Computer Modeling

Shimon Panfil

April 30, 2016

1 What is all about?

Computer simulation is a simulation, run on a single computer, or a network of computers, to reproduce behavior of a system. The simulation uses an abstract model (a computer model, or a computational model) to simulate the system. Computer simulations have become a useful part of mathematical modeling of many natural systems in physics (computational physics), astrophysics, climatology, chemistry and biology, human systems in economics, psychology, social science, and engineering. Simulation of a system is represented as the running of the system's model. It can be used to explore and gain new insights into new technology and to estimate the performance of systems too complex for analytical solutions. (Wikipedia)

Computer modeling definition: Constructing and manipulating abstract (mathematical and/or graphical) representations of economic, engineering, manufacturing, social, and other types of situations and natural phenomenon, simulated with the help of a computer system. Also called computer simulation. (http://www.businessdictionary.com/definition/computer-modeling.html)

There are much more definitions and descriptions all of them are pretty useless and circular. However everybody working in high tech industry understand (I hope) what is all about. Basically computer modeling comprises the following hierarchy of the models:

Physical model derived form First Principles (Physical Laws) with the aid of some reasonable assumptions;

Mathematical model describing the physical one;

Numerical model comprising algorithms and software for (numerically) solving mathematical model;

Presentation model comprising UI, methods for information storage and retrieval etc. . .

The boundaries between models in this scheme are rather artificial and not very clear, the strongest interconnections exits between first two models, last two are also interleaved. Different parts in this hierarchy are developed in different times and by different (groups of) people with all usual problems of communications, project management and all that. There is also the class of problems specific for computer modeling (and to scientific programming in general) — we can not easily decide are the results right or wrong and even if know them to be wrong we do not know have the error appeared from programming bug or from bad algorithm or from mathematical model inaccuracy or physical model incompleteness (or some combination of above).

2 Maxwell's equations

Light propagation and scattering is a part of classical electrodynamics and as such it is fully described by Maxwell equations. Let us go all way from Maxwell equation in vacuum to our model of light propagation and scattering and fix all the assumptions we need. In vacuum microscopic Maxwell's equations have the form:

$$
\begin{aligned}
\nabla \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t} &= 0 \\
\nabla \cdot \vec{B} &= 0 \\
\nabla \times \vec{B} - \frac{1}{c}\frac{\partial \vec{E}}{\partial t} &= \frac{4\pi}{c}\vec{j} \\
\nabla \cdot \vec{E} &= 4\pi\rho
\end{aligned}
\tag{1}
$$

here $\vec{E}$, $\vec{B}$ are electric and magnetic fields and $\vec{j}$, ρ current and charge densities. I use physical system of units where all fields have the same dimensions, in engineering system SI these equations and all other formulas look slightly differently.

These equations are nice and exact but refer to any particle in our system not even molecules or atoms but every electron and proton. To be practical

we should use some averaged fields described collective effects. Averaging does not change homogeneous (first and second) equations in the Maxwell's equation set (1) but $\vec{B}$ is replace by $\vec{H}$ and $\vec{E}$ by $\vec{D}$ in second pair. ρ and $\vec{j}$ now describe the charges and currents which are external to the media.

$$
\begin{aligned}
\nabla \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t} &= 0 \\
\nabla \cdot \vec{B} &= 0 \\
\nabla \times \vec{H} - \frac{1}{c}\frac{\partial \vec{D}}{\partial t} &= \frac{4\pi}{c}\vec{j} \\
\nabla \cdot \vec{D} &= 4\pi\rho
\end{aligned}
\tag{2}
$$

Though equations (1) and (2) seem very similar they describe completely different physics. Surely we need also so called constituent equations connecting physical $(\vec{E}, \vec{B})$ and auxiliary fields $(\vec{D}, \vec{H})$. The simplest such equation is

$$
\begin{aligned}
\vec{D} &= \epsilon\vec{E} \\
\vec{j} &= \sigma\vec{E} \\
\vec{B} &= \mu\vec{H}
\end{aligned}
\tag{3}
$$

where ϵ, σ and μ are some constants. We shall consider only electrically neutral ($\rho = 0$) and non-magnetic ($\mu = 1$) media.

Let us recall that the size of the atom is $\sim 0.1nm$ and characteristic distance between atoms in molecules is $\sim 0.1 - 0.2nm$ so that large polymer molecules have the size comparable with 1 nm. This in turn means that validity of simple constituent equations (3) in nanometer scale is not evident and should be checked. Probably the first and second equations in (3) should be replaced by anisotropic ones

$$
\begin{aligned}
D_i &= \epsilon_{jk}E_k \\
j_i &= \sigma_{jk}E_k
\end{aligned}
\tag{4}
$$

where tensors ϵ_{jk} and σ_{jk} might be function of coordinates and time.

We may be not interested in transient effects and consider harmonic time dependence only so any time dependent variable can be written as

$$
V(\vec{r}, t) = v(\vec{r})e^{-i\omega t}
\tag{5}
$$

3

and Maxwell's equations take the form:

$$\nabla \times \vec{E} - \frac{i\omega}{c}\vec{H} = 0$$

$$\nabla \times \vec{H} + \frac{i\omega n^2(\vec{r},\omega)}{c}\vec{E} = 0 \qquad (6)$$

here $n(\vec{r},\omega)$ is complex refraction index which determines optical properties of the media. Note that two remaining equations in (2) are satisfied automatically on solutions of equations (6).

$$n(\vec{r},\omega) = \alpha(\vec{r},\omega) + i\beta(\vec{r},\omega)$$

$$\alpha(\vec{r},\omega) > 0$$

$$\beta(\vec{r},\omega) \geq 0$$

$$\alpha^2(\vec{r},\omega) = \frac{\epsilon(\vec{r},\omega)}{2}\left(1 + \sqrt{1 + \frac{16\pi^2\sigma^2(\vec{r},\omega)}{\omega^2\epsilon^2(\vec{r},\omega)}}\right)$$

$$\beta^2(\vec{r},\omega) = \frac{\epsilon(\vec{r},\omega)}{2}\left(-1 + \sqrt{1 + \frac{16\pi^2\sigma^2(\vec{r},\omega)}{\omega^2\epsilon^2(\vec{r},\omega)}}\right) \qquad (7)$$

It is important also to remember that equations (6) and (7) are obtained with constituent equations (3), on using equations (4) much more complicated expressions appear. Usually optical media are bad conductors so $\alpha \gg \beta$.

To complete the derivation of mathematical model we should add necessary boundary conditions to system (6). This mathematical problem can be represented in other of forms e.g. vector wave equation:

$$\nabla \times \nabla \times \vec{E} - \frac{\omega^2 n^2(\vec{r},\omega)}{c^2}\vec{E} = 0 \qquad (8)$$

which is equivalent to

$$\nabla^2\vec{E} + \frac{\omega^2 n^2(\vec{r},\omega)}{c^2}\vec{E} = 0$$

$$\nabla \cdot \vec{E} = 0 \qquad (9)$$

It can be also represented as the system of integral equations.

Generally this problem is hopelessly difficult to solve[1]. Fortunately in most of real world problems media may be divided into a number of homogeneous regions where refraction index is constant. So equations (6,8 or 9) have known analytical solutions. The only (but serious problem) is to apply correct boundary conditions: tangential component of both fields should be continuous for (6) and tangential component of electric field together with its normal derivative should be continuous for (8, 9). Note that the sharp boundary approximation may not be applicable in nanometer scale[2]. Integral equations are also simplified greatly, all volume integrals disappear, only integral over boundaries survive.

3   Boundary Integrals

Let us describe boundary integrals for arbitrary shaped region with constant refraction index in so called Franz form. Let V be a region in three dimensional space and S be its boundary. Let $\vec{y}$ be a point inside V and $\vec{x}$ be a point on S. Let $\vec{\nu}(\vec{x})$ be internal normal to S. Let

$$\begin{aligned}
\vec{E}_t(\vec{x}) &= \vec{\nu}(\vec{x}) \times \vec{E}(\vec{x}) \\
\vec{H}_t(\vec{x}) &= \vec{\nu}(\vec{x}) \times \vec{H}(\vec{x})
\end{aligned} \tag{10}$$

be tangential component of fields defined on the surface S. It can be proved that these components generate the field

$$\begin{aligned}
\vec{E}(\vec{y}) &= \nabla \times \int_S G(R)\vec{E}_t(\vec{x})ds + \frac{ic}{\omega n^2}\nabla \times \nabla \times \int_S G(R)\vec{H}_t(\vec{x})ds \\
\vec{H}(\vec{y}) &= \nabla \times \int_S G(R)\vec{H}_t(\vec{x})ds - \frac{ic}{\omega}\nabla \times \nabla \times \int_S G(R)\vec{E}_t(\vec{x})ds \tag{11}
\end{aligned}$$

where

$$\begin{aligned}
G(R) &= \frac{e^{i\frac{\omega}{c}nR}}{4\pi R} \\
R &= |\vec{x} - \vec{y}|
\end{aligned} \tag{12}$$

inside the volume V and zero fields outside it.

---

[1] Not only becouse of limited computer power but and mainly becouse we can not obtain necessary data

[2] See discussion of molecular size above

To complete definition of our models we should formulate more accurately what we mean when we are talking about material boundary. Physically more accurate description would be: there is some thin intermediate layer where bulk material properties change their values. We can **assume** that the thickness of these intermediate layers are all much smaller than minimal characteristic dimension of our problem (e.g. light wavelength) and results do not depend on what exactly is going inside or near intermediate layers. However this assumption only tells us nothing about the shape and position of the boundary: e.g. is it plane, curved , rough etc... *Boundary properties are fixed by additional assumptions and their correctness should be checked.* Putting this in other words it is "effective boundary" and its location and shape need not be exactly same as that of geometric boundary which is not defined at nanometer scale. Note that sharp angles and other non-smooth parts of boundary lead to the field singularities and bad convergence.
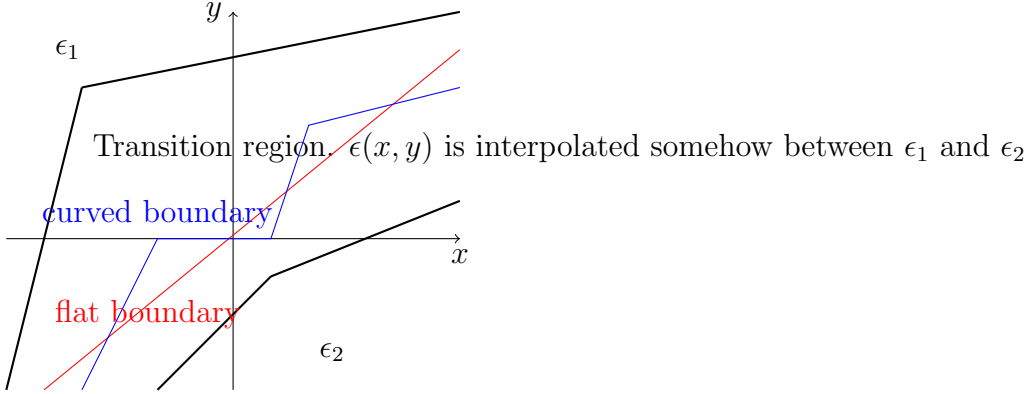

Figure 1: Transition layer and effective boundary

4   Algorithms and Methods

Let us suppose that me managed to divide our volume of interest into a number of parts with constant refraction indices and the condition that allow to neglect transition layers are fulfilled. We can use quite a few methods to solve the equations (6) or (11). Let us consider them briefly.

6

4.1 RCWA

Rigorous coupled-wave analysis (RCWA) is a semi-analytical method in computational electromagnetics that is most typically applied to solve scattering from periodic dielectric structures. It is a Fourier-space method so devices and fields are represented as a sum of spatial harmonics. The method is based on the Floquet's theorem that the solutions of periodic differential equations can be expanded with Floquet functions (or sometimes referred as Bloch wave, especially in solid-state physics community). A device is divided into layers that are each uniform in the z direction. A staircase approximation is needed for curved devices with properties such as dielectric permittivity graded along the z-direction. The electromagnetic modes in each layer are calculated and analytically propagated through the layers. The overall problem is solved by matching boundary conditions at each of the interfaces between the layers using a technique like scattering matrices. To solve for the electromagnetic modes, which are decided by the wave vector of the incident plane wave, in periodic dielectric medium, the Maxwell's equations (in partial differential form) as well as the boundary conditions are expanded by the Floquet functions and turned into infinitely large algebraic equations. With the cutting off of higher order Floquet functions, depending on the accuracy and convergence speed one needs, the infinitely large algebraic equations become finite and thus solvable by computers.

Being a Fourier-space method it suffers several drawbacks. Gibbs phenomenon is particularly severe for devices with high dielectric contrast. Truncating the number of spatial harmonics can also slow convergence and techniques like fast Fourier factorization (FFF) should be used. FFF is straightforward to implement for 1D gratings, but the community is still working on a straightforward approach for crossed grating devices. The difficulty with FFF in crossed grating devices is that the field must be decomposed into parallel and perpendicular components at all of the interfaces. This is not a straightforward calculation for arbitrarily shaped devices.

Boundary conditions must be enforced at the interfaces be-

tween all the layers. When many layers are used, this becomes too large to solve simultaneously. Instead, we borrow from network theory and calculate scattering matrices. This lets us solve the boundary conditions one layer at a time.

This long quotation from Wikipedia describes the method well enough. I would like only add some comments. All interfaces should be exactly plane and parallel and all structures should be exactly periodical this is both advantage — no need to tune the method for particular case and disadvantage also — we can not model even small deviations. Staircase approximation may produce severe artifacts I will not explain all the theory behind but show simple example . Let us make staircase approximation for the hypotenuse of right triangle, see fig. (2), if we compute the area of triangle all is fine, but on computing the length of hypotenuse we will get that it is sum of catheti!
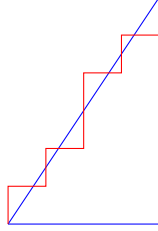

Figure 2: Right triangle

4.2 FDFD

Finite Differences Frequency Domain (FDFD) method is based on finite difference approximation of differential equations (6) or (8) [3]. So the equation at hands takes the form $Ax = b$, where matrix A is large but sparse. Feasibility and efficiency of the method really are that of software for working with sparse matrices. Good open source (e.g. SuperLU http://crd-legacy.lbl.gov/ xiaoye/SuperLU/)and commercial packages do exist.

4.3 DDA

The discrete dipole approximation (DDA) is a method for computing scattering of radiation by particles of arbitrary shape and by periodic structures.

---

[3]Actually any differential equation

Simply stated, the DDA is an approximation of the continuum target by a finite array of polarizable points. The points acquire dipole moments in response to the local electric field. For a finite array of point dipoles the scattering problem may be solved exactly, so the only approximation that is present in the DDA is the replacement of the continuum target by an array of N-point dipoles. The replacement requires specification of both the geometry (location of the dipoles) and the dipole polarizabilities. For monochromatic incident waves the self-consistent solution for the oscillating dipole moments may be found; from these the absorption and scattering cross sections are computed. If DDA solutions are obtained for two independent polarizations of the incident wave, then the complete amplitude scattering matrix can be determined. DDA may be considered as particular sort of Finite Volume Element Method. Good open source exists (http://www.ddscat.org/)

4.4   BEM

Boundary Element Method (BEM) is simply discretization of surface integral equations (11) or similar ones. Matrices which appear in BEM are considerably smaller than that in FDFD or FEM but they are not sparse. Well known software packages e.g. LAPACK exist for such matrices.

5   Physical Optics

In previous sections we consider optics as particular case of electromagnetic problem, nothing special was used. It is well known however that for most practical problems the solution of Maxwell's equation is not feasible and some kind of physically motivated approximation is needed. The simplest and most used approximation is geometrical optics where wave properties of light are neglected completely. Physical optics is next in line, it takes wave properties in consideration only when it is necessary: e.g. when the distance from the boundary is of order of wave length (diffraction) or we trying to see something which has size comparable with wave length. Mathematically speaking physical optics is short wave length approximation to Maxwell's equations. The most natural form of such approximation is so called eikonal equations.