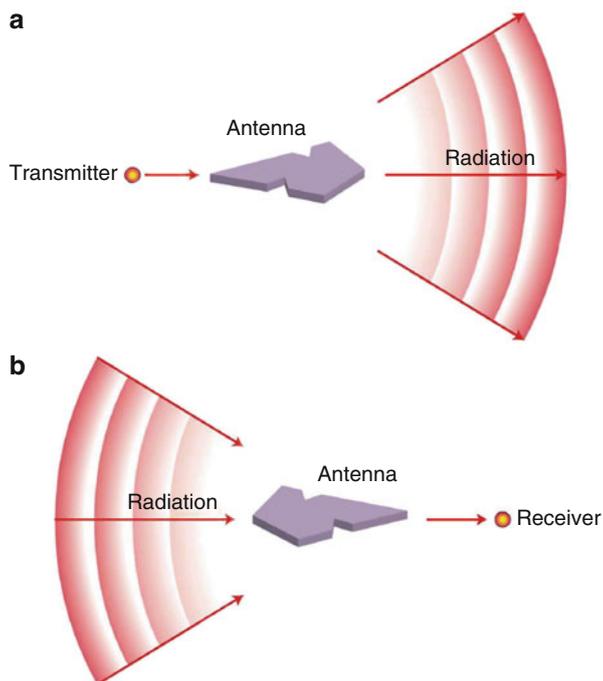


Fig. 12 Schematic patterns of nanoantennas. **(a)** Transmitting nanoantenna. **(b)** Receiving nanoantenna. *Arrows* indicate the direction of energy flow. The two configurations are related by the principle of reciprocity (The figure is taken from Novotny and Hulst (2011))

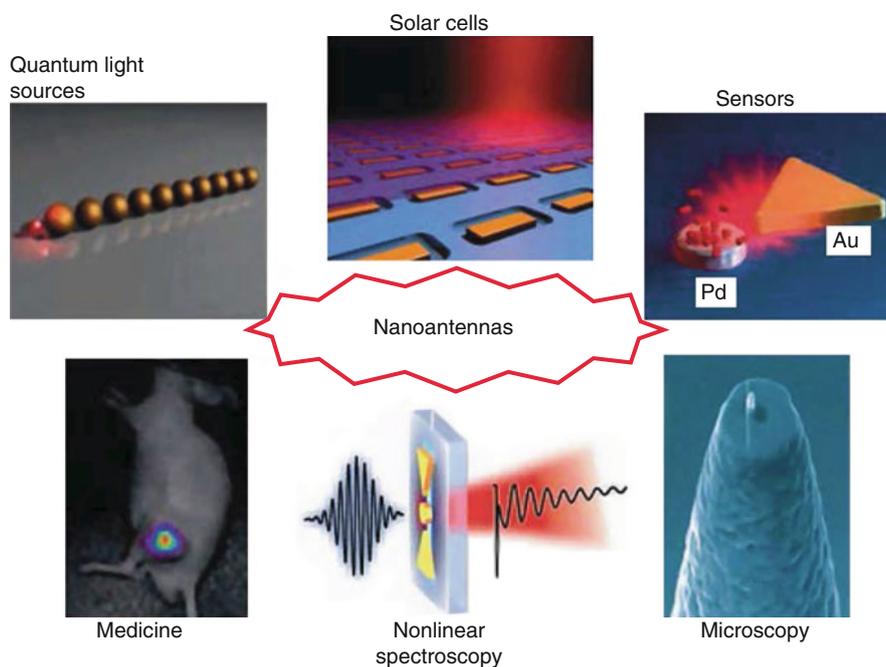


Fig. 13 Modern applications of nanoantennas (The figure is taken from Krasnok et al. (2013))

Time-Domain Methods Versus Frequency-Domain Methods

Most optical materials are dispersive; therefore, a recursive convolution method (Luebbers et al. 1990) or a piecewise linear recursive convolution method (Kelley and Luebbers 1996) must be adopted for time-

domain methods. For noble metals with plasmonic effects in the visible light range, such as silver and gold, the complex refractive index has to be described by a large number of summation terms in the Lorentz–Drude model leading to a long calculation time. However, for frequency-domain methods, an experimentally tabulated refractive index of the dispersive material is incorporated directly. Another difficulty in time-domain methods is the treatment of periodic boundary conditions particularly for the oblique incidence caused by the anticausal property of the Floquet theorem. Hence, the ability of frequency-domain methods to handle the case of oblique incidence is clearly an important advantage over time-domain methods (Veysoglu et al. 1993). Moreover, time-domain methods suffer from numerical dispersion and stability problems in contrast to frequency-domain methods. This drawback becomes serious if a three-dimensional large-scale nanoantenna structure is investigated. A significant merit associated with a time-domain method is a broadband simulation at the frequencies of interest. The frequency-domain method can employ parallel computing technique to circumvent the problem. Compared to frequency-domain methods, time-domain methods cannot obtain accurate eigenvalue solutions but save considerable computer resources. Additionally, time-domain methods are easier and more convenient to model nonlinear and multiphysics effects in nanoantennas.

Integral Equation Methods Versus Differential Equation Methods

Differential equation methods involving finite-difference and finite-element algorithms (Jin 2014; Taflove and Hagness 2005) can treat a variety of inhomogeneous boundary conditions conveniently. The methods have a powerful ability to model a complex nanoantenna structure. The resultant matrix by differential equation methods is sparse due to the “local” nature of differential operators of Maxwell’s equations or wave equations. The method consumes the memory cost of $O(N)$ and complexity of $O(N)$ per matrix-vector multiplication in the Krylov subspace iteration algorithm (Vandervorst 1992). Moreover, multifrontal or multigrid methods (Davis and Duff 1997) can speed up the solution process of the differential equations. To simulate the interaction between light and nanoantennas, an efficient absorption boundary condition (Berenger 1994; Chew and Weedon 1994) and additional volume grids enclosing the nanoantenna have to be adopted.

In comparison with differential equation methods, integral equation methods (Chew 1999, 2008) connect field components to equivalent currents by using “global” integral operators represented with dyadic Green’s functions. As a result, integral equation methods always guarantee higher accuracy but lead to full dense matrix. Fortunately, matrix-free fast algorithms (Chew et al. 2000), such as fast Fourier transform (Catedra et al. 1989) and fast multipole methods (Greengard and Rokhlin 1987; Song et al. 1997), can significantly reduce computer resources required by solving the dense matrix. Thanks to the Green’s tensor, the integral equation methods automatically satisfy the radiation boundary condition but need singularity treatments. In particular, the surface integral equation method having a unique feature of surface triangulation produces much smaller unknowns. However, the method can only be employed to analyze a homogeneous or piecewise-homogeneous structure. For an arbitrary inhomogeneity or complex environment encountered in nanoantennas, the near-field calculation by the surface integral equation method is hard to implement.

Mode-Matching Methods

The mode-matching method (Chew 1999; Bienstman and Baets 2001) is a commonly used technique for the formulation of optical problems, especially for structures consisting of two or more separated regions. It is based on expanding the fields and matching them at the boundaries of different regions and thus lends itself naturally to the analysis of some nanoantenna structures. The most representative of mode-matching methods involves rigorous coupled-wave analysis (Moharam et al. 1995), scattering matrix method (Yonekura et al. 1999), plane-wave expansion method (Johnson and Joannopoulos 2001), and T-matrix

method (Mishchenko et al. 2010). Using cheap computer resources, these methods are particularly useful in characterizing the optical response of periodic or multilayered nanoantennas, Yagi-Uda nanoantennas, and nanoantenna array. Absorption and radiation (scattering) characteristics of nanoantennas can be obtained with a program. However, mode-matching methods are not well suited for characterizing plasmonic effects because a large quantity of modes are required to describe the plasmon coupling and hybridization.

Optical Nanoantennas Versus Classical Antennas

This section introduces the basic principles and physical parameters that govern the operation of nanoantennas. The differences between classical antennas and optical nanoantennas are summarized.

Source and Receiver

Classical antennas use current (voltage) and electrical load as a source and receiver, respectively. However, the source and receiver of nanoantennas become quantum objects such as molecules, atoms, ions, or quantum dots. At this length scale, the light–matter interaction becomes quantized (Novotny and Hulst 2011).

Resonant Transducer

The design of classical antennas relates to the wavelength λ of incident or radiative electromagnetic waves (Novotny and Hulst 2011). For example, a half-wave antenna has a length L of $\lambda/2$, and a Yagi-Uda antenna has fixed separations between elements at fractions of λ . At microwave and radiowave regimes, metals can be regarded as perfect electric conductors (PECs). Thus, the scaling law was adopted to design antennas from one frequency to another. However, the scaling breaks down at optical frequencies, where metals (gold, silver, etc.) are highly lossy and dispersive materials and the skin depth of metals becomes comparable to the nanoantenna size. The permittivity of metals at optical regime can be expressed by a superposition of the Lorentz–Drude model (Maier 2007). Although an effective wavelength (Novotny and Hulst 2011) or a low-frequency circuit theory has been utilized (Alu and Engheta 2008) to simplify the nanoantenna design, a rigorous solution to Maxwell’s equations is essential to capture the wave interaction between nanoantennas and sources/receivers. Particularly, the size of a resonant nanoantenna is much smaller than the wavelength of radiation, which breaks the half-wavelength limit.

Absorption, Scattering, and Extinction Cross Sections

The far-field radiation pattern adopted in classical antennas is still very useful to describe the angular response of nanoantenna radiation or scattering. However, the total absorption and scattering properties of a nanoantenna are easier to measure and calculate for optimized designs.

Given a plane-wave excitation (laser excitation can be well approximated as a plane-wave excitation in nanoantenna experiments), the scattering cross section, which estimates the total scattering strength of a nanoantenna as a function of frequency, is defined as (Tsang et al. 2000)

$$\sigma_s = \frac{\int_S \Re e \left[\frac{1}{2} \mathbf{E}_{sca} \times \mathbf{H}_{sca}^* \right] \cdot d\mathbf{S}}{|\mathbf{S}_{inc}|} \quad (131)$$

where $*$ denotes the complex conjugation, S is an arbitrary surface enclosing the nanoantenna, $\mathbf{S}_{inc} = \frac{1}{2} [\mathbf{E}_{inc} \times (\mathbf{H}_{inc}^*)]$ is the incident energy flux, and \mathbf{E}_{sca} and \mathbf{H}_{sca} are the scattered electric and magnetic fields, respectively. Considering the total power absorbed by a nanoantenna, absorption cross section is of the form (Tsang et al. 2000)

$$\sigma_a = -\frac{\int_S \frac{1}{2} \Re[\mathbf{E} \times \mathbf{H}^*] \cdot d\mathbf{S}}{|\mathbf{S}_{inc}|} = \frac{\int_V k_0 \epsilon_r''(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2 dV}{|\mathbf{E}_{inc}|^2} \quad (132)$$

where ϵ_r'' is the imaginary part of relative permittivity and k_0 is the wave number of free space (background medium). The extinction cross section describes the intrinsic losses of a nanoantenna system including both absorption loss and scattering (or leaky) loss. Hence, the sum of the scattering and absorption cross sections is the extinction cross-section σ_e

$$\sigma_e = \sigma_a + \sigma_s \quad (133)$$

Moreover, another mathematical expression for the extinction cross section is to employ the optical theorem at the far-field limit (Tsang et al. 2000)

$$\sigma_e = \frac{4\pi}{k_0} \Im m[\mathbf{e}_{inc} \cdot \bar{\mathbf{F}} \cdot \mathbf{e}_{inc}] \quad (134)$$

and

$$\mathbf{E}_{sca} = \frac{\exp(ik_0 r)}{r} \bar{\mathbf{F}} \cdot \mathbf{E}_{inc}, \quad r \rightarrow \infty \quad (135)$$

where i is the imaginary unit and \mathbf{e}_{inc} is the polarization unit vector of the incident electric field \mathbf{E}_{inc} . The absorption, scattering, and extinction cross sections can be obtained by a postprocessing procedure after rigorously solving Maxwell's equations (He et al. 2012).

Directivity, Gain, and Reciprocity

Nanoantenna gain is a product of radiation efficiency η_r and directivity D , i.e., $G = \eta_r D$, which is the same as classical antennas. The radiation efficiency is defined as the ratio of the far-field power P_r radiated by the emitter with a nanoantenna to the total delivered power, i.e., $\eta_r = P_r / (P_r + P_l)$, where P_l is the total ohmic loss by nanoantenna absorption. This is also identical to classical antennas. However, considering internal losses of the emitter itself, an internal efficiency is defined as (Krasnok et al. 2013)

$$\eta_{in} = \frac{P_0}{P_0 + P_{0l}} \quad (136)$$

where P_{0l} and P_0 are the internal losses and the power radiated by the emitter in the absence of the nanoantenna, respectively. Finally, the radiation efficiency is expressed as

$$\eta_r = \frac{P_r}{P_r + P_l + P_0(1 - \eta_{in})/\eta_{in}} \quad (137)$$

From the above equations, the introduction of nanoantennas could significantly improve the radiation efficiency of a poor emitter ($\eta_{in} \ll 1$). Because of a resonant enhancement, the radiated power P_r with the nanoantenna is much larger than the radiated power P_0 without the nanoantenna. Similar to classical

antennas, reciprocity holds true for nanoantennas if the medium is the reciprocal medium ($\bar{\epsilon}_r = \bar{\epsilon}_r^t, \bar{\mu}_r = \bar{\mu}_r^t$). Considering two nanoantenna systems with different sources \mathbf{J}_1 and \mathbf{J}_2 , which respectively produce the fields \mathbf{E}_1 and \mathbf{E}_2 , then $\int_V \mathbf{E}_1 \cdot \mathbf{J}_2 dV = \int_V \mathbf{E}_2 \cdot \mathbf{J}_1 dV$.

Input Impedance Versus Local Density of States

For classical antennas, guided structures (transmission line, coplanar waveguide, microstrip line, etc.) are utilized to control the amplitude and phase of feeding source. The input impedance of antennas should be matched with the characteristic impedance of guided structures to minimize the reflection. However, the guided structure is not needed for nanoantennas if the source is a quantum emitter or their clusters, which are located around the feeding gap (element) of nanoantennas. In this situation, the impedance matching concept is not applicable to the nanoantenna design. Typically, a focused laser beam is used to excite or feed nanoantennas. The excitation field to quantum emitters is enhanced by the nanoantenna. Meanwhile, radiation field from quantum emitters is again enhanced by the nanoantenna. For surface-enhanced Raman scattering (SERS), the enhancement factor is written as (Giannini et al. 2011)

$$G_{SERS} = \frac{|\mathbf{E}^p(\mathbf{r}_0, \omega_i)|^2}{|\mathbf{E}_{inc}^p(\mathbf{r}_0, \omega_i)|^2} \cdot \frac{|\mathbf{E}^d(\mathbf{r}_0, \omega_r)|^2}{|\mathbf{E}_{inc}^d(\mathbf{r}_0, \omega_r)|^2} \quad (138)$$

where \mathbf{r}_0 is the emitter point and $|\mathbf{E}|^2/|\mathbf{E}_{inc}|^2$ are due to local field enhancements. ω_i and ω_r are the angular frequencies respectively for the incident (pump) laser beam and Raman-shifted radiation field. The enhancement factor can be easily obtained by a near-field simulation of nanoantennas. It should be noted that $\mathbf{E}_{inc}^p(\mathbf{r}_0, \omega_i)$ is the incident E-field from the laser beam (that can be regarded as a plane wave) while $\mathbf{E}_{inc}^d(\mathbf{r}_0, \omega_r)$ is the incident E-field from the emitter (that can be regarded as a Hertzian dipole).

To let quantum emitters efficiently radiate electromagnetic waves, photon local density of states (LDOS) at the emitter point or spontaneous emission (decay) rate of the emitter should be enhanced. The spontaneous emission rate counts the number of spontaneously emitted photons per unit time. The LDOS counts the number of electromagnetic modes at the emitter point. Each electromagnetic mode can be taken as a decay channel. The more decay channels there are, the easier it is for an excited atom to emit photons via returning to its ground state (Novotny and Hulst 2011). Therefore, more decay channels mean a larger spontaneous emission rate. Nanoantennas could significantly boost spontaneous emissions (named Purcell effect) due to the highly localized near-field enhancement induced by plasmonic effects. For example, the gap plasmonic mode supported at metallic bow-tie antennas enhances the single-molecule fluorescence (Kinkhabwala et al. 2009).

The LDOS in reciprocal, lossless, inhomogeneous, and anisotropic medium can be connected to dyadic Green's functions or Green's tensor

$$\rho(\mathbf{r}_0, \omega_0) = \frac{k_0^2}{\pi\omega_0} \text{Tr} \left\{ \bar{\epsilon}_r^{1/2} \cdot \Im m [\bar{\mathbf{G}}^e(\mathbf{r}_0, \mathbf{r}_0; \omega_0)] \cdot \bar{\epsilon}_r^{1/2} + \bar{\mu}_r^{1/2} \cdot \Im m [\bar{\mathbf{G}}^m(\mathbf{r}_0, \mathbf{r}_0; \omega_0)] \cdot \bar{\mu}_r^{1/2} \right\} \quad (139)$$

where $\bar{\mathbf{G}}^e$ and $\bar{\mathbf{G}}^m$ are the electric and magnetic dyadic Green's functions in inhomogeneous medium. $\bar{\epsilon}_r$ and $\bar{\mu}_r$ are the relative permittivity and relative permeability tensors. For lossless, isotropic, inhomogeneous, and nonmagnetic medium, the above can be simplified to (Novotny and Hecht 2006)

$$\begin{aligned}\rho(\mathbf{r}_0, \omega_0) &= \sum_k |\mathbf{u}_k(\mathbf{r}_0)|^2 \delta(\omega_k - \omega_0) \\ &= \frac{2\omega_0}{\pi c^2} \text{Tr}\{\Im[\overline{\mathbf{G}}^e(\mathbf{r}_0, \mathbf{r}_0; \omega_0)]\}\end{aligned}\quad (140)$$

where $c = 1/\sqrt{\mu\epsilon}$ is the speed of light in the medium. Here, \mathbf{u}_k are the eigenmodes of E-field in inhomogeneous medium satisfying an orthogonal relation $\int_v \mathbf{u}_k^\dagger(\mathbf{r})\epsilon_r(\mathbf{r})\mathbf{u}_k(\mathbf{r})d\mathbf{r} = \delta_{k',k}$. The LDOS can be separated into radiative and non-radiative parts, which are connected to radiation efficiency in Eq. 137. According to Fermi's golden rule (Novotny and Hecht 2006), the spontaneous emission rate γ_{se} is proportional to the LDOS

$$\gamma_{se} = \frac{\pi\omega_0}{3\hbar\epsilon_0} |\mathbf{p}|^2 \rho(\mathbf{r}_0, \omega_0) \quad (141)$$

where \mathbf{p} is the electric dipole moment of the atomic (molecular) transition.

Computational electromagnetics plays an important role in obtaining the Green's tensor in arbitrary inhomogeneous environment. The Green's tensor in inhomogeneous medium can be found by solving the following vector wave equations:

$$\nabla \times \overline{\mu}^{-1} \cdot \nabla \times \overline{\mathbf{G}}^e - \omega_0^2 \overline{\epsilon} \cdot \overline{\mathbf{G}}^e = \overline{\mathbf{I}}\delta(\mathbf{r} - \mathbf{r}') \quad (142)$$

$$\nabla \times \overline{\epsilon}^{-1} \cdot \nabla \times \overline{\mathbf{G}}^m - \omega_0^2 \overline{\mu} \cdot \overline{\mathbf{G}}^m = \nabla \times [\overline{\epsilon}^{-1} \cdot \overline{\mathbf{I}}\delta(\mathbf{r} - \mathbf{r}')] \quad (143)$$

A finite-difference method has been developed to discretize the delta function (monopole source) and derivative of delta function (dipole source) (Qiao et al. 2011). Taking a 2D line source as an example, one gets

$$\delta(x - m'\Delta_x, y - n'\Delta_y)|_{x=m'\Delta_x, y=n'\Delta_y} \approx \frac{1}{\Delta_x\Delta_y} \quad (144)$$

$$\begin{aligned}\frac{\partial\delta(x - m'\Delta_x, y - n'\Delta_y)}{\partial y} &\approx -\frac{\delta(x - m'\Delta_x, y - (n' + 0.5)\Delta_y)}{\Delta_y} \\ &\quad + \frac{\delta(x - m'\Delta_x, y - (n' - 0.5)\Delta_y)}{\Delta_y}\end{aligned}\quad (145)$$

where Δ_x and Δ_y are the spatial steps along the x and y directions and (m', n') is the grid index corresponding to the line source point. Using the above techniques and the finite-difference implementation, the numerical dyadic Green's functions can be found accordingly. Recently, using multilayered Green's functions and integral equation approach, the numerical Green's tensor around an arbitrarily shaped metallic nanostructure deposited on a multilayered structure has been calculated (Chen et al. 2012b).

Weak Coupling and Strong Coupling Regimes

Consider a nanoantenna operating at a single resonant mode \mathbf{u}_k , such as a metallic dipole antenna. In this situation, spontaneous emission rate in Eq. 141 can be rewritten as (Andreani et al. 1999)

$$\gamma_{se} = \frac{2}{3\hbar\epsilon_0} |\mathbf{p}|^2 \frac{Q}{V} \quad (146)$$

where Q is the quality factor of the resonant antenna and $V = 1/|\mathbf{u}_{\mathbf{k}}(\mathbf{r}_0)|^2$ is the mode volume. According to Jaynes–Cummings model, strong and weak coupling regimes can be distinguished by using the atom-field coupling constant (Andreani et al. 1999)

$$\kappa = \sqrt{\frac{|\mathbf{p}|^2 \omega}{2\hbar\epsilon_0 V}} \quad (147)$$

Weak coupling obeys the condition $\kappa \ll \gamma_{se}$. In this irreversible regime, the effect of vacuum fluctuation fields on the emitter can be treated as a perturbation. Due to large radiation and absorption losses, spontaneously emitted photons cannot react back upon the emitter. Strong coupling obeys the condition $\kappa \gg \gamma_{se}$. In this reversible regime, photons remain in the cavity for a long time and are absorbed and re-emitted many times, which is called vacuum Rabi oscillation. Strong coupling allows the electron and photon entangled to be a polariton, which is a quasiparticle. The strong coupling regime is quite important to cavity quantum electrodynamics and quantum computing. In Eqs. 146 and 147, the important characteristic parameters Q and V can be rigorously calculated by an eigenvalue analysis. Calculation of the mode volume V is trivial if the eigenmode $\mathbf{u}_{\mathbf{k}}$ around the (resonant) working frequency ω of the nanoantenna has been obtained. The quality factor can be extracted from corresponding complex eigenfrequency via the equation $Q = \Re(\omega_k)/[2\Im(\omega_k)]$. Several fast and universal eigenvalue solvers have been developed for an arbitrary electromagnetic system (Dai et al. 2012, 2013, 2014; Sha et al. 2014).

Design Rules for Optical Nanoantennas

Broadband Nanoantennas

A broadband nanoantenna can be an absorber for light harvesting or be a (near-field) concentrator for light enhancement (Atwater and Polman 2010). It also can be a far-field reflector for light trapping (Yu et al. 2010). Tapered nonresonant structures (Maksymov et al. 2011) support propagating (not standing) waves and thus a broadband response. Alternatively, the multisized nanoantenna (Cui et al. 2011) is another way to obtain a broadband absorption or scattering. The mode coupling (hybridization) between individual resonators induces a spectral overlap between eigenmodes (Ye and He 2010). Meanwhile, high-order modes could be excited via a strong mode coupling or symmetry breaking (Giannini et al. 2011). According to the coupled-mode theory, a single-mode resonator will be operated at the critical coupling regime if the radiation loss is balanced with the absorption loss (Yu et al. 2010), which is similar to a maximum power transfer condition in electrical engineering. Under this condition, a maximum absorption can be achieved but with a narrow bandwidth. Additionally, a broadband reflector can be realized by resonators operated at the over-coupling regime (Yu et al. 2010), where the radiation loss is much larger than the absorption (ohmic) loss.

Wavelength-Selective Nanoantennas

Wavelength selectivity by nanoantennas is required to resolve and match vibrational modes of a target molecule. Radiation and ohmic losses can be reduced to construct a narrow-band resonator with a high-quality factor. A defect in a periodic structure can induce the wavelength selectivity as well (Joannopoulos et al. 2008). The wavelength selectivity has been realized by a single dielectric sphere, where the

constructive and destructive interferences between the broad dipole eigenmode (first term of the spherical harmonics in Mie series) and the narrow quadrupole eigenmode (second term of the spherical harmonics in Mie series) induce an extraordinary Fano resonance with an asymmetric and narrow spectral line (Liu et al. 2012). The Fano spectral line is fundamentally different from the Lorentz spectral line achieved by a single-mode resonator.

Directional Nanoantennas

Directional nanoantennas are useful for far-field detection and sensing. A high directivity can be obtained by the following rules: (1) far-field interference by using an antenna array (Liu et al. 2011); (2) near-field interference by using the reflector, such as Yagi-Uda antennas (Taminiau et al. 2008); (3) using optical couplers, such as gratings, Fabry–Perot directive antenna, etc. (Martin-Moreno et al. 2003); and (4) using high-order modes and mode interference, such as a single sphere antenna, whispering gallery antenna, etc. (Gerard et al. 2009; Liu et al. 2012; Staude et al. 2013).

Metallic Nanoantennas Versus Dielectric Nanoantennas

Metallic nanoantennas have an unprecedented ability to concentrate and confine electromagnetic fields at a deep subwavelength scale. Dielectric nanoantennas (Pellegrini et al. 2009) have a larger geometric size but with a lower ohmic loss. Dielectric nanoantennas could easily excite high-order eigenmodes or exhibit both electric and magnetic resonances (Gerard et al. 2009; Staude et al. 2013) and thus show a strong scattering capability. Also, they can be semiconducting materials themselves and are more compatible to optoelectronic devices (Chen et al. 2012a). Dielectric nanoantennas can be integrated with metallic ones to achieve better performances (Devilez et al. 2010).

Nonlinear Optical Nanoantennas

Metals at optical regime have nonlinear susceptibility. Strong electromagnetic fields in plasmonic nanoantennas significantly enhance nonlinear processes including second harmonic generation, Kerr effect, four-wave mixing, etc. (Kauranen and Zayats 2012). Far-field detection of high-order harmonic generation depends on (1) pump field intensity at the frequency of fundamental mode, (2) spatial overlap between fundamental eigenmode and high-order harmonic eigenmode, and (3) outcoupling (far-field radiation) strength of high-order harmonic eigenmode. There are many designs in literatures to enhance the nonlinear effects of nanoantennas (Kauranen and Zayats 2012). The nonlinear response of nanoantennas can be rigorously modeled with the self-consistent solution to coupled Maxwell's equations and hydrodynamic equation (Corvi and Schaich 1986). Using a surface susceptibility tensor, a surface integral equation approach has been proposed to model the second harmonic generation from symmetric metallic objects, where the surface contribution is much larger than the bulk one (Makitalo et al. 2011; Xiong et al. 2014). It is worthy of mentioning that symmetry breaking is an efficient way to enhance the second harmonic generation from plasmonic nanoantennas (Pu et al. 2010).

Control of Light Angular Momentum by Nanoantennas

Similar to an electron, a photon has spin and orbital angular momenta, which are respectively defined as (Bliokh et al. 2013)

$$\mathbf{P}_s = \frac{1}{8\omega} \nabla \times \mathfrak{J}m \left[\frac{\epsilon_0}{\mu_r} (\mathbf{E}^* \times \mathbf{E}) + \frac{\mu_0}{\epsilon_r} (\mathbf{H}^* \times \mathbf{H}) \right] \quad (148)$$

$$\mathbf{P}_o = \frac{1}{4\omega} \Im m \left[\frac{\epsilon_0}{\mu_r} (\mathbf{E}^* \cdot \nabla \mathbf{E}) + \frac{\mu_0}{\epsilon_r} (\mathbf{H}^* \cdot \nabla \mathbf{H}) \right] \quad (149)$$

The summation of the spin and orbital angular momenta is equal to the electromagnetic momentum density

$$\mathbf{P} = \mathbf{P}_s + \mathbf{P}_o = \frac{\text{Re}(\mathbf{E}^* \times \mathbf{H})}{2c^2} \quad (150)$$

Moreover, the time-averaged helicity density, which characterizes the difference between the number of right-hand and left-hand circularly polarized photons, is defined as (Bliokh et al. 2013)

$$h = \frac{-1}{2\omega c_0} \Im m (\mathbf{E}^* \cdot \mathbf{H}) \quad (151)$$

From Eqs. 148 and 151, linear polarized light has zero spin angular momentum and zero helicity. The left-handed and right-handed circular polarizations have different spin angular momentum and helicity. From Eq. 149, the orbital angular momentum of a light beam depends on the field spatial distribution (not on the polarization). It can be split into an internal and an external orbital angular momenta. The internal orbital angular momentum is an origin-independent angular momentum of a light beam that can be associated with a helical or twisted wavefront. The external orbital angular momentum is the origin-dependent (center of the beam-dependent) angular momentum. The total angular momentum is conserved if light goes through a rotationally symmetric nanostructure. To rotate circularly polarized light, chiral metamaterials as building blocks have been constructed via mirror symmetry breaking (Kwon et al. 2008; Ye and He 2010; Ye et al. 2010). Consequently, if a rotationally symmetric nanoantenna constructed by chiral metamaterials converts the right-hand circularly polarized light to the left-hand circularly polarized light, both spin and orbital angular momenta of light will be changed based on the total angular momenta conservation. Additionally, all the above physical quantities including spin and orbital angular momenta and helicity density can be computed by classical computational electromagnetics.

Tunable Active Nanoantennas

Optical responses of nanoantennas involving resonance frequency, bandwidth, radiation pattern, near-field distribution, etc., can be dynamically tunable and switchable by using optical, electrical, and mechanical ways. Plasmonic effects could induce a strong optical force to reconfigure the geometry of nanoantennas (Bonakdar et al. 2012) or enhance nonlinear (Kerr) effects to modify refractive indices of antennas or loads (Chen and Alu 2010). Permittivity of liquid crystals (Berthelot et al. 2009) and graphene (Yao et al. 2013; Ren et al. 2013) can be tunable with electrostatic fields. Hence, they have been introduced to electrostatically tune optical responses of nanoantennas. Nanomechanical control of an optical antenna was also reported in literatures (Merlein et al. 2008).

Modal Analysis for Antenna Design

Modal analysis, as the preoccupation of many researchers, has been successfully adopted in the investigation of conducting enclosures such as waveguides and resonators at different frequency bands (Collin 1991; Van Bladel 1985). Recently, characteristic mode analysis gains popularity among antenna engineers

as it provides a systematic approach for antenna design which is no longer experience or intuition based. Initiated by Garbacz and refined by Harrington and Mautz (Garbacz and Turpin 1971; Harrington and Mautz 1971), characteristic mode theory was popularized by the work of Cabedo-Fabres within the last decade (Cabedo-Fabres et al. 2007). On the other hand, the well-documented literature on natural mode analysis mainly focuses on bounded or semi-bounded systems (even dielectric resonator antennas are normally enclosed by impenetrable boundaries when their modal shapes are sought for). Recent effort addresses bounded and unbounded problems using a general framework based on natural mode expansion, which offers useful physical insight into antenna operation, as well as an alternative of CMA for modal design and engineering (Dai et al. 2012, 2014).

Characteristic Mode Analysis

Considering the electric field integral equation (EFIE) for a perfect electric conductor (PEC) object, the characteristic mode analysis (CMA), formulated on top of the method of moments (MoM), aims to obtain the characteristic current modes \mathbf{J}_n supported on the PEC surface S by solving the generalized eigenvalue equation

$$\bar{\mathbf{Z}} \cdot \mathbf{J}_n = v_n \bar{\mathbf{R}} \cdot \mathbf{J}_n \quad (152)$$

or

$$\bar{\mathbf{X}} \cdot \mathbf{J}_n = \lambda_n \bar{\mathbf{R}} \cdot \mathbf{J}_n, \quad (153)$$

where $\bar{\mathbf{Z}} = \bar{\mathbf{R}} + i\bar{\mathbf{X}}$, and $v_n = 1 + i\lambda_n$. Meanwhile, $\bar{\mathbf{Z}}$ is the impedance matrix for a given frequency, whose entries are calculated as

$$\begin{aligned} (\bar{\mathbf{Z}})_{mn} &= i\omega\mu \langle \mathbf{f}_m, \bar{\mathbf{G}}, \mathbf{f}_n \rangle_S \\ &= i\omega\mu \int_S d\mathbf{r} \mathbf{f}_m(\mathbf{r}) \cdot \int_S \bar{\mathbf{G}}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{f}_n(\mathbf{r}') d\mathbf{r}', \end{aligned} \quad (154)$$

where $\bar{\mathbf{G}}(\mathbf{r}, \mathbf{r}')$ is the dyadic Green's function and $\mathbf{f}_n(\mathbf{r})$ are RWG basis functions.

Since $\bar{\mathbf{R}}$ and $\bar{\mathbf{X}}$ are dense matrices, the use of fast algorithms for full matrix-vector product is indispensable to solve for λ_n and \mathbf{J}_n iteratively. Otherwise, CMA can only be used to analyze small objects. The characteristic value (eigenvalue) λ_n is important as its magnitude is proportional to the object's reactive power. When $\lambda_n = 0$, the correspondent characteristic mode \mathbf{J}_n is at resonance which is efficient in receiving or radiating energy. When $\lambda_n > 0$ ($\lambda_n < 0$), mode \mathbf{J}_n is storing electric (magnetic) energy. The characteristic angle can be calculated as $\phi_n = 180^\circ - \arctan(\lambda_n)$. Hence, mode \mathbf{J}_n is at resonance when $\phi_n = 180^\circ$ and stores energy only when $\phi_n = 90^\circ$ or 270° .

When an incident field \mathbf{E}_{inc} exists, the induced current on the PEC object can be calculated by solving

$$\bar{\mathbf{Z}} \cdot \mathbf{I}_{ind} = \mathbf{V} \quad (155)$$

where

$$(\mathbf{V})_n = -\langle \mathbf{f}_n, \mathbf{E}_{inc} \rangle_S. \quad (156)$$

Since the orthogonal property of characteristic modes holds for $\lambda_m \neq \lambda_n$

$$\mathbf{J}_m^T \cdot \bar{\mathbf{R}} \cdot \mathbf{J}_n = 0, \quad (157)$$

one can easily expand \mathbf{I}_{ind} in terms of \mathbf{J}_n as

$$\mathbf{I}_{ind} = \sum_n \alpha_n \mathbf{J}_n \quad (158)$$

where the modal expansion coefficients are

$$\alpha_n = \frac{1}{1 + i\lambda_n} \frac{\mathbf{J}_n^T \cdot \mathbf{V}}{\mathbf{J}_n^T \cdot \bar{\mathbf{R}} \cdot \mathbf{J}_n} \quad (159)$$

Natural Mode Expansion

Natural mode analysis, when formulated in terms of integral equations, differs from CMA in that it is closely related to the problem of seeking for \mathbf{J} such that $\bar{\mathbf{Z}} \cdot \mathbf{J} = 0$. This requires one to find natural resonances of an open system or, equivalently, roots of $\det [\bar{\mathbf{Z}}(\omega_n)] = 0$ on the complex plane. On the other hand, CMA should always yield real λ_n as a function of given frequencies along the real axis, while the correspondent \mathbf{J}_n can never satisfy $\bar{\mathbf{Z}} \cdot \mathbf{J}_n = 0$.

It is quite challenging to find natural modes with integral equation solvers as the complexity of calculating $\det [\bar{\mathbf{Z}}]$ is $O(N^3)$ where N is the number of unknowns. In the singularity expansion-based studies, natural modes are extracted from the measured data rather than being directly calculated (Baum et al. 1991). However, natural mode analysis can be easily carried out if it is formulated in terms of differential equations.

Bounded Case

Consider an arbitrary inhomogeneity (where $\bar{\epsilon}$ and $\bar{\mu}$ are tensors) bounded in domain V by impenetrable boundaries at surface S . The electric field \mathbf{E}_n and magnetic field \mathbf{H}_n of a natural mode are governed by the linear eigenvalue equations

$$\nabla \times \bar{\mu}^{-1} \cdot \nabla \times \mathbf{E}_n = \omega_n^2 \bar{\epsilon} \cdot \mathbf{E}_n \quad (160)$$

and

$$\nabla \times \bar{\epsilon}^{-1} \cdot \nabla \times \mathbf{H}_n = \omega_n^2 \bar{\mu} \cdot \mathbf{H}_n, \quad (161)$$

respectively. On the surface S , the field satisfies $\hat{n} \times \mathbf{E}_n = 0$ or $\hat{n} \times \mathbf{H}_n = 0$ for PEC or PMC boundary, respectively, where \hat{n} is the unit normal vector of S .

Taking the case of \mathbf{E}_n as an example, similar analysis applies to the case of \mathbf{H}_n . When the medium is lossless, namely, $\bar{\epsilon} = \bar{\epsilon}^\dagger$ and $\bar{\mu} = \bar{\mu}^\dagger$, or reciprocal, namely, $\bar{\epsilon} = \bar{\epsilon}^T$ and $\bar{\mu} = \bar{\mu}^T$, one can show that for $\omega_m^2 \neq \omega_n^2$,

$$\langle \mathbf{E}_m^\gamma, \bar{\epsilon} \cdot \mathbf{E}_n \rangle_V = \int_V \mathbf{E}_m^\gamma(\mathbf{r}) \cdot \bar{\epsilon} \cdot \mathbf{E}_n(\mathbf{r}) \, d\mathbf{r} = 0, \quad (162)$$

where $\gamma = *$ for the lossless case and omitted for the reciprocal case. Nontrivial degenerate modes satisfying $\omega_m^2 = \omega_n^2 \neq 0$ can be orthogonalized by applying the Gram–Schmidt process.

The operator $(\nabla \times \bar{\mu}^{-1} \cdot \nabla \times)$ has a null space spanned by countably infinite irrotational eigenmodes \mathbf{E}_q associated with zero eigenvalues. They are orthogonal to nontrivial modes \mathbf{E}_n and are expressible as $\mathbf{E}_q = \nabla \psi_q$, where ψ_q are solutions of eigenvalue equation

$$\nabla \cdot \bar{\epsilon} \cdot \nabla \psi_q(\mathbf{r}) = k_q^2 \psi_q(\mathbf{r}) \quad (163)$$

associated with eigenvalues k_q^2 . Under prescribed boundary conditions (Dirichlet or Neumann) of a bounded V , for $k_p^2 \neq k_q^2$, the orthogonality property $\langle \psi_p^\gamma, \psi_q \rangle_V = 0$ or $\langle \mathbf{E}_p^\gamma, \bar{\epsilon} \cdot \mathbf{E}_q \rangle_V = 0$ holds, where γ is defined for the lossless case and reciprocal case, accordingly.

By using the eigenbasis \mathbf{E}_n including \mathbf{E}_q , the solution of Eq. 8 in V can be expanded as

$$\mathbf{E}(\mathbf{r}) = \sum_n^\infty \alpha_n \mathbf{E}_n(\mathbf{r}) \quad (164)$$

where modal expansion coefficients α_n can be obtained using the modal orthogonality as

$$\alpha_n = \frac{1}{\omega_n^2 - \omega^2} \frac{\langle \mathbf{E}_n^\gamma, i\omega \mathbf{J} \rangle_V}{\langle \mathbf{E}_n^\gamma, \bar{\epsilon} \cdot \mathbf{E}_n \rangle_V} \quad (165)$$

In low-frequency applications, \mathbf{E}_q are crucial in expanding the field. If the source \mathbf{J} is not divergence-free, \mathbf{E}_q may need to be included in the modal expansion solution.

When the inhomogeneity is lossy, namely, $\bar{\epsilon} \neq \bar{\epsilon}^\dagger$ or $\bar{\mu} \neq \bar{\mu}^\dagger$, or nonreciprocal, namely, $\bar{\epsilon} \neq \bar{\epsilon}^T$ or $\bar{\mu} \neq \bar{\mu}^T$, Eq. 162 is no longer valid. An approach for constructing orthogonal eigenmodes is to introduce an auxiliary system to the original one as follows (Chen and Lien 1980):

$$\Delta \times [\bar{\mu}^{-1}]^\nu \cdot \nabla \times \mathbf{E}_n^a = (\omega_n^2)^\gamma \bar{\epsilon}^\nu \cdot \mathbf{E}_n^a \quad (166)$$

where $\nu = \dagger$ and $\nu = T$ for lossy and nonreciprocal cases, respectively, and γ is defined the same as before. Equation 166 should be constructed in the same domain V as in Eq. 160, and the auxiliary fields \mathbf{E}_n^a should satisfy the same boundary conditions as in Eq. 160 but with different material media. Therefore, one can obtain the nondegenerate orthogonality property as

$$\langle \mathbf{E}_m^{a\gamma}, \bar{\epsilon} \cdot \mathbf{E}_n \rangle_V = 0. \quad (167)$$

Similar to the original problem, degenerate curl-free modes of the auxiliary system satisfy $\mathbf{E}_q^a = \nabla \psi_q^a$ where ψ_q^a are eigensolutions of

$$\nabla \cdot \bar{\epsilon} \cdot \nabla \psi_q^a(\mathbf{r}) = (k_q^2)^\gamma \psi_q^a(\mathbf{r}) \quad (168)$$

Again, for $k_p^2 \neq k_q^2$, one has $\langle \mathbf{E}_p^{a\gamma}, \bar{\epsilon} \cdot \mathbf{E}_q \rangle_V = 0$. Therefore, the field excited by any current \mathbf{J} can be expanded in terms of the complete set (assumed to be) of eigenbasis \mathbf{E}_n where α_n are in a simple form of

$$\alpha_n = \frac{1}{\omega_n^2 - \omega^2} \frac{\langle \mathbf{E}_n^{a\gamma}, i\omega \mathbf{J} \rangle_V}{\langle \mathbf{E}_n^{a\gamma}, \bar{\epsilon} \cdot \mathbf{E}_n \rangle_V}. \quad (169)$$

Unbounded Case

A straightforward extension of natural mode expansion (NME) to unbounded problems is to let S tend to infinity. A small loss has to be introduced to the medium in V , which is equivalent to requiring the wave to be outgoing at infinity. The small loss also guarantees the uniqueness of the field solution due to a given source (Chew 1990).

Based on the modal behavior, natural modes can be approximately classified into trapped modes and exterior modes. The former resonate due to the inhomogeneity with most energy confined, while the latter resonate between the inhomogeneity and the far boundaries. Note that the energy confinement for a trapped mode in 3D problems is never perfect, unless the inhomogeneity is a perfectly enclosed cavity or a dielectric with an infinitely high refractive index. In certain geometries, some trapped modes are coupled to external radiation, which are also regarded as leaky modes.

As S approaches infinity, the spectra of trapped modes slightly broaden while the modal shapes remain almost unchanged. On the other hand, exterior modes are not “immune” to external variations. When the boundaries expand, the distribution of exterior modes on the complex ω plane becomes increasingly dense, which eventually yields a continuum. Thus, one may intuitively write the modal expansion solution of unbounded fields in the form of

$$\mathbf{E}(\mathbf{r}) = \sum_{n=1}^N \alpha_n \mathbf{E}_n(\mathbf{r}) + \int d\omega' \alpha(\omega') \mathbf{E}(\omega' \mathbf{r}). \quad (170)$$

The discrete summation in Eq. 170 corresponds to trapped modes, while the integral corresponds to continuous exterior modes where $\alpha(\omega')$ should include the density of states.

In numerical studies, one can replace PEC/PMC boundaries with PMLs (Berenger 1994; Chew and Weedon 1994) to emulate unbounded media. When PMLs are implemented, the eigenspectra become discrete from which trapped modes can be easily distinguished from exterior ones since the quality factors

$$Q = \left| \frac{\Re[\omega_n]}{2\Im[\omega_n]} \right| \quad (171)$$

of the former are obviously larger. Moreover, exterior modes in this case include resonances inside PMLs which have large $\Im[\omega_n]$. Using PMLs enables one to correctly capture the physics of trapped modes which are important in characterizing the field behavior of an excited system. Therefore, the model order can be greatly reduced through NME using only a few important modes.

The governing Eqs. 160 and 161 can be discretized with conventional differential equation solvers, such as FDM and FEM. The resultant generalized eigenvalue equations only have sparse matrices and therefore can be easily solved with iterative eigensolvers, such as Lanczos and Arnoldi methods (Lehoucq and Sorensen 1996). Shift-invert techniques can be used to calculate interior spectrum of the system, from which one may extract trapped modes that are physically important. For example, the -1 st-, 0 th-, and $+1$ st-order modes of a composite right-/left-handed (CRLH) waveguide slot are found as shown in Fig. 14a–c. Consider an open-ended CRLH waveguide slot antenna (Fig. 15a) (Dong and Itoh 2010), where the substrate has a thickness of $h = 1.27$ mm and a relative permittivity of Eq. 44. Parameters in Fig. 15a are set as $\omega_1 = 0.3$ mm, $\omega_2 = 0.45$ mm, $\omega_3 = 9.6$ mm, and $\omega_s = 3.9$ mm. Figure 15b shows 50 natural frequencies searched around 11 GHz. One can easily identify five trapped modes as they have

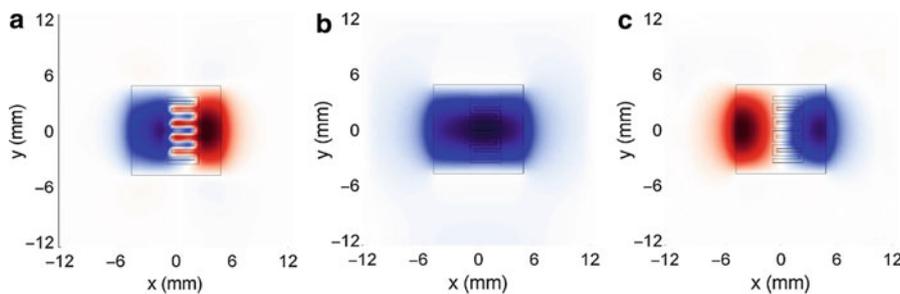


Fig. 14 Simulated modal shapes $\Re\{E_z\}$ on an xy plane for several trapped modes: (a) –1st-order mode. (b) 0th-order mode. (c) +1st-order mode

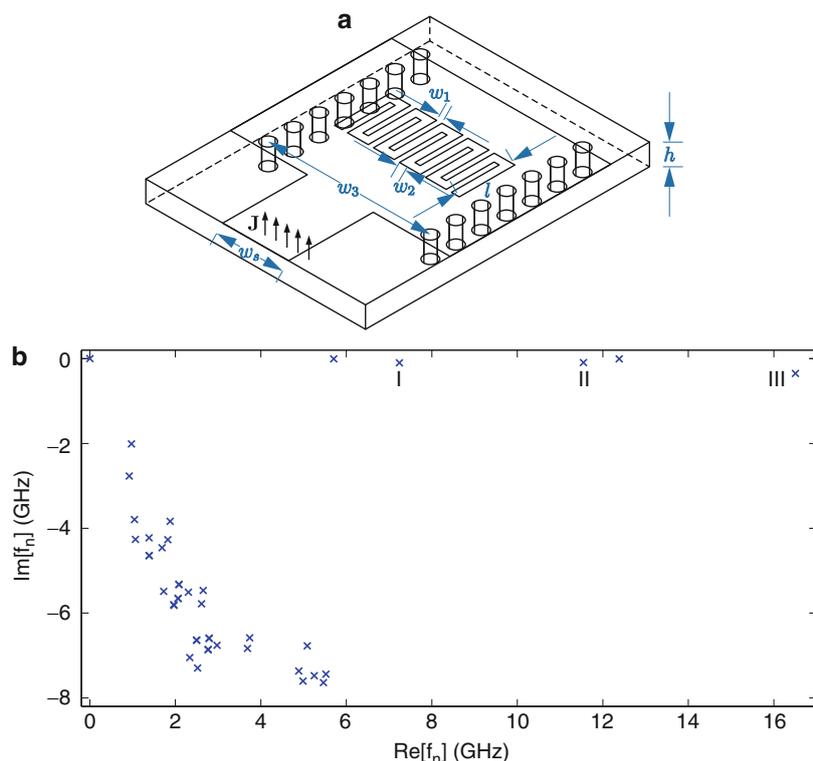


Fig. 15 (a) Open-ended CRLH-SIW slot antenna. (b) Computed complex eigenfrequencies around 11 GHz

small $\Im m[f_n]$, among which I, II, and III are –1st-, 0th-, and +1st-order resonances, respectively. The rest with large $\Im m[f_n]$ are exterior modes, which are well separated from trapped ones due to the existence of PMLs.

When the antenna is excited by a z -polarized current sheet operating at 5.9 GHz, the excited field $\Im m[E_z^{\text{dir}}]$ can be approximately expanded using five trapped modes as good agreement is observed between Fig. 16a, b. More quantitatively, comparisons of the directly calculated field and the expanded fields with different numbers of trapped modes along $y = 0.36$ mm are given in Fig. 16c. The relative error is shown in Fig. 16d with respect to the number of modes used to expand the field. The small error near the source (Fig. 16c) may be reduced by taking the irrotational modes into account as the source is not divergence-free. A reduced modal approximation of the excited field in a properly operating antenna is achieved since only a few trapped modes dominate the expanded solution. Provided that trapped modes can be solved for by the shift-invert technique in a preprocessing stage, NME is well suited for a fast wideband field

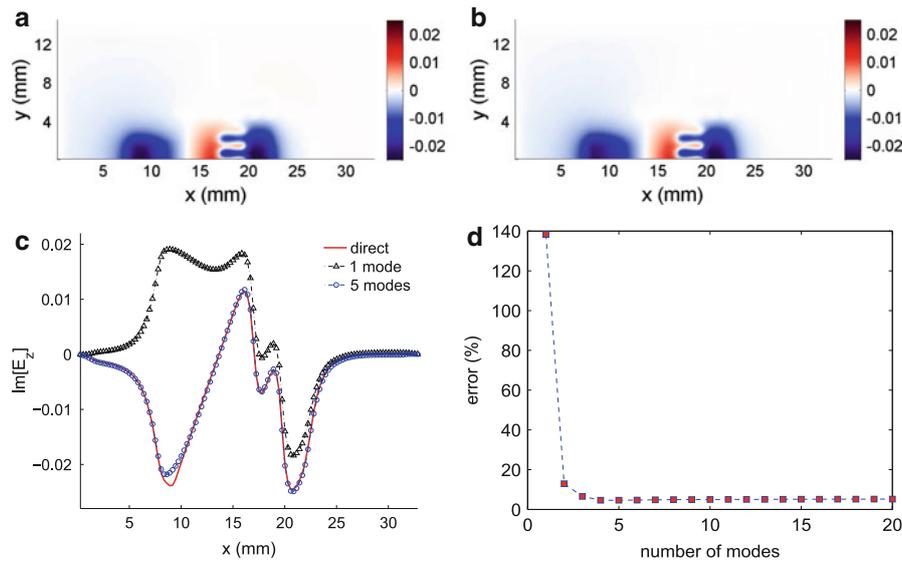


Fig. 16 A current sheet operating at 5.9 GHz is introduced to the CRLH waveguide slot antenna. $\mathcal{I}m[E_z]$ of (a) directly calculated field on xy plane at $z = 0.635$ mm. (b) Expanded field using trapped modes on the same plane. (c) Comparisons of directly calculated and expanded $\mathcal{I}m[E_z]$ along an observation line $y = 0.36$ mm. (d) Relative error with respect to the number of modes used to expand the field

approximation and for problems with different source configurations since the field reconstruction with a few modes consumes negligible CPU time.

Incorporation of PMLs in natural mode analysis leads to linear eigenvalue problems since PMLs behave as a wideband absorber for electromagnetic waves, whose working frequency can always be predetermined. This greatly simplifies the calculation of natural modes and renders NME a much simpler expression than that in Baum et al. (1991). Besides indicating modal radiation ability by modal quality factors, natural mode analysis also provides guidance on source placement for the excitation of desired modes. Furthermore, NME offers more flexibility in modeling complicated antennas than CMA where dielectric parts are included.

Equivalence Principle Algorithm for Antennas

The lack of analytical methods in terms of closed-form formulas or approximate asymptotic expressions has been a real motivation for antenna engineers to invoke computational electromagnetic methods to analyze and design antennas. Although the currently available numerical methods can address many problems in antenna structures modeling, there are other issues mostly related to multi-scale problems which cannot be properly addressed by these methods (Li and Chew 2007).

Multi-scale Problems

In general, traditional modeling techniques mainly focus on one scale (e.g., macroscopic or microscopic scales) and the physics change significantly over these scales. For example, when the macroscopic scale (or macroscale) is used, only the effect of the larger scales (sizes) is of interest and the effect of the smaller scales is modeled by some constitutive relations (Weinan 2012). It is known that macroscopic scale is efficient and microscopic scale is accurate. Hence, simultaneous incorporation of models at different scales helps to exploit the advantages offered by all of those scales. This so-called multi-scale concept for the electromagnetic modeling has been the research focus in the past two decades. A typical multi-scale

electromagnetic problem can be raised because of the nature of the electromagnetic fields. In fact, electromagnetic physics can be divided into: circuit physics, wave physics, and ray physics. From a numerical point of view, to model a multi-scale problem where both low-frequency circuit physics and high-frequency wave physics coexist, disparate meshing is not avoidable. This leads to a large number of unknowns reducing the convergence rate of the iterative solvers. As an example, consider a small patch antenna mounted on roof of a car. The antenna dimensions are in centimeters while the dimensions of the car are in meters. These are examples of multi-scale problems. To address these problems, a few ideas have been suggested where the domain decomposition method (DDM), the most popular one, will be discussed next.

Domain Decomposition Method

The main idea of domain decomposition method relies on dividing a larger solution domain into smaller nonoverlapping subdomains. Each of these individual subdomains can be solved independently while the interaction between adjacent subdomains is included through interface boundary conditions to guarantee the unique solution. In fact, the solutions of each subdomain are stitched together using some transmission condition (e.g., Robin-type transmission condition) and iterative algorithm to derive the solution of the large problem (Lu et al. 2008). In this manner, the solution of each subdomain is independent of the other subdomains. Hence, the meshing strategy can be optimally selected for each subdomain. The freedom to select the mesh strategy helps to facilitate the parallelization of the solution and its reuse and also to improve the condition number of the matrix system. These advantages have encouraged the combination of DDM with other electromagnetic solvers such as finite-element method and finite-difference frequency-domain method. Then, the domain decomposition method was used for solving three-dimensional large finite periodic problems such as photonic or electromagnetic bandgap structures and frequency-selective surfaces. In addition, relatively large dielectric aperture array, patch antenna array, and slotted waveguide array were successfully analyzed using DDM. To extend the application of DDM for integral equation-based solvers, the equivalence principle algorithm (EPA) was introduced by Chew (Chew and Lu 1993; Lu and Chew 1995).

Equivalence Principle Algorithm

Based on EPA, the entire solution domain is decomposed into many subdomains by defining appropriate closed surfaces which are called equivalence surfaces. The subdomains are enclosed by the equivalence surfaces and communicate mutually only through the closed surfaces. Consider a simple scenario where there are only two objects, namely, Object1 and Object2, and they are enclosed by two separate surfaces, namely, S_1 and S_2 , respectively (Fig. 17). The entire EPA procedure for this case can be cast into four steps:

1. Outside-in propagation

Initially, the collective effect of every other exterior sources is modeled by equivalent magnetic and/or electric currents on the equivalence surface S_1 . These equivalent sources generate the original fields inside the S_1 and zero field outside. These equivalent currents, called incident currents are denoted as \mathbf{J}_1^{inc} and \mathbf{M}_1^{inc} .

2. Solving for the induced current on enclosed object

Having the incident field on Object1 produced by the equivalence currents and by invoking a solution method (e.g., MoM) the induced currents on Object1 denoted as \mathbf{J}_1 and \mathbf{M}_1 , can be found.

3. Inside-out propagation

Having found the currents on Object1, the scattered field distribution can be readily calculated anywhere inside or on the surface of S_1 . Then, the equivalence principle can be used to calculate the

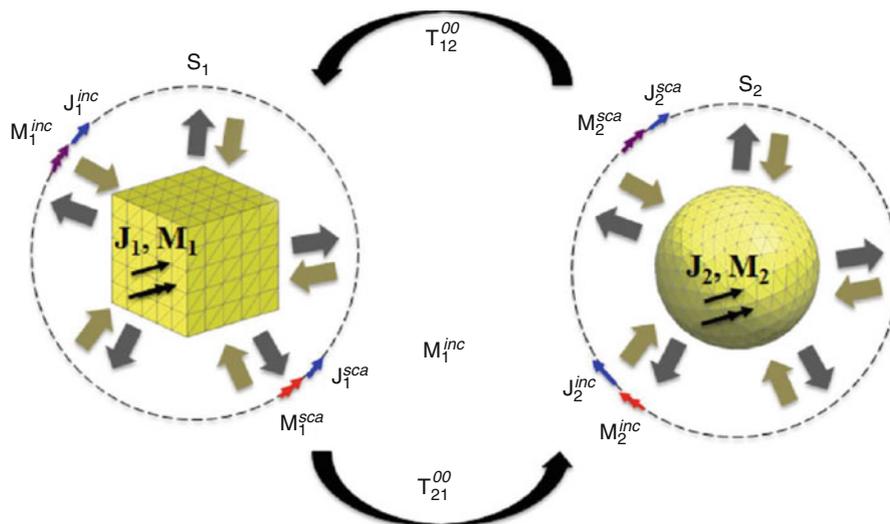


Fig. 17 A simple two-object scenario to explain the equivalence principle algorithm

equivalent currents on the surface (by defining the region of interest to be outside of S_1). These equivalent currents, called scattered currents are denoted as \mathbf{J}_1^{sca} and \mathbf{M}_1^{sca} .

A similar procedure is applied to the other defined equivalence surfaces and the enclosed objects in a parallel manner. However, the procedure will not capture the whole physics until the interaction between the surfaces is also considered.

4. Interaction between equivalence surfaces

The interaction between S_1 and other equivalence surfaces (i.e., S_2 in this case) is modeled by introduction of an operator called the translation operator. Its main function is to transfer the effect of the equivalent currents from one equivalence surface to another equivalence surface. In contrast with domain decomposition methods for differential equation-based solvers, the subdomains do not have to touch each other (i.e., white space between the subdomains is allowed).

The entire EPA procedure for the two-object two-equivalence surface scenario can be described through the following equations:

$$\begin{bmatrix} \mathbf{J}_m^{sca} \\ \mathbf{M}_m^{sca} \end{bmatrix} = \begin{bmatrix} -\hat{n} \times \mathcal{K}_H^s \\ -\hat{n} \times \mathcal{L}_E^s \end{bmatrix} \cdot [\mathcal{L}_{\text{solver}}^{-1}] \cdot \begin{bmatrix} -\mathcal{L}_E^s & -\mathcal{K}_H^s \end{bmatrix} \cdot \begin{bmatrix} \mathbf{J}_m^{inc} \\ \mathbf{M}_m^{inc} \end{bmatrix} = S_{mm} \cdot \begin{bmatrix} \mathbf{J}_m^{inc} \\ \mathbf{M}_m^{inc} \end{bmatrix} \quad (172)$$

$$\begin{bmatrix} \mathbf{J}_m^{sca} \\ \mathbf{M}_m^{sca} \end{bmatrix} = \begin{bmatrix} -\hat{n} \times \mathcal{K}_H^s & -\hat{n} \times \mathcal{L}_H^s \\ -\hat{n} \times \mathcal{L}_E^s & -\hat{n} \times \mathcal{K}_E^s \end{bmatrix} \cdot \begin{bmatrix} \mathbf{J}_n^{sca} \\ \mathbf{M}_n^{sca} \end{bmatrix} = T_{mn}^{oo} \cdot \begin{bmatrix} \mathbf{J}_n^{sca} \\ \mathbf{M}_n^{sca} \end{bmatrix} \quad (173)$$

where \mathcal{L}_E , \mathcal{K}_E , \mathcal{L}_H , and \mathcal{K}_H are integral operators and they are well defined in the literature (Li and Chew 2007). Also $\mathcal{L}_{\text{solver}}^{-1}$ represents the current solver part. Moreover, S_{mm} represents scattering operator for equivalence surface S_m and T_{mn}^{oo} shows the translation operator between surface m and n where $m, n = \{1, 2\}$. Combination of these two equations results in a system of equations as

$$\begin{bmatrix} \mathbf{J}_1^{sca} \\ \mathbf{M}_1^{sca} \end{bmatrix} - S_{11} \cdot T_{12}^{oo} \cdot \begin{bmatrix} \mathbf{J}_2^{sca} \\ \mathbf{M}_2^{sca} \end{bmatrix} = S_{11} \cdot \begin{bmatrix} \mathbf{J}_1^{inc} \\ \mathbf{M}_1^{inc} \end{bmatrix} \quad (174)$$

$$-S_{22} \cdot T_{21}^{oo} \cdot \begin{bmatrix} \mathbf{J}_1^{sca} \\ \mathbf{M}_1^{sca} \end{bmatrix} + \begin{bmatrix} \mathbf{J}_2^{sca} \\ \mathbf{M}_2^{sca} \end{bmatrix} = S_{22} \cdot \begin{bmatrix} \mathbf{J}_2^{inc} \\ \mathbf{M}_2^{inc} \end{bmatrix} \quad (175)$$

By solving this system of equations, the unknown scattered equivalent currents on equivalence surfaces will be calculated. Similar equations can be derived for more objects with more equivalence surfaces.

EPA for Connected Regions

The EPA procedure works well for objects which are not intercepted by the equivalence surface. However, in the antenna applications and many other electromagnetic problems, unconnected (isolated) region models are seldom used and one region may be connected with some other regions through wires, lumped elements, switches, and ground plane. In this case, the original continuous currents will be cut by the equivalence surface which leads into electric charge accumulation on both sides at the connection. This charge accumulation and current discontinuity result in a singular field generation which will induce a singular equivalence current on the intercepting equivalence surfaces and affects the accuracy of the results. Moreover, having objects very close to the equivalence surface requires computationally costly singularity subtraction and near-field treatments. In the literature, three methods, namely, singularity matching approach, fictitious junction basis method, and the tap basis scheme have been suggested to retain the current continuity and avoid the singularity. Tap basis scheme has been shown to model the continuous current accurately and no singular integral will be involved (Li and Chew 2007, 2008).

Tap Basis Scheme

For scenarios where the enclosed objects by an equivalence surface are connected to other objects enclosed by other equivalence surfaces, a tap basis scheme has been introduced in (Li and Chew 2007, 2008) to keep the current flowing smoothly in and out of the equivalence surfaces. Assume a metallic strip which is divided into two sections by defining two-equivalence surfaces (Fig. 18a). The strip is then divided into three regions by defining a transition region or tap basis region (Fig. 18b). To account for the tap basis, Eqs. 174 and 175 should be modified as

$$\begin{bmatrix} \mathbf{J}_1^{sca} \\ \mathbf{M}_1^{sca} \\ \mathbf{t}_{t1} \end{bmatrix} - S_{11} \cdot \begin{bmatrix} T_{12}^{oo} \cdot \begin{bmatrix} \mathbf{J}_2^{sca} \\ \mathbf{M}_2^{sca} \end{bmatrix} \\ -\mathcal{L}_{tt}^{-1} \cdot (\mathbf{t}_{t1} + \mathbf{t}_{t2}) \end{bmatrix} = S_{11} \cdot \begin{bmatrix} \mathbf{J}_1^{inc} \\ \mathbf{M}_1^{inc} \\ -\mathcal{L}_{tt} \cdot \mathbf{E}_t^i \end{bmatrix} \quad (176)$$

$$\begin{bmatrix} \mathbf{J}_2^{sca} \\ \mathbf{M}_2^{sca} \\ \mathbf{t}_{t2} \end{bmatrix} - S_{22} \cdot \begin{bmatrix} T_{21}^{oo} \cdot \begin{bmatrix} \mathbf{J}_1^{sca} \\ \mathbf{M}_1^{sca} \end{bmatrix} \\ -\mathcal{L}_{tt}^{-1} \cdot (\mathbf{t}_{t1} + \mathbf{t}_{t2}) \end{bmatrix} = S_{22} \cdot \begin{bmatrix} \mathbf{J}_2^{inc} \\ \mathbf{M}_2^{inc} \\ -\mathcal{L}_{tt} \cdot \mathbf{E}_t^i \end{bmatrix} \quad (177)$$

where \mathcal{L}_{tt}^{-1} and \mathbf{E}_t^i are the current solver operator and the incident field on the tap basis part, respectively. Moreover, $\mathbf{t}_{ti} = \mathcal{L}_{ti} \cdot \mathbf{J}_i$ represents the induced field by current on i -th metallic object (i.e., \mathbf{J}_i) on the tap region. Expressions Eqs. 176 and 177 constitute six equations for six sets of unknowns $[\mathbf{J}_1^{sca}, \mathbf{M}_1^{sca}, \mathbf{t}_{t1}, \mathbf{J}_2^{sca}, \mathbf{M}_2^{sca}, \mathbf{t}_{t2}]$. The equations can be set up for the N object case. If no tap basis is involved with the object, the bottom rows of the equations disappear, and one reverts back to the ordinary EPA (Eqs. 174 and 175).

Antenna Modeling Using EPA in Literature

In the past two decades, the equivalence principle algorithm has been successfully used to solve multi-scale electromagnetic problems including antenna problems. Some of the reported works in the literature for analyzing antennas using EPA are summarized in Tables 1 and 2. The applications of EPA to analyze

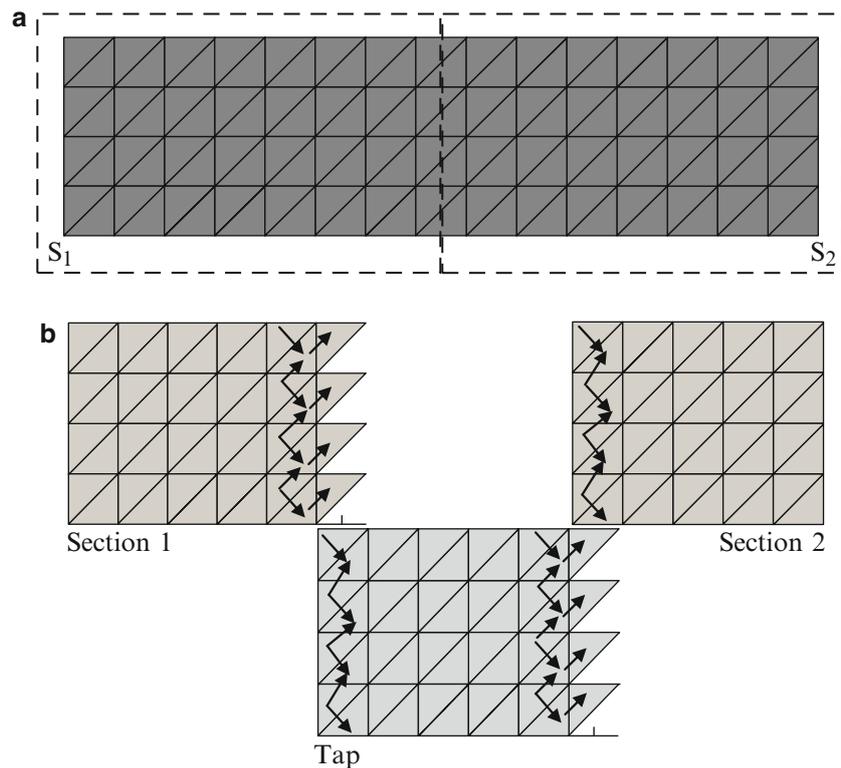


Fig. 18 Metallic strip. (a) Divided into two sections by two-equivalence surfaces, (b) introduction of tap basis to keep the current continuity

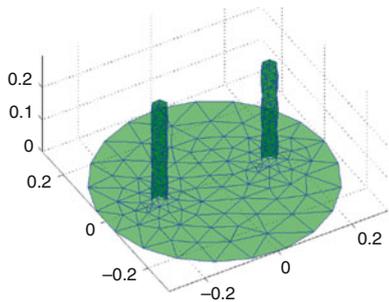
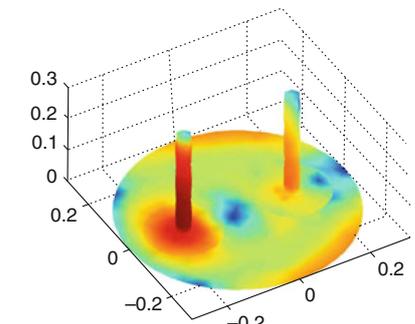
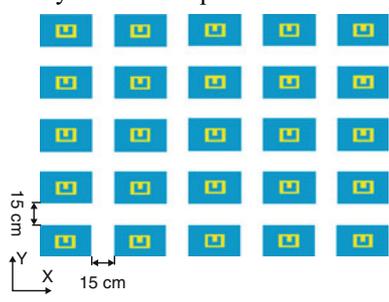
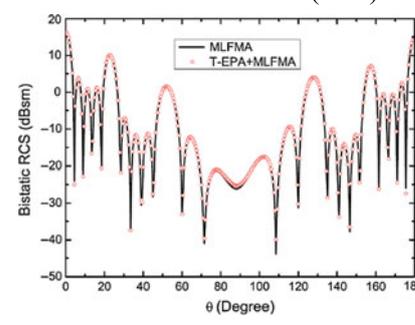
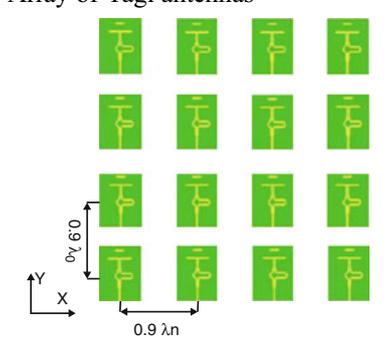
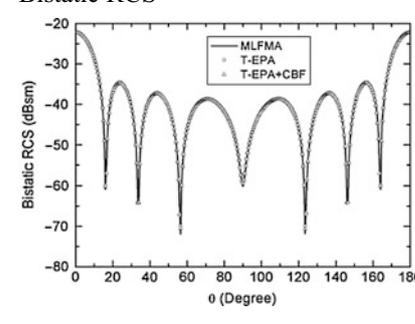
array antennas are listed in Table 1. Array of monopoles, U-shaped slot microstrips, and Yagi antennas have been successfully modeled and solved using EPA. In Table 2, applications of EPA for some other special cases are listed where EPA has been applied for analyzing an antenna mounted on the roof of a car, reconfigurable pixelated antennas, and a bow-tie antenna. Overall, the application of EPA helped significantly reduce the computational burden and accelerate the calculations.

Recently, several ideas have been suggested to improve the accuracy and convergence rate of EPA. Most of these works try to address problems associated with the equivalence surfaces. Application of multilevel fast multipole algorithm to calculate the interactions between the equivalence surfaces significantly accelerated the whole EPA procedure. Also, in Shao et al. (2014), utilization of grid-robust higher-order vector basis (GRHOVB) instead of curvilinear RWG function for the equivalence surfaces has been proposed to improve the accuracy of EPA. In Shao et al. (2012), instead of solving for both electric and magnetic currents on the equivalence surfaces, the concept of generalized impedance boundary condition (GIBC) has been utilized to define a relationship between these currents and reduce the number of unknowns. Despite all of these endeavors, EPA is still an open area for further research. The goal of the future works is to improve the accuracy and the speed of convergence as well as modifying EPA for newly emerging applications such as solar cells, implantable devices, and nanoscale structures.

High-Frequency Techniques for Antennas

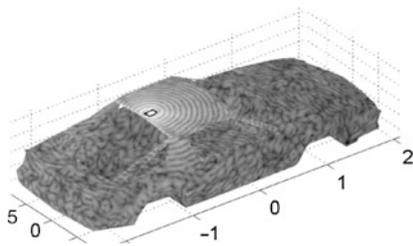
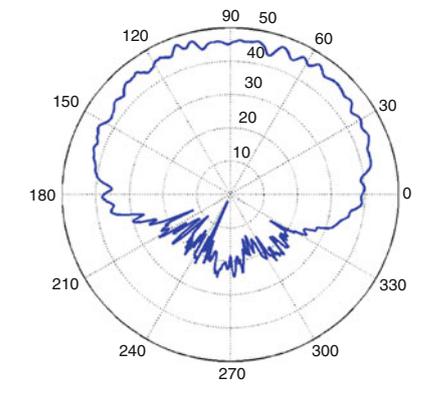
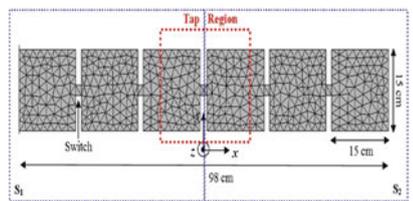
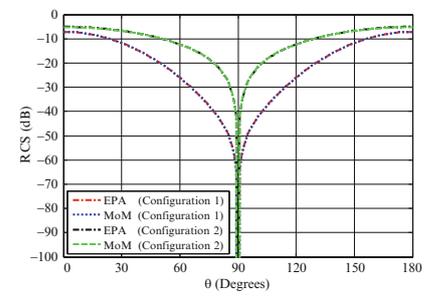
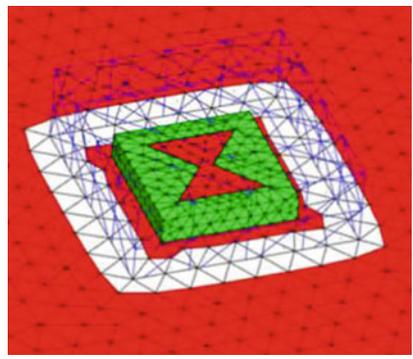
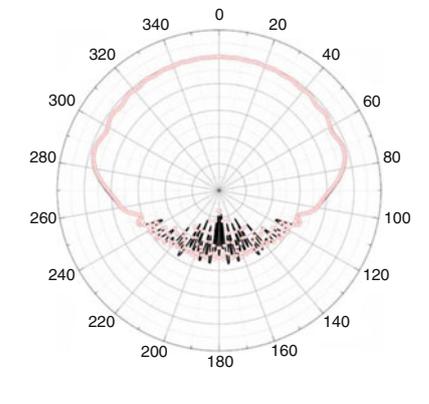
For antennas with a reflector or a platform that is large compared to the wavelength, high-frequency techniques can be used to simulate electromagnetic scattered fields very efficiently. They are important in three aspects (Samii and Israel 1980; Yaghjian 1984; Chou et al. 2005). First, high-frequency methods are

Table 1 Application of EPA to model array antennas (Li and Chew 2007; Shao et al. 2011, 2012, 2013)

Antenna type	Desired output
Two monopoles on ground plane	Current distribution
	
Array of microstrip antennas	Bistatic radar cross section (RCS)
	
Array of Yagi antennas	Bistatic RCS
	

much faster than full-wave methods on calculating electromagnetic scattered fields from electrically large antennas. Second, high-frequency methods could provide physical insights for the understanding of the radiation mechanisms of antennas. Typical high-frequency insights lead to wave physical critical points: the shadow and reflection boundaries, the caustics from surfaces of antennas, etc. Third, for complicated antenna structures, high-frequency methods have potential to be combined with the full-wave method well to solve radiation problems. Specifically, wave fields in electrically large parts exhibit the high-frequency ray physics phenomenon. While in electrically small and moderate parts, wave fields have circuit and wave physics behaviors, respectively. Thus, once the hybrid solvers in terms of high-frequency methods and full-wave solvers are developed, the efficiency of the electromagnetic wave simulation could be greatly enhanced.

Table 2 Application of EPA to solve antenna problems (Li and Chew 2008; Fallahpour et al. 2014; Shao et al. 2014)

Antenna type	Desired output
<p data-bbox="129 295 502 331">Antenna mounted on roof of a car</p> 	<p data-bbox="1037 295 1244 331">Radiation pattern</p> 
<p data-bbox="129 750 478 786">Reconfigurable pixelled antenna</p> 	<p data-bbox="1037 750 1101 786">RCS</p> 
<p data-bbox="129 1120 446 1155">Bow tie antenna on PEC plate</p> 	<p data-bbox="1037 1120 1244 1155">Radiation pattern</p> 

When the product of the wave frequency k and the diameter of the antenna a becomes tens to thousands, the electromagnetic waves assume the high-frequency localization phenomenon. In this sense, the high-frequency scattered fields have the high-frequency wave physics phenomenon and ray physics. Thus, high-frequency wave fields propagate like rays. With the aid of the high-frequency ray physics theory, various high-frequency methods were developed in the last century. For instance, for calculating the reflected and transmitted wave fields, the high-frequency methods contain the geometric optics (GO) method (Ling et al. 1989) and the physical optics (PO) method (Macdonald 1913). While for computing the diffracted wave fields in Fig. 19a, these methods correspond to the geometry theory of diffraction (GTD) (Keller 1962; Hansen 1981), uniform theory of diffraction (UTD) (Kouyoumjian and

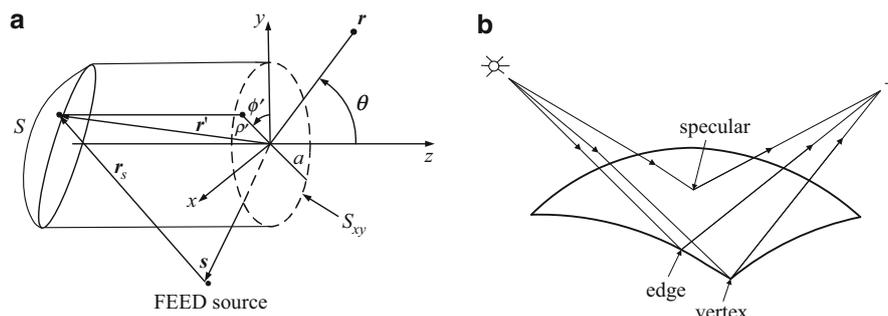


Fig. 19 (a) A shaped reflector antenna system. (b) High-frequency ray fields from the scatterer, including the reflected fields from the specular points and the diffracted fields from the edge and vertex

Pathak 1974), uniform asymptotic diffraction (UAT) (Lee and Deschamps 1976), physical theory of diffraction (PTD) (Ufimtsev 2005), and incremental length diffraction coefficients (ILDC) (Yaghjian et al. 1996). In the following, these high-frequency methods are introduced in an elegant way. Furthermore, fast solvers for the high-frequency scattered fields are addressed, including the high-frequency asymptotic method (HFA) (Conde et al. 2001; Carluccio et al. 2010) and the numerical steepest descent path method (NSDP) (Wu et al. 2012a, b, 2013a, b, c, 2014).

Radiation from Antennas

For the reflector antenna system with a location of the feed as shown in Fig. 19a, the electric scattered far fields could be expressed as

$$\mathbf{E}^{(s)}(\mathbf{r}) \approx \frac{ike^{ikr}}{4\pi r} \sqrt{\frac{\mu}{\epsilon}} (\bar{\mathbf{I}} - \hat{r}\hat{r}) \cdot \int_S d\mathbf{r}' e^{-ik\hat{r}\cdot\mathbf{r}'} \mathbf{J}(\mathbf{r}') \quad (178)$$

Here, $\mathbf{J}(\mathbf{r}')$ is the induced electric current on the surface of the antenna, S , and \mathbf{r} and \mathbf{r}' denote the observation and integration source points. The projection aperture of S in the xy plane is the circular region, S_{xy} , with radius a as shown in Fig. 19a. Then, the radiation integrals $\mathbf{F}(\mathbf{r})$ could be obtained from Eq. 178, with the formulation

$$\mathbf{F}(\mathbf{r}) = \int_S d\mathbf{r}' e^{-ik\hat{r}\cdot\mathbf{r}'} \mathbf{J}(\mathbf{r}') = \int_0^a \int_0^{2\pi} e^{-ik\hat{r}\cdot\mathbf{r}'} \mathbf{J}(\mathbf{r}') \rho' T(\rho', \phi') d\rho' d\phi' \quad (179)$$

Here, $T(\rho', \phi')$ is the Jacobian transformation function for the antenna surface between (x, y, z) and (ρ, ϕ, z) coordinate systems.

The vector surface integrals $\mathbf{F}(\mathbf{r})$ in Eq. 179 are called the radiation integrals. Furthermore, the surface integral contains the phase information of the dyadic Green's function in the far-field region, the induced electric current, and the integration of the curvature from S .

High-Frequency Techniques

Geometric Optics Method

The ray path between any two points \mathbf{p}_1 and \mathbf{p}_2 in three-dimensional space follows Fermat's principle, that is,

$$\delta \int_{\mathbf{p}_1}^{\mathbf{p}_2} n(s) ds = 0 \quad (180)$$

Here, $n(s)$ is the refraction index of the medium and δ is the variation operator in variational calculus. Equation 180 expresses that the ray follows the path that minimizes the optical distances between \mathbf{p}_1 and \mathbf{p}_2 . On invoking Fermat's principle in Eq. 180, the reflected rays could be determined from the incident field. This method is called the geometrical optics (GO) method. Furthermore, based on the ray tube theory and conservation of energy flux for light intensity, the reflected field $\mathbf{E}^{(r)}(\mathbf{r})$ at the distance s from the reflection point \mathbf{p}_r of the scatterer could be expressed as

$$\mathbf{E}^{(r)}(\mathbf{r}) = \mathbf{E}^{(i)}(\mathbf{p}_r) \cdot \bar{\mathbf{R}} \sqrt{\frac{\rho_1(\mathbf{p}_r)\rho_2(\mathbf{p}_r)}{(\rho_1(\mathbf{p}_r) + s)(\rho_2(\mathbf{p}_r) + s)}} e^{iks} \quad (181)$$

Here, $\bar{\mathbf{R}}$ is the dyadic reflection coefficient, $\rho_1(\mathbf{p}_r)$ and $\rho_2(\mathbf{p}_r)$ are the principle radii of S at the point \mathbf{p}_r of reflection, and s is the distance from \mathbf{p}_r to \mathbf{r} along the direction of reflected field. For the antenna with penetrable material, the transmitted wave fields also exist. The typical GO method for calculating the wave fields from the complicated cavity structure is the shooting and bouncing ray method (SBR) (Ling et al. 1989). Actually, by adopting the SBR method, multiple scattering wave fields are captured and calculated efficiently.

The GO method for calculating the scattered field has close connection with the scattered field in Eq. 178. On invoking the high-frequency asymptotic method for Eq. 178 in section “Basic Circuit Laws from Maxwell's Equations” in the following, one can obtain that the leading term in Eq. 178 corresponds to the GO ray field at the specular reflection point.

Geometric Theory of Diffraction

When the rays come to the wedges, the diffracted wave fields and the reflection and shadow boundaries come out. In this situation, for the case that the observation point lies in the transition and deep shadow regions, the GO method fails to capture the scattered wave fields very well. To remedy this aforementioned failure, Sommerfeld addressed the early high-frequency scattering problem from the perfect conducting half-plane (Sommerfeld 1896). The scattered field was expressed in terms of the Fresnel integral. By performing the asymptotic expansion of the Fresnel integral, the scattered far field could be easily evaluated. In the high-frequency diffracted wave current like the fringe current, the shadow and reflection boundaries were clearly demonstrated. Moreover, the regions for the observation point lying around the reflection and shadow boundaries are called transition regions.

The calculation of the diffracted ray field from the perfectly conducting curved wedge was developed by Keller (1962). This method was called the geometry theory of diffraction (GTD). The diffracted wave fields from the point \mathbf{p}_d along the curved wedge has the similar formulation as (Eq. 181):

$$\mathbf{E}^{(d)}(\mathbf{r}) = \mathbf{E}^{(i)}(\mathbf{p}_d) \cdot \bar{\mathbf{D}} \sqrt{\frac{\rho_d(\mathbf{p}_d)}{s(\rho_d(\mathbf{p}_d) + s)}} e^{iks} \quad (182)$$

Here, $\bar{\mathbf{D}}$ is the dyadic diffraction coefficient in (Keller 1962), $\rho_d(\mathbf{p}_d)$ is the principle radius of the curved wedge at the point \mathbf{p}_d of reflection, and s is the distance from \mathbf{p}_d to \mathbf{r} along the direction of the reflected field.

Furthermore, the diffraction coefficients $\bar{\mathbf{D}}$ could be obtained from the asymptotic expansion of the radiation integrals of the Sommerfeld half-plane diffraction problems. Thus, the fields from the GTD method contain both reflected and transmitted wave fields from GO method and the diffracted wave fields. Therefore, compared to the GO method, the GTD method provides a relatively mature way for calculating the high-frequency wave fields.

Again, on invoking the high-frequency asymptotic method in section “[Basic Circuit Laws from Maxwell’s Equations](#)” for Eq. 178, the higher-order terms correspond to the diffracted ray fields. Moreover, these obtained diffracted ray fields agree well with the results generated from the GTD method. However, at the transition regions, these GTD diffraction coefficients become infinite; thus, the GTD diffracted wave fields become singular. Therefore, when the observation point lies around the transition regions, the GTD method breaks down.

Uniform Theory of Diffraction

To remedy the failure of the GTD method at the transition region, people developed the efficient uniform theory of diffraction (UTD) (Kouyoumjian and Pathak 1974) and uniform asymptotic diffraction (UAT) methods (Lee and Deschamps 1976). With the high-frequency wave physics aspect, on invoking the asymptotic technique, the radiation integrals for the diffracted wave fields from the wedge could be expressed as the superposition of contributions from high-frequency critical points. This superposition of critical-point contributions could also be regarded as the diffracted rays with the ray physics aspect. However, when the observation points lie in transition regions, these critical points will coalesce. In this situation, the asymptotic approximation of the radiation integrals fails to approximate these integrals well.

To account for the coupling of high-frequency critical points, the uniform asymptotic approximation of radiation integral could be adopted to remedy the breakdown of the GTD method at transitions. This method was called the UTD method, which provides the UTD diffraction coefficients in terms of the UTD transition function. The UTD transition function with the Fresnel function involved is used to compensate the GTD breakdown and makes the wave fields continuous around the transition regions. Meanwhile, with the observation point far away from the transition regions, the wave field results from the UTD and GTD methods could agree well with each other. Once Eq. 178 is derived in the uniform asymptotic way, higher accuracy for the diffracted wave fields will result in near-transition regions. The UTD method follows the asymptotic of the radiation integral. Thus, the breakdown problem of the GTD method around the transition regions could be well solved by the UTD method.

Physical Optics Method

The physical optics (PO) approximation serves as an efficient approach for analyzing the scattered electromagnetic fields. Early in 1913, Macdonald introduced the PO approximation concept by approximating the induced current on the surface of the consider scatterer (Macdonald 1913). The surface PO-induced current is approximated by doing the local tangent plane approximation, that is, the integration small surface patch along the electrically large object is assumed to be locally flat and smooth under the external high-frequency condition (Ufimtsev 2008). In this sense, for the incident high-frequency electromagnetic waves on the electrically large real-world objects, the PO approximation has revealed itself as an efficient way to calculate the scattered field (Lee and Mittra 1983). However, the PO current is not correct around the shadow boundary (Ufimtsev 2005), which leads to the fact that the physical optics method does not satisfy the reciprocity theorem.

With ka going from tens to thousands, by making use of the local tangential plane approximation for Eq. 179, the induced current in the lit region of considered scatterer could be approximated as

$$\mathbf{J}_{PO}(\mathbf{r}) = 2\hat{n}(\mathbf{r}) \times \mathbf{H}^{(i)}(\mathbf{r}) \quad (183)$$

Equation 183 is known as the PO current. Here, $\hat{n}(\mathbf{r})$ is the outward unit normal to the surface and $\mathbf{H}^{(i)}(\mathbf{r})$ is the incident field on the reflector antenna from the feed source. Here, $\mathbf{H}^{(i)}(\mathbf{r})$ in Eq. 183 is directly incident field generated by the feed to location \mathbf{r} with the formulation

$$\mathbf{H}^{(i)}(\mathbf{r}) \simeq C \frac{e^{ikr_s}}{r_s} \mathbf{g}(\theta, \varphi) \quad (184)$$

with C as a known complex constant and $\mathbf{g}(\theta, \varphi)$ as the vector radiation function from the feed source, and it is assumed to be known.

After substituting \mathbf{J}_{PO} in Eq. 183 into \mathbf{J} in Eq. 178, one can obtain the PO scattered field from the reflector antenna. It could be represented by three radiation surface integrals. More specifically, the phase function in Eq. 178 contains the information from the dyadic Green's function, the shape function for the surface of the reflector antenna, and the incident magnetic field from the feed source. Owing to the highly oscillatory nature of the radiation integrand in Eq. 179, the direct numerical scheme for evaluating the integral makes the computational workload increase vastly when ka is large. The problem is more serious when there are many the observation points. Therefore, one needs to seek the efficient techniques to evaluate these PO integrals (Burkholder and Lee 2005).

Physical Theory of Diffraction

To remedy the limitation of the PO current, Ufimtsev introduced the edge wave current to correct the PO current, known as the physical theory of diffraction (PTD) method (Ufimtsev 2005). The PTD current, denoted as \mathbf{J}_{PTD} , takes the expression

$$\mathbf{J}_{PTD}(\mathbf{r}) = \mathbf{J}_{PO}(\mathbf{r}) + \mathbf{J}_U(\mathbf{r}) \quad (185)$$

Here, $\mathbf{J}_U(\mathbf{r})$ is the edge wave current which is used to correct the limited accuracy of $\mathbf{J}_{PO}(\mathbf{r})$ around the shadow boundary and in the shadow regions. Actually, $\mathbf{J}_U(\mathbf{r})$ could be obtained from the scattered wave solution of the canonical perfect conducting wedge scatterer (Ufimtsev 2005). After substituting $\mathbf{J}_{PTD}(\mathbf{r})$ in Eq. 185 into Eq. 178, one can obtain the scattered fields from the PTD method. The resultant scattered fields contain the PO scattered fields and the edge diffracted wave fields, which could also be regarded as the radiation integrals with highly oscillatory integrands. Compared to the GTD and the UTD method, the PTD method gives more accurate results than the GTD and the UTD methods via the integral formulations. Meanwhile, it is time consuming to adopt the PTD method by the direct adaptive numerical scheme for these radiation integrals.

Incremental Length Diffraction Coefficient Technique

The incremental length diffraction technique introduced by Shore and Yaghjian (Hansen and Shore 1998; Yaghjian et al. 1996) provides an efficient way to calculate the high-frequency diffracted wave fields from the perfect conducting scatterer. The nonuniform (NU) diffracted wave field could be obtained by the integration of the product of the Green's function with the NU current around the surface of the scatterer, S . Furthermore, the NU-diffracted wave field could be captured by simplifying the integration domain S into $C(\mathbf{r}_s, s)$ and dl (Hansen and Shore 1998), with dl as the shadow boundary of the scatterer. In the high-frequency regime, $C(\mathbf{r}_s, s)$ conforms locally to the shape of the scatterer around the shadow boundary point \mathbf{r}_s . This technique for the determination of the NU-diffracted wave field is called the incremental length diffraction coefficient (ILDC) technique (Hansen and Shore 1998; Yaghjian et al. 1996).

In particular, for the scatterer with convex shadow boundary, the NU-diffracted wave current is called the NU-Fock current (Fock 1946). The NU-diffracted wave field could be obtained from the integration of

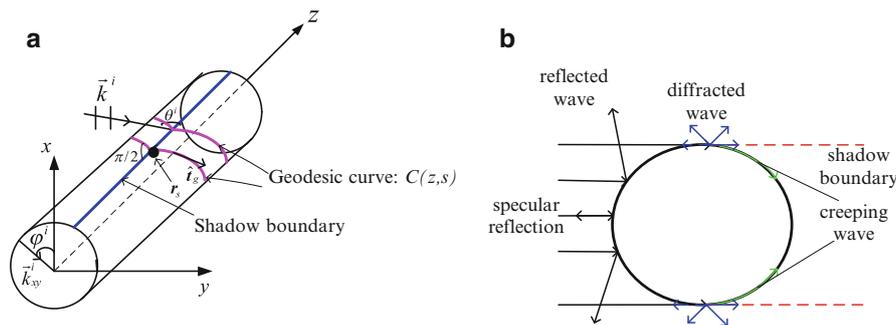


Fig. 20 (a) The obliquely incident plane wave impinges on the convex cylinder with finite length and the shadow boundary is shown by the blue line on the cylinder. The NU-Fock current in the shadow region follows the direction of the geodesic curve $C(z, s)$ by the magenta color. (b) PO current in the illuminated region and NU-Fock current around the shadow boundary and in the deep shadow region for the NU-diffracted and creeping waves

the product of the free-space Green's function and the NU-Fock current on S . Then, the resultant electric scattered field has the expression

$$\mathbf{E}^{(q)}(\mathbf{r}) = i\omega\mu \left(\bar{\mathbf{I}} + \frac{1}{k^2} \nabla \nabla \right) \cdot \int_S d\mathbf{r}' \frac{\mathbf{J}^{(q)}(\mathbf{r}') e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \quad (186)$$

The index q stands for the TE and TM cases (horizontal and vertical polarizations), and $\mathbf{J}^{(q)}(\mathbf{r}')$ is the NU-Fock current. In the far-field zone, Eq. 186 could be reformulated as

$$\mathbf{E}^{(q)}(\mathbf{r}) \approx \frac{ike^{ikr}}{4\pi r} \sqrt{\frac{\mu}{\epsilon}} (\bar{\mathbf{I}} - \hat{r}\hat{r}) \cdot \int_S d\mathbf{r}' e^{-ik\hat{r}\cdot\mathbf{r}'} \mathbf{J}^{(q)}(\mathbf{r}') \quad (187)$$

One can suppose that the integration surface S in Fig. 20 is given by $C(z, s) \times [z_0, z_1]$. Here, $C(z, s)$ is the integration curve and $[z_0, z_1]$ is the integration domain along the z direction. Then, with the strip of thickness dz as consideration, the far scattered field could be formulated as

$$d\mathbf{E}^{(q)}(\mathbf{r}) \approx dz \frac{ike^{ikr}}{4\pi r} \sqrt{\frac{\mu}{\epsilon}} (\bar{\mathbf{I}} - \hat{r}\hat{r}) \cdot \int_{C(z,s)} ds C'(z, s) e^{-ik\hat{r}\cdot C(z,s)} \mathbf{J}^{(q)}(C(z, s)) \quad (188)$$

In the illuminated region, the above $C(z, s)$ is normal to the shadow boundary. That is, $C(z, s)$ is the line along the \mathbf{k}^i direction. In the shadow region, $C(z, s)$ is the geodesic curve which forms the constant angle with the z axis as θ^i . The above Eq. 188 is known as the ILDC formulation (Hansen and Shore 1998).

Fast Solvers for High-Frequency Scattered Fields

High-Frequency Asymptotic Method

To present the high-frequency asymptotic (HFA) method in an elegant way, one can study contour integrals of the type

$$I(k) = \int_C f(z) e^{ikg(z)} dz \quad (189)$$

Here, \mathcal{C} is an integration domain in the real axis or a contour on the complex plane. Moreover, Eq. 189 could be treated as the simplification of the surface radiation integral Eq. 179 after some mathematical manipulations.

Due to the localization nature of the high-frequency wave field with the wave frequency k or the angular frequency ω approaching infinity, one can obtain an asymptotic approximation of the radiation integral in Eq. 189. This could be done by the “method of stationary phase” (or the “saddle point method”). That is, the major contribution of $I(k)$ comes from the neighborhood of the saddle point z_s satisfying $g'(z_s) = 0$. Here, one can assume that there is one saddle point z_s inside \mathcal{C} . Then, the leading term for the asymptotic expansion of Eq. 189 could be obtained by the expansion of $g(z)$ with the Taylor series and terminating at the second-order term, with the expression

$$I(k) \simeq f(z_s) e^{ikg(z_s)} \left(\frac{2\pi i}{kg''(z_s)} \right)^{1/2} \quad (190)$$

Furthermore, if $g''(z_s) = 0$, the above expansion Eq. 190 shall be modified with the Taylor series expansion of $g(z)$ to the third term. As is seen from Eq. 190, $I(k)$ is of $O(k^{-1/2})$ and becomes increasingly accurate when k goes to infinity. The saddle point in two dimensions is called the “stationary phase point.” Detailed discussion could be found in Conde et al. (2001) and Carluccio et al. (2010). By the same token, the integrand of the radiation integral in Eq. 189 is approximated by several leading terms. With the high-frequency wave physics aspect, the leading term in Eq. 190 corresponds to the GO ray fields at the specular reflection point, as shown in Fig. 19a. From Eq. 190, it is easy to see that the traditional HFA approach could provide the calculation of the radiation integral with frequency-independent workload.

Numerical Steepest Descent Path Method

Owing to the asymptotic process of the HFA method for radiation integrals in Eq. 179, the generated high-frequency scattered field results usually lose accuracy, in particular, when the parameter ka is not extremely large but lies in the high-frequency regime.

In the following, the application of the NSDP method for the PO scattered fields is considered. Since the PO scattered fields contain phase terms that have rapid variations in space, the resultant PO integrands are highly oscillatory in nature. To efficiently evaluate these surface PO integrals, the numerical steepest descent path (NSDP) method was proposed (Vico-Bondia et al. 2010; Zhang et al. 2014; (Wu et al. 2012a, b, 2013a, b, c). On invoking the NSDP method, the original PO real integration domain is deformed into the complex NSDPs on complex plane. In this manner, the PO integrands decay exponentially on these complex NSDPs, and they can be integrated with workload independent of frequency. In contrast to the HFA method, the only approximation in the NSDP method is the numerical integration of the exponential decay PO integrand. Hence, the proposed NSDP method improves the PO scattered field accuracy.

The definition of the NSDP follows that, for the given integration domain $[L_1, L_2]$ with the integrand $e^{ikg(x)}$, one can define the path functions $x = x_{L_m}(p)$, $m = 1, 2$, $p \in [0, \infty)$, satisfying three conditions (Chew 1990):

- (a) $x_{L_m}(0) = L_m$, that is, the paths start at L_m .
- (b) $\Re e(g(x_{L_m}(p))) = \Re e(g(x_{L_m}(0))) \equiv C$, where C is a constant.
- (c) $\Im m(g(x_{L_m}(p))) = p$.

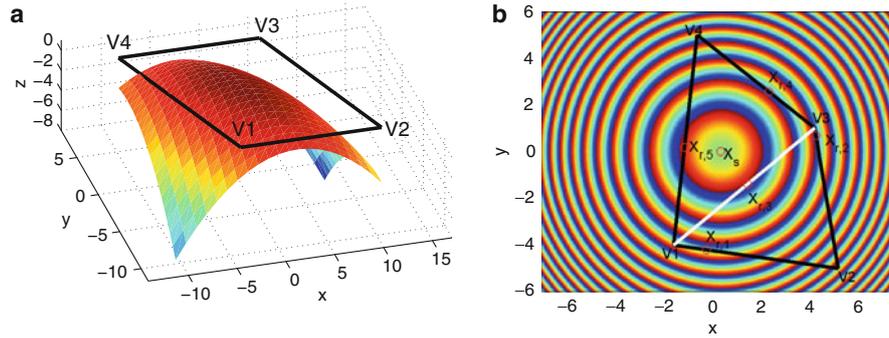


Fig. 21 (a) A parabolic PEC patch. (b) The projected x - y quadrilateral domain $V_1V_2V_3V_4$ of the parabolic PEC patch

Then, after substituting $x = x_{L_m}(p)$ into the phase function $g(x)$, one sees that $e^{ikg(x_{L_m}(p))} = e^{ikg(L_m) - kp}$ decreases exponentially as $O(e^{-kp})$ when kp goes large. Hence, the path functions $x = x_{L_m}(p)$ are called the NSDPs. Through (a), (b), and (c), the NSDPs $x_{L_m}(p)$ for the end points L_m could be obtained (Chew 1990; Wu et al. 2012a, 2013a, b).

However, the phase behavior of $g(x)$ around the stationary phase point x_s is different from that around the end points. The corresponding NSDP shall also change. One notice that $g(x) - g(x_s)$ is a quadratic function around x_s after the Taylor expansion of the phase function at x_s (Chew 1990). Therefore, one can change the above condition (c) with $\Im m(g(x_{L_m}(p))) = p^2$ and define the corresponding NSDP, $x = x_0(p)$, $p \in (-\infty, \infty)$.

In the following, the application of the NSDP method for the PO scattered fields from the quadratic parabolic PEC patches as shown in Fig. 21 is considered. The projected xy quadrilateral domain is S_{xy} , denoted as $V_1V_2V_3V_4$. The edges are governed by $y = a_mx + b_m$, $m = 1, 2, 4, 5$. To capture the high-frequency critical-point phenomenon for PO scattered fields, one can consider that the PO integrand has the quadratic variation for the phase functions on the quadrilateral patches. After substituting \mathbf{J}_{PO} into Eq. 178, the three PO surface integrals could be obtained. For simplicity, one can denote them as \tilde{I} . In the process for the reduction of the PO surface integral to several line integrals $I_2^{(a_m, b_m)}$, the complementary error function comes out in the integrand (Wu et al. 2012a, b). On invoking Stokes' phenomenon of the complementary error function, the phase functions on the edges are $g^{(m)}(x) = x^2 + (a_mx + b_m)^2$, and Stokes' lines are $y = -x - \frac{b_m}{a_m}$. Then, the NSDPs on the edges of the quadrilateral domain are (Chew 1990)

$$x_{L_m}(p) = \frac{\text{sgn}(L'_m)}{\sqrt{1 + a_m^2}} \sqrt{L_m'^2 + ip} + x_s, p \in [0, \infty) \quad (191)$$

$$x_{0,m}(p) = \frac{e^{(i\frac{\pi}{4})} p}{\sqrt{1 + a_m^2}} + x_s, p \in (-\infty, \infty) \quad (192)$$

where $x_s = -\frac{a_m b_m}{1 + a_m^2}$ is the stationary phase point of phase function $g^{(m)}(x)$ defined on edges and $L'_m = \sqrt{1 + a_m^2}(L_m - x_s)$. These NSDPs are demonstrated in Fig. 22. After changing the original PO integration path to the NSDPs and on invoking the Cauchy's integral theorem, the PO integral \tilde{I} can be represented in terms of the NSDPs as

$$\tilde{I} = -I_2^{(a_1, b_1)} + I_2^{(a_2, b_2)} + I_2^{(a_4, b_4)} + I_2^{(a_5, b_5)} \quad (193)$$

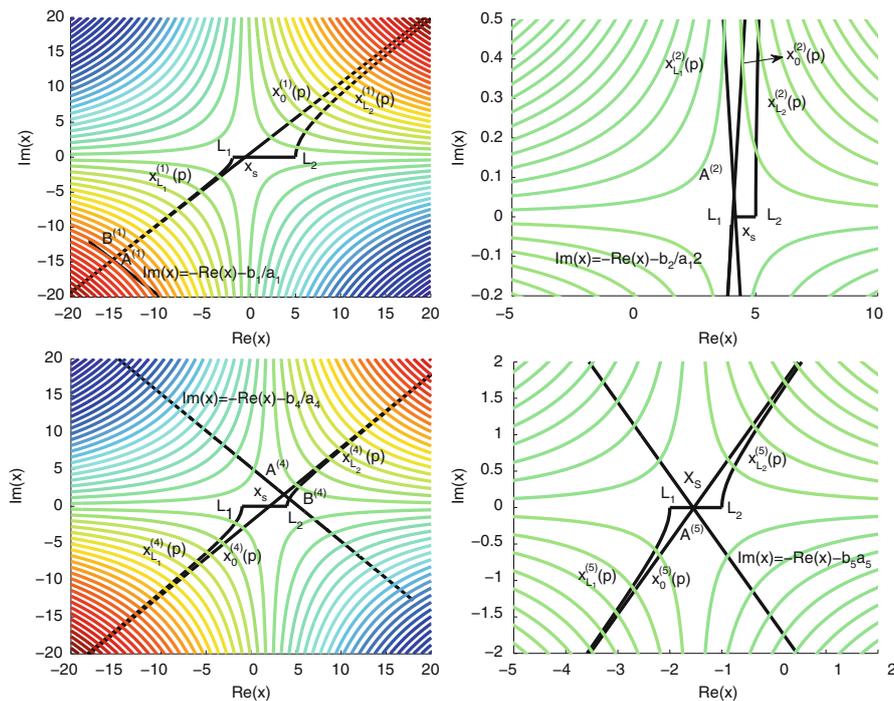


Fig. 22 The four sub-figures correspond to the NSDPs for the integrand of $I_2^{(a_m, b_m)}$, $m = 1, 2, 4, 5$, defined on the four edges of the quadrilateral domain $\mathbf{V}_1\mathbf{V}_2\mathbf{V}_3\mathbf{V}_4$ in Fig. 21

where

$$I_2^{(a_1, b_1)} = I_{L_1}^{(a_1, b_1)} - I_{L_2}^{(a_1, b_1)} + I_{x_s}^{(a_1, b_1)} + K((\mathbf{V}_2(1), 0)) - K((\mathbf{V}_1(1), 0)) - K(\mathbf{B}^{(1)}) + K(\mathbf{A}^{(1)}) \quad (194)$$

$$I_2^{(a_2, b_2)} = I_{L_1}^{(a_2, b_2)} + I_{x_s}^{(a_2, b_2)} - I_{L_2}^{(a_2, b_2)} + K((\mathbf{V}_2(1), (0))) - K(\mathbf{A}^{(2)}) \quad (195)$$

$$I_2^{(a_4, b_4)} = I_{L_1}^{(a_4, b_4)} + I_{x_s}^{(a_4, b_4)} - I_{L_2}^{(a_4, b_4)} + K(\mathbf{B}^{(4)}) - K(\mathbf{A}^{(4)}) \quad (196)$$

$$I_2^{(a_5, b_5)} = I_{L_1}^{(a_5, b_5)} + I_{x_s}^{(a_5, b_5)} - I_{L_2}^{(a_5, b_5)} + K(\mathbf{A}^{(5)}) - K((\mathbf{V}_1(1), 0)) \quad (197)$$

Originally, $I_2^{(a_m, b_m)}$, $m = 1, 2, 4, 5$, are the highly oscillatory PO line integrals $I_2^{(a, b)}$ defined on the four edges $\overrightarrow{\mathbf{V}_1\mathbf{V}_2}$, $\overrightarrow{\mathbf{V}_2\mathbf{V}_3}$, $\overrightarrow{\mathbf{V}_3\mathbf{V}_4}$, and $\overrightarrow{\mathbf{V}_1\mathbf{V}_4}$ in Fig. 21, and in Eqs. 195, 196, and 197, $K(x)$ comes from Stokes' phenomenon of $\text{erfc}(z)$'s different asymptotic behaviors on different domains, which has a closed-form formula. At this point, PO integral is expressed in terms of NSDPs with an exponentially decay integrand, which could be evaluated in a frequency-independent workload.

Compared to the HFA method, one can see the advantage of the NSDP method after studying the high-frequency critical-point contributions from both methods. With the ray physics aspect, the high-frequency critical-point contributions from the stationary phase point, resonance point, and vertex point correspond to the ray fields from the specular reflection point, the boundary specular reflection point, and the vertex point, respectively. The leading terms of the HFA approach for the high-frequency critical-point

contributions including the stationary phase point, resonance point, and vertex point (Carluccio et al. 2010) are of $O(k^{-1})$, $O(k^{-\frac{3}{2}})$, $O(k^{-2})$, respectively. On the other hand, these critical-point contributions could also be rigorously expressed via the PO integrand defined on the NSDPs. Since the process for evaluating the PO integral by the NSDP method uses less approximation, these PO integral results gain high accuracy from those produced by the HFA method. Meanwhile, the frequency-independent computation effort is achieved (Wu et al. 2012a, b, 2013a, b, c).

Conclusions

In this chapter, an overview of different computational electromagnetics (CEM) methods is given that are used to solve antenna problems. As computers become more powerful, these methods are more prevalent for antenna designs. Many software tools have been developed to enable better antenna designs. Hence, it is important for antenna designers and engineers to understand the working principles of different CEM methods or future methods that will be incorporated into these software tools. With better understanding, better antennas can be designed.

As antenna engineering often involves circuit design, it is also important that antenna designers understand how electromagnetic theory morphs into circuit theory when the feature sizes are much smaller than the wavelength. Antenna structures are designed to be of the order of wavelength or larger in order to receive the signals that are transmitted as waves. However, as these signals are received by the antennas, they invariably have to be routed through different systems using a combination of circuit and waveguide theories. It is instrumental that antenna engineers understand both the electromagnetic and circuit theory well.

As the signal is received, the signal is routed through a system where wave physics becomes circuit physics. Hence, it is also important that computation be done correctly in the “twilight” zone between the circuit theory and electromagnetic full wave theory.

Due to the advanced made in nanofabrication technology, it is now possible to make nanoantennas that are subwavelength even in the optical regime. These nanoantennas find applications in optical links in computer systems as front-end communication bus. They can provide the high data rate needed. They also can be used to enhance the optical emissions given by atoms and molecules or couple light energy to materials to enhance the generation of photoelectricity.

Many antenna designers use software to gain physical insight into the working of the antennas. It will be increasingly important that this software provides output parameters that offer a deeper insight into the workings of an antenna. This can be seen in the increasing use of characteristic mode analysis for antenna design. The future can be the use of natural mode analysis for better understanding and design of antennas.

Also, as systems become highly complex, it is prudent to decompose a larger system as a sum of smaller systems for ease of finding the solution. So the domain decomposition method comes in useful. The equivalence principle algorithm (EPA) is useful to achieve this decomposition, even though much research work is still needed.

As a final note, high-frequency scattering has always been important in design of reflector antennas. The improvement of the accuracy and speed of existing high-frequency techniques is an area that warrants further research.

General research directions are new numerical analysis tools for better antenna design and engineering. Fast CEM algorithms, robust and valid over a broad frequency range, are future goals. These future tools should handle multi-scale structures efficiently and robustly, with the least use of memory and computer resource. Moreover, coupling to multiphysics simulation will become increasingly important as antennas

are integrated into complex environments. New metrics for antenna design and performance need to be formulated as the complexity becomes higher.

References

- Adams RJ (2004) Physical and analytical properties of a stabilized electric field integral equation. *IEEE Trans Antennas Propag* 52(2):362–372
- Alu A, Engheta N (2008) Tuning the scattering response of optical nanoantennas with nanocircuit loads. *Nat Photonics* 2:307–310
- Andreani LC, Panzarini G, Gerard JM (1999) Strong-coupling regime for quantum boxes in pillar microcavities: theory. *Phys Rev B* 60:13276–13279
- Andriulli FP (2012) Loop-star and loop-tree decompositions: analysis and efficient algorithms. *IEEE Trans Antennas Propag* 60(5):2347–2356
- Andriulli FP, Cools K, Bađci H, Olyslager F, Buffa A, Christiansen S, Michielssen E (2008) A multiplicative Calderón preconditioner for the electric field integral equation. *IEEE Trans Antennas Propag* 56(8):2398–2412
- Andriulli FP, Bađci H, Vipiana F, Vecchi G, Michielssen E (2009) Analysis and regularization of the TD-EFIE low-frequency breakdown. *IEEE Trans Antennas Propag* 57(7):2034–2046
- Atwater HA, Polman A (2010) Plasmonics for improved photovoltaic devices. *Nat Mater* 9:205–213
- Axelsson O, Barker VA (1984) Finite element solution of boundary value problems: theory and computation. Academic Press, New York
- Bađci H, Andriulli FP, Cools K, Olyslager F, Michielssen E (2009) A Calderón multiplicative preconditioner for the combined field integral equation. *IEEE Trans Antennas Propag* 57(10):3387–3392
- Balanis CA (2012a) Advanced engineering electromagnetics, 2nd edn. Wiley, New York
- Balanis CA (2012b) Antenna theory: analysis and design. Wiley, New York
- Baum C, Rothwell E, Chen K-M, Nyquist D (1991) The singularity expansion method and its application to target identification. *Proc IEEE* 79:1481–1492
- Benzi M, Golub GH, Liesen J (2005) Numerical solution of saddle point problems. *Acta Numer* 14:1–137
- Berenger JP (1994) A perfectly matched layer for the absorption of electromagnetic waves. *J Comput Phys* 114:185–200
- Berthelot J, Bouhelier A, Huang CJ et al (2009) Tuning of an optical dimer nanoantenna by electrically controlling its load impedance. *Nano Lett* 9:3914–3921
- Bharadwaj P, Deutsch B, Novotny L (2009) Optical antennas. *Adv Opt Photon* 1:438–483
- Biagioni P, Huang JS, Hecht B (2012) Nanoantennas for visible and infrared radiation. *Rep Prog Phys* 75:024402
- Bienstman P, Baets R (2001) Optical modelling of photonic crystals and VCSELs using eigenmode expansion and perfectly matched layers. *Opt Quant Electron* 33:327–341
- Bliokh KY, Bekshaev AY, Nori F (2013) Dual electromagnetism: helicity, spin, momentum and angular momentum. *New J Phys* 15:033026
- Bonakdar A, Kohoutek J, Dey D, Mohseni H (2012) Optomechanical nanoantenna. *Opt Lett* 37:3258–3260
- Borup DT, Gandhi OP (1984) Fast-Fourier transform method for calculation of SAR distributions in finely discretized inhomogeneous models of biological bodies. *IEEE Trans Microw Theory Tech* 32(4):355–360

- Buffa A, Christiansen S (2007) A dual finite element complex on the barycentric refinement. *Math Comput* 76(260):1743–1769
- Burke GJ, Poggio AJ (1981) “Numerical Electromagnetic Code-2,” Ver. 5.7.5, Arie Voors
- Burkholder RJ, Lee TH (2005) Adaptive sampling for fast physical optics numerical integration. *IEEE Trans Antennas Propag* 53(5):1843–1845
- Cabedo-Fabres M, Antonino-Daviu E, Valero-Nogueira A, Bataller MF (2007) The theory of characteristic modes revisited: a contribution to the design of antennas for modern applications. *IEEE Antennas Propag Mag* 49:52–68
- Carluccio G, Albani M, Pathak PH (2010) Uniform asymptotic evaluation of surface integrals with polygonal integration domains in terms of UTD transition functions. *IEEE Trans Antennas Propag* 58(4):1155–1163
- Catedra MF, Gago E, Nuno L (1989) A numerical scheme to obtain the RCS of three-dimensional bodies of resonant size using the conjugate gradient method and the fast Fourier transform. *IEEE Trans Antennas Propag* 37:528–537
- Celik M, Pileggi L, Odabasioglu A (2002) *IC interconnect analysis*. Kluwer Academic, Dordrecht
- Chen PY, Alu A (2010) Optical nanoantenna arrays loaded with nonlinear materials. *Phys Rev B* 82:235405
- Chen CH, Lien C-D (1980) The variational principle for non-self-adjoint electromagnetic problems. *IEEE Trans Microw Theory Tech* 28:878–886
- Chen QL, Wilton DR (1990) Electromagnetic scattering by three-dimensional arbitrary complex material/conducting bodies. *IEEE Int Symp Antennas Propag* 2:590–593
- Chen LZ, Sha WEI, Choy WCH (2012a) Light harvesting improvement of organic solar cells with self-enhanced active layer designs. *Opt Express* 20:8175–8185
- Chen YP, Sha WEI, Choy WCH, Jiang LJ, Chew WC (2012b) Study on spontaneous emission in complex multilayered plasmonic system via surface integral equation approach with layered medium Green’s function. *Opt Express* 20:20210–20221
- Chew WC (1990) *Waves and fields in inhomogeneous media*. Van Nostrand Reinhold, New York, Reprinted by Piscataway, NJ: IEEE Press, 1995
- Chew WC, Kong JA (1981) Asymptotic formula for the resonant frequencies of a circular microstrip antenna. *J Appl Phys* 52(8):5365–5369
- Chew WC, Lu CC (1993) The use of Huygens equivalence principle for solving the volume integral equation of scattering. *IEEE Trans Antennas Propag* 41(7):897904
- Chew WC, Weedon WH (1994) A 3-D perfectly matched medium from modified Maxwell’s equations with stretched coordinates. *Microw Opt Technol Lett* 7(13):599–604
- Chew WC, Jin JM, Michielssen E, Song J (2000) *Fast and efficient algorithms in computational electromagnetics*. Artech House, London
- Chew WC, Jin JM, Michielssen E, Song JM (eds) (2001) *Fast and efficient algorithms in computational electromagnetics*. Artech House, Boston
- Chew WC, Tong MS, Hu B (2008) *Integral equation methods for electromagnetic and elastic waves*. Morgan Claypool, San Francisco
- Chew WC, Tong MS, Hu B (2009) *Integral equations methods for electromagnetic and elastic waves*. Morgan & Claypool, San Rafael
- Chou HT, Ho HK, Chung TY (2005) A discrete-time uniform geometrical theory of diffraction for the fast transient analysis of scattering from curved wedges. *IEEE Trans Antennas Propag* 53(11):3633–3643
- Clark B (1985) Well logging apparatus and method using transverse magnetic mode. U.S. Patent 4,553,097

- Coifman R, Rokhlin V, Wandzura S (1993) The fast multipole method for the wave equation: a pedestrian prescription. *IEEE Antennas Propag Mag* 35(3):7–12
- Collin RE (1991) *Field theory of guided waves*. IEEE Press, Piscataway
- Collins RE (1966) *Foundations for microwave engineering*. McGraw-Hill, New York
- Conde OM, Pérez J, Cátedra MF (2001) Stationary phase method application for the analysis of radiation of complex 3-D conducting structures. *IEEE Trans Antennas Propag* 49(5):724–731
- Cools K, Andriulli FP, Olyslager F, Michielssen E (2009) Improving the MFIE's accuracy by using a mixed discretization. In: *IEEE antennas and propagation society international symposium*, pp 1–4, Charleston, SC
- Corvi M, Schaich WL (1986) Hydrodynamic-model calculation of second-harmonic generation at a metal surface. *Phys Rev B* 33:3688
- Cui YX, Xu J, Fung KH et al (2011) A thin film broadband absorber based on multi-sized nanoantennas. *Appl Phys Lett* 99:253101
- Dai QI, Chew WC, Lo YH, Liu YG, Jiang LJ (2012) Generalized modal expansion of electromagnetic field in 2-D bounded and unbounded media. *IEEE Antennas Wirel Propag Lett* 11:1052–1055
- Dai QI, Lo YH, Chew WC, Jiang LJ (2013) An efficiently preconditioned eigenanalysis of inhomogeneously loaded rectangular cavities. *IEEE Antennas Wirel Propag Lett* 12:58–61
- Dai QI, Lo YH, Chew WC, Liu YG, Jiang LJ (2014) Generalized modal expansion and reduced modal representation of 3-D electromagnetic fields. *IEEE Trans Antennas Propag* 62:783–793
- Davis TA, Duff IS (1997) An unsymmetric-pattern multifrontal method for sparse LU factorization. *SIAM J Matrix Anal Appl* 18:140–158
- Devilez A, Stout B, Bonod N (2010) Compact metallo-dielectric optical antenna for ultra directional and enhanced radiative emission. *ACS Nano* 4:3390–3396
- Dong YD, Itoh T (2010) Miniaturized substrate integrated waveguide slot antennas based on negative order resonance. *IEEE Trans Antennas Propag* 58:3856–3864
- Fallahpour M, Ma Z, Li MK, Chew WC (2014) Using equivalence principle algorithm to analyze and design reconfigurable pixelled antennas. In: *Allerton antenna applications symposium*, IL, USA
- Fock VA (1946) The distributions of currents induced by a plane wave on the surface of a conductor. *J Phys* 10:130–136
- Garbacz RJ, Turpin RH (1971) A generalized expansion for radiated and scattered fields. *IEEE Trans Antennas Propag* 19:348–358
- Gerard D, Devilez A, Aouani H et al (2009) Efficient excitation and collection of single-molecule fluorescence close to a dielectric microsphere. *J Opt Soc Am B* 26:1473–1478
- Giannini V, Fernandez-Dominguez AI, Heck SC, Maier SA (2011) Plasmonic nanoantennas: fundamentals and their use in controlling the radiative properties of nanoemitters. *Chem Rev* 111:3888–3912
- Graglia RD (1993) On the numerical integration of the linear shape functions times the 3D Green's function or its gradient on a plane triangle. *IEEE Trans Antennas Propag* 41(10):1448–1456
- Greengard L, Rokhlin V (1987) A fast algorithm for particle simulations. *J Comput Phys* 73:325–348
- Hansen RC (ed) (1981) *Geometric theory of diffraction*. IEEE Press, Piscataway
- Hansen TB, Shore RA (1998) Incremental length diffraction coefficients for the shadow boundary of a convex cylinder. *IEEE Trans Antennas Propag* 46(10):1458–1466
- Harrington RF, Mautz JR (1971) Theory of characteristic modes for conducting bodies. *IEEE Trans Antennas Propag* 19:622–628
- He SQ, Sha WEI, Jiang LJ, Choy WCH, Chew WC, Nie ZP (2012) Finite-element-based generalized impedance boundary condition for modeling plasmonic nanostructures. *IEEE Trans Nanotechnol* 11:336–345

- Heaviside O (1888) On electromagnetic waves, especially in relation to the vorticity of the impressed forces, and the forced vibration of electromagnetic systems. *Philos Mag* 25:130–156. Also see Nahin PJ (1990) Oliver Heaviside. *Scientific American* 122–129
- Ho C, Ruehli A, Brennan P (1975) The modified nodal approach to network analysis. *IEEE Trans Circ Syst CAS-32*:504–509
- Hsiao GC, Kleinman RE (1997) Mathematical foundations for error estimation in numerical solutions of integral equations in electromagnetics. *IEEE Trans Antennas Propag* 45(3):316–328
- Jin JM (2002) *The finite element method in electromagnetics*, 2nd edn. Wiley-IEEE Press, New York
- Jin JM (2014) *The finite element method in electromagnetics*, 3rd edn. Wiley-IEEE Press, Hoboken
- Joannopoulos JD, Johnson SG, Winn JN, Meade RD (2008) *Photonic crystals: molding the flow of light*. Princeton University Press, Princeton
- Johnson SG, Joannopoulos JD (2001) Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt Express* 8:173–190
- Kalkbrenner T, Hakanson U, Schadle A et al (2005) Optical microscopy via spectral modifications of a nanoantenna. *Phys Rev Lett* 95:200801
- Kauranen M, Zayats AV (2012) Nonlinear plasmonics. *Nat Photonics* 6:737–748
- Keller JB (1962) Geometrical theory of diffraction. *JOSA* 52(2):116–130
- Kelley DF, Luebbers RJ (1996) Piecewise linear recursive convolution for dispersive media using FDTD. *IEEE Trans Antennas Propag* 44:792–797
- King RWP (1956) *The theory of linear antennas*. Harvard University Press, Cambridge, MA
- Kinkhabwala A, Yu ZF, Fan SH et al (2009) Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. *Nat Photonics* 3:654–657
- Kouyoumjian RG, Pathak PH (1974) A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface. *Proc IEEE* 62(11):1448–1461
- Krasnok AE, Maksymov IS, Denisyuk AI et al (2013) Optical nanoantennas. *Phys Usp* 56:539–564
- Kwon DH, Werner PL, Werner DH (2008) Optical planar chiral metamaterial designs for strong circular dichroism and polarization rotation. *Opt Express* 16:11802–11807
- Lee R, Cangellaris AC (1992) A study of discretization error in the finite element approximation of wave solution. *IEEE Trans Antennas Propag* 40(5):542–549
- Lee SW, Deschamps GA (1976) A uniform asymptotic theory of electromagnetic diffraction by a curved wedge. *IEEE Trans Antennas Propag* 24(1):25–34
- Lee SW, Mittra R (1983) Fourier transform of a polygonal shape function and its application in electromagnetics. *IEEE Trans Antennas Propag* 31(1):99–103
- Lee JF, Lee R, Burkholder RJ (2003) Loop star basis functions and a robust preconditioner for EFIE scattering problems. *Trans Antennas Propag* 51(8):1855–1863
- Lehoucq R, Sorensen D (1996) Deflation techniques for an implicitly restarted Arnoldi method. *SIAM J Matrix Anal Appl* 17:789–821
- Li M-K, Chew WC (2007) Wave-field interaction with complex structures using equivalence principle algorithm. *IEEE Trans Antennas Propag* 55(1):130–138
- Li M-K, Chew WC (2008) Multiscale simulation of complex structures using equivalence principle algorithm with high-order field point sampling scheme. *IEEE Trans Antennas Propag* 56(8):2389–2397
- Ling H, Chou RC, Lee SW (1989) Shooting and bouncing rays: calculating the RCS of an arbitrarily shaped cavity. *IEEE Trans Antennas Propag* 37(2):194–205
- Liu YG, Li Y, Sha WEI (2011) Directional far-field response of a spherical nanoantenna. *Opt Lett* 36:2146–2148

- Liu YG, Choy WCH, Sha WEI, Chew WC (2012) Unidirectional and wavelength-selective photonic sphere-array nanoantennas. *Opt Lett* 37:2112–2114
- Liu QS, Sun S, Chew WC (2014) Convergence of low-frequency EFIE-based systems with weighted right-hand-side effect. *IEEE Trans Antennas Propag* 62(10):5108–5116
- Lo YT, Solomon D, Richards W (1979) Theory and experiment on microstrip antennas. *Antennas Propag IEEE Trans* 27(2):137–145
- Lo YH, Jiang LJ, Chew WC (2013) Finite-width feed and load models. *IEEE Trans Antennas Propag* 61(1):281–289
- Lu CC, Chew WC (1995) The use of Huygens equivalence principle for solving 3-D volume integral equation of scattering. *IEEE Trans Antennas Propag* 43(5):500–507
- Lu Z-Q, An X, Hong W (2008) A fast domain decomposition method for solving three dimensional large-scale electromagnetic problems. *IEEE Trans Antennas Propag* 56(8):2200–2210
- Luebbers R, Hunsberger FP, Kunz KS et al (1990) A frequency-dependent finite-difference time-domain formulation for dispersive materials. *IEEE Trans Electromagn Compat* 32:222–227
- Macdonald HM (1913) The effect produced by an obstacle on a train of electric waves. *Philos Trans R Soc Lond A Math Phys Sci* 212:299–337
- Maier SA (2007) *Plasmonics: fundamentals and applications*. Springer, New York.
- Makitalo J, Suuriniemi S, Kauranen M (2011) Boundary element method for surface nonlinear optics of nanoparticles. *Opt Express* 19:23386–23399
- Maksymov IS, Davoyan AR, Kivshar YS (2011) Enhanced emission and light control with tapered plasmonic nanoantennas. *Appl Phys Lett* 99:083304
- Manges JB, Cendes ZJ (1995) A generalized tree-cotree gauge for magnetic field computation. *IEEE Trans Magn* 31(3):1342–1347
- Martin-Moreno L, Garcia-Vidal FJ, Lezec HJ, Degiron A, Ebbesen TW (2003) Theory of highly directional emission from a single subwavelength aperture surrounded by surface corrugations. *Phys Rev Lett* 90:167401
- Maxwell JC (1865) A dynamical theory of the electromagnetic field. *Phil Trans R Soc Lond* 155:459–512
- Merlein J, Kahl M, Zuschlag A et al (2008) Nanomechanical control of an optical antenna. *Nat Photonics* 2:230–233
- Michiels B, Bogaert I, Fostier J, Zutter DD (2014) A well-scaling parallel algorithm for the computation of the translation operator in the MLFMA. *IEEE Trans Antennas Propag* 62(5):2679–2687
- Mishchenko MI, Travis LD, Mackowski DW (2010) T-Matrix computations of light scattering by nonspherical particles: a review. *J Quant Spectrosc Radiat Transf* 111:1704–1744
- Moharam MG, Grann EB, Pommet DA, Gaylord TK (1995) Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings. *J Opt Soc Am A* 12:1068–1076
- Muhlschlegel P, Eisler HJ, Martin OJF et al (2005) Resonant optical antennas. *Science* 308:1607–1609
- Nagel LW, Pederson DO (1973) SPICE (Simulation Program with Integrated Circuit Emphasis). Memorandum No. ERL-M382, University of California, Berkeley
- Novotny L, Hecht B (2006) *Principles of nano-optics*. Cambridge University Press, New York
- Novotny L, Hulst NV (2011) Antennas for light. *Nat Photonics* 5:83–90
- Pellegrini G, Mattei G, Mazzoldi P (2009) Light extraction with dielectric nanoantenna arrays. *ACS Nano* 3:2715–2721
- Peng Z, Wang XC, Lee JF (2011) Integral equation based domain decomposition method for solving electromagnetic wave scattering from non-penetrable objects. *IEEE Trans Antennas Propag* 59(9):3328–3338

- Pillegi L, Rohrer R, Visweswariah C (1995) *Electronic circuits and system simulation methods*. McGraw-Hill, New York
- Pozar DM (2011) *Microwave engineering*, 4th edn. Wiley, New York
- Pu Y, Grange R, Hsieh CL, Psaltis D (2010) Nonlinear optical properties of core-shell nanocavities for enhanced second-harmonic generation. *Phys Rev Lett* 104:207402
- Qian ZG, Chew WC (2008) An augmented electric field integral equation for low frequency electromagnetics analysis. In: *IEEE international symposium on antennas and propagation*, San Diego
- Qian Z-G, Chew WC (2009) Fast full-wave surface integral equation solver for multiscale structure modeling. *IEEE Trans Antennas Propag* 57(11):3594–3601
- Qian ZG, Chew WC (2010) Enhanced A-EFIE with perturbation method. *IEEE Trans Antennas Propag* 58(10):3256–3264
- Qiao PF, Sha WEI, Choy WCH, Chew WC (2011) Systematic study of spontaneous emission in a two-dimensional arbitrary inhomogeneous environment. *Phys Rev A* 83:043824
- Rahmat-Samii Y, Galindo-Israel V (1980) Shaped reflector antenna analysis using the Jacobi–Bessel series. *IEEE Trans Antennas Propag* 28(4):427–432
- Rao SM, Wilton GR, Glisson AW (1982) Electromagnetic scattering by surfaces of arbitrary shape. *IEEE Trans Antennas Propag* 30(3):409–418
- Ren XG, Sha WEI, Choy WCH (2013) Tuning optical responses of metallic dipole nanoantenna using graphene. *Opt Express* 21:31824–31829
- Rokhlin V (1990) Rapid solution of integral equations of scattering theory in two dimensions. *J Comput Phys* 86(2):414–439
- Rong A, Cangellaris AC (2001) Generalized PEEC models for three-dimensional interconnect structures and integrated passives of arbitrary shapes. *Digest Electr Perf Electron Packag*, Boston, MA 10:225–228
- Ruehli AE (1972) Inductance calculations in a complex integrated circuit environment. *IBM J Res Dev* 16(5):470–481
- Ruehli AE (1974) Equivalent circuit models for three dimensional multiconductor systems. *IEEE Trans Microw Theory Tech* MTT-22(3):216–221
- Ruehli AE, Brennan PA (1973) Efficient capacitance calculations for three-dimensional multiconductor systems. *IEEE Trans Microw Theory Tech* 21(2):76–82
- Ruehli AE, Antonini G, Esch J, Ekman J, Mayo A, Orlandi A (2003) Non-orthogonal PEEC formulation for time and frequency domain EM and circuit modeling. *IEEE Trans Electromagn Compat* 45(2):167–176
- Ruehli A, Antonio G, Jiang LJ (2013) Skin-effect loss models for time- and frequency-domain PEEC solver. *Proc IEEE* 101(2):451–472
- Sacks ZS, Kingsland DM, Lee R, Lee JF (1995) A perfectly matched anisotropic absorber for use as an absorbing boundary condition. *IEEE Trans Antennas Propag* 43(12):1460–1463
- Sha WEI, Meng LL, Choy WCH, Chew WC (2014) Observing abnormally large group velocity at the plasmonic band edge via a universal eigenvalue analysis. *Opt Lett* 39:158–161
- Shao H, Hu J, Nie ZP, Han G, He S (2011) Hybrid tangential equivalence principle algorithm with MLFMA for analysis of array structures. *Prog Electromagn Res* 113:127–141
- Shao H, Hu J, Guo H, Ye F, Lu W, Nie Z (2012) Fast simulation of array structures using T-EPA with hierarchical LU decomposition. *IEEE Antennas Wirel Propag Lett* 11:1556–1559
- Shao H, Hu J, Chew WC (2013) Solving array structures using single-source equivalence principle algorithm. In: *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, Torino, pp 774–777

- Shao H, Hu J, Nie Z, Jiang L (2014) Simulation of multiscale structures using equivalence principle algorithm with grid-robust higher order vector basis. *J Electromagn Waves Appl* 28(11):1333–1346. doi:10.1080/09205071.2014.921123
- Sommerfeld A (1896) *Mathematische Theorie der Diffraction*. *Math Ann* 47(s319):317–374
- Song JM, Lu CC, Chew WC (1997) Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects. *IEEE Trans Antennas Propag* 45:1488–1493
- Staude I, Miroshnichenko AE, Decker M et al (2013) Tailoring directional scattering through magnetic and electric resonances in subwavelength silicon nanodisks. *ACS Nano* 7:7824–7832
- Sun S, Jiang LJ, Chew WC (2013a) Enhanced A-EFIE with Calderón multiplicative preconditioner. In: *Symposium on antennas and propagation (IEEE APS13)*, Orlando, FL
- Sun S, Liu YG, Chew WC, Ma ZH (2013b) Calderón multiplicative preconditioned EFIE with perturbation method. *IEEE Trans Antennas Propag* 61(1):247–255
- Taflove A (1995) *Computational electrodynamics: the finite difference time domain method*. Artech House, Boston
- Taflove A, Hagness SC (2005) *Computational electrodynamics: the finite-difference time-domain method*, 3rd edn. Artech House, Boston
- Taminiau TH, Stefani FD, Hulst NFV (2008) Enhanced directional excitation and emission of single emitters by a nano-optical Yagi-Uda antenna. *Opt Express* 16:10858–10866
- Tong MS, Chew WC, Rubin BJ, Morsey JD, Jiang L (2009) On the dual basis for solving electromagnetic surface integral equations. *IEEE Trans Antennas Propag* 57(10):3136–3146
- Trefethen LN, Bau D (1997) *Numerical Linear Algebra*, No. 50, SIAM
- Tsang L, Kong JA, Ding KH (2000) *Scattering of electromagnetic waves: theories and applications*. Wiley, New York
- Ufimtsev PY (2005) *Backscatter*. Wiley, New York
- Ufimtsev PY (2008) New insight into the classical macdonald physical optics approximation. *IEEE Antennas Propag Mag* 50:11–20
- Valdés F, Andriulli FP, Bağcı H, Michielssen E (2011) A Calderón-preconditioned single source combined field integral equation for analyzing scattering from homogeneous penetrable objects. *IEEE Trans Antennas Propag* 59(6):2315–2328
- Van Bladel J (1985) *Electromagnetic fields*. Hemisphere, New York
- Vandervorst HA (1992) Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM J Sci Stat Comput* 13:631–644
- Vecchi G (1999) Loop-star decomposition of basis functions in the discretization of EFIE. *IEEE Trans Antennas Propag* 47(2):339–346
- Veysoglu ME, Shin RT, Kong JA (1993) A finite-difference time-domain analysis of wave scattering from periodic surfaces-oblique-incidence case. *J Electromagn Waves Appl* 7:1595–1607
- Vico-Bondia F, Ferrando-Bataller M, Valero-Nogueira A (2010) A new fast physical optics for smooth surfaces by means of a numerical theory of diffraction. *IEEE Trans Antennas Propag* 58(3):773–789
- Volakis JL, Chatterjee A, Kempel LC (1998) *Finite element method electromagnetics: antennas, microwave circuits, and scattering applications*. Wiley-IEEE Press, New York
- Wagner RL, Otto GP, Chew WC (1993) Fast waveguide mode computation using wavelet-like basis functions. *IEEE Microw Guid Wave Lett* 3(7):208–210
- Warwick C (2009) Everything you always wanted to know about SPICE* (*But were afraid to ask). *EMC J* (Nutwood UK Limited) 82:27–29
- Weinan E (2012) *Principles of multiscale modeling*. Cambridge University Press, Cambridge, UK
- Wu YM, Jiang LJ, Chew WC (2012a) An efficient method for computing highly oscillatory physical optics integral. In: *Symposium on antennas and propagation (IEEE APS12)*, Chicago, IL

- Wu YM, Jiang LJ, Chew WC (2012b) An efficient method for computing highly oscillatory physical optics integral. *Prog Electromagn Res PIER* 127:211–257
- Wu YM, Jiang LJ, Chew WC (2013a) Computing highly oscillatory physical optics integral on the polygonal domain by an efficient numerical steepest descent path method. *J Comput Phys* 236:408–425
- Wu YM, Jiang LJ, Chew WC (2013b) Reducing computational workload of electromagnetic scattered fields from electrically large quadratic surface at high frequency. In: *Symposium on antennas and propagation (IEEE APS13)*, Orlando, FL
- Wu YM, Jiang LJ, Sha WEI, Chew WC (2013c) The numerical steepest descent path method for calculating physical optic integrals on smooth conducting surfaces. *IEEE Trans Antennas Propag* 61(8):4183–4193
- Wu YM, Jiang LJ, Chew WC (2014) The contour deformation method for calculating the high frequency scattered fields by the Fock current on the surface of the 3-D convex cylinder. In: *Symposium on antennas and propagation (IEEE APS14)*
- Xiong XYZ, Jiang LJ, Lo YH, Chew WC (2014) Second-harmonic generation in metal nanoparticles modeling by surface integral equation. In: *2014 I.E. international symposium on antennas and propagation and USNC-URSI radio science meeting*, Memphis
- Yaghjian AD (1984) Equivalence of surface current and aperture field integrations for reflector antennas. *IEEE Trans Antennas Propag* 32(12):1355–1358
- Yaghjian AD, Shore RA, Woodworth MB (1996) Shadow boundary incremental length diffraction coefficients for perfectly conducting smooth, convex surfaces. *Radio Sci* 31(12):1681–1695
- Yagi H, Uda S (2014) Projector of the sharpest beam of electric waves. *Proc Imp Acad Jpn (Imp Acad)* 2(2):49–52. Retrieved Sept 11, 2014
- Yao Y, Kats MA, Genevet P et al (2013) Broad electrical tuning of graphene-loaded plasmonic antennas. *Nano Lett* 13:1257–1264
- Ye YQ, He SL (2010) 90 degrees polarization rotator using a bilayered chiral metamaterial with giant optical activity. *Appl Phys Lett* 96:203501
- Ye YQ, Jin Y, He SL (2010) Omnidirectional, polarization-insensitive and broadband thin absorber in the terahertz regime. *J Opt Soc Am B* 27:498–504
- Yee KS (1966) Numerical solution of initial boundary value problems involving Maxwell's equation in isotropic media. *IEEE Trans Antennas Propag* 14:302–307
- Yonekura J, Ikeda M, Baba T (1999) Analysis of finite 2-D photonic crystals of columns and lightwave devices using the scattering matrix method. *J Light Technol* 17:1500–1508
- Yu ZF, Raman A, Fan SH (2010) Fundamental limit of nanophotonic light trapping in solar cells. *Proc Natl Acad Sci* 107:17491–17496
- Zhang Y, Cui TJ, Chew WC, Zhao JS (2003) Magnetic field integral equation at very low frequencies. *IEEE Trans Antennas Propag* 51(8):1864–1871
- Zhang J, Xu B, Cui TJ (2014) An alternative treatment of saddle stationary phase points in physical optics for smooth surfaces. *IEEE Trans Antennas Propag* 62(2):986–991
- Zhao JS, Chew WC (2000) Integral equation solution of Maxwell's equations from zero frequency to microwave frequency. *IEEE Trans Antennas Propag* 48(10):1635–1645
- Zhao K, Vouvakis M, Lee JF (2004) Application of the multilevel adaptive cross-approximation on ground plane designs. *IEEE Int Symp Electromagn Compat* 1:124–127
- Zhu Y, Cangellaris AC (2006) *Multigrid finite element methods for electromagnetic field modeling*. Wiley-IEEE Press, Hoboken