Nuclear Instruments and Methods in Physics Research A 645 (2011) 273–277

Contents lists available at ScienceDirect

Nuclear Instruments and Methods in Physics Research A

journal homepage: www.elsevier.com/locate/nima

The CPO programs and the BEM for charged particle optics

Frank H. Read [a,b,*], Nicholas J. Bowring [a,c]

[a] CPO Ltd., Cumberland House, 9 Jordangate, Macclesfield SK10 1EE, UK
[b] School of Physics and Astronomy, University of Manchester, Manchester, UK
[c] Department of Engineering and Technology, Manchester Metropolitan University, Manchester, UK

ARTICLE INFO

Available online 28 December 2010

Keywords:
BEM
Charged particle optics
Computer simulations
Space-charge systems

ABSTRACT

The Boundary Element Method for solving the Laplace and Poisson equations for electrostatic systems will be outlined, with the emphasis on 3D systems and the commercial CPO programs. Some applications to charged particle optics will be described. Since the BEM is a charge-based method it is ideally suited for systems that include space-charge and/or cathodes. It is also exceptionally accurate and can deal easily with electrodes of very different sizes. Several 'benchmark tests will be presented, in which systems with known analytic solutions are used to illustrate the accuracy and versatility of the programs.

1. Introduction and historical context

The Boundary Element Method (BEM), also known as the Integral Equation Method, has been used in many fields, including those of electrostatics and charged particle optics (where it has also been called the Charge Density Method and the Surface Charge Method). In this review we shall concentrate on the use of the BEM in the CPO programs of CPO Ltd. [1] but will start with a brief description of the historical context and will include a brief review of other applications of the BEM to charged particle optics.

The first application of the BEM to electrostatics seems to be given by Maxwell [2,3] in 1878, who used it to find the capacitance of a thin square plate with sides of unit length. He subdivided the plate into 36 square segments that hold 6 independent charges and manually adjusted those charges to try to obtain a uniform unit potential at the centers of all the segments, using Coulomb's Law. The main approximation that he made was that each segment has a uniform charge density but nevertheless he obtained a value that is accurate to 0.8% according to the latest BEM calculations [4] (and interestingly there is still no analytical solution).

The first use of the BEM in charged particle optics was given by the Manchester group, who used it to calculate the focal properties of two-tube electrostatic lenses [5]. This early work culminated in a standard work on electrostatic lenses [6] and has continued since then.

2. Outline of the BEM

The BEM, as applied to electrostatic systems, is based on the fact that when voltages are applied to conducting electrodes, charges appear on the surfaces of those electrodes. These charges are the only sources of all the potentials and fields in the system. So in the BEM the surfaces of the electrodes are effectively replaced by these sheets of surface charge. Only the surfaces have to be modeled and it is not necessary to create a grid of points in the space enclosed by the electrodes, nor is it necessary to enclose a system.

In the version of the BEM used in the CPO programs 3D electrode surfaces are divided into flat rectangular or triangular segments, each of which has a uniformly distributed charge over its surface. The flatness of the segments and the uniformity of their charge densities represent the only significant approximations (but see below the recent use of non-uniform charge distributions). As with Maxwell, the segment charges are adjusted to make the potentials at the centers of the segments equal to the voltages applied to the parent electrodes (that is, Dirichlet boundary conditions are used – dielectrics and Neumann boundary conditions will be dealt with below). This adjustment is carried out computationally by setting up the matrix that connects the mid-center potential of each segment to the charges on all the segments and then inverting the matrix to obtain the charges that exist for the given set of applied voltages. Potentials and fields can then be calculated anywhere in space.

The errors in setting up the matrix are essentially negligible because the potential distributions of uniformly charged 3D triangles and rectangles are known analytically [7], including at the surface, and also the potential distributions of 2D segments with planar or rotational symmetry can be calculated very accurately [8]. Similarly there is essentially negligible error in inverting the matrix. The errors in calculating the final potentials, which are continuous, are also essentially negligible while the corresponding fields are obtained by numerical differentiation when necessary. The errors due the flatness and charge uniformity of the segments remain of course and methods of reducing these are discussed below.

It is not necessary to invert the matrix for each new set of applied voltages. Instead the program pre-calculates the sets of charges that

* Corresponding author at: School of Physics and Astronomy, University of Manchester, Manchester, UK
E-mail address: cpo@electronoptics.com (F.H. Read).

0168-9002/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.nima.2010.12.163

correspond to one of the applied voltages (of which the number is usually much smaller than the number of segments) being unity while all the others are zero. These sets are called the 'unit charges' and the appropriate linear combinations of them can be used for particular sets of applied voltages.

The BEM is mathematically and computationally more complex than the more traditional Finite Difference and Finite Element Methods (FDM and FEM), but the BEM offers several advantages. Perhaps the most important of these is the inherent accuracy of the method; for example it was shown some years ago [9,10] that for non-space-charge electrostatic 3D problems the CPO programs are typically one or two orders of magnitude more accurate than two representative FDM and FEM programs for the same computing time (or typically one or two orders of magnitude faster for the same accuracy), although this comparison has not been carried out for the more recent versions of all these methods. A comparison has also been carried out [11] for the space-charge simulation of a CERN laser ion source, where it was found that the CPO, KOBRA and IGUN codes exhibited good agreement, but no statements were made about the comparative computing times.

Another obvious advantage of the BEM is that it can deal easily with electrodes of very different sizes, such as nano-sized electrodes in centimeter-sized systems.

As a charge-based method another strength of the BEM lies in dealing with systems that include space-charge and/or cathodes, as illustrated below.

A disadvantage is that the BEM tends to integrate trajectories more slowly, but a technique for alleviating this is discussed below. Another slight disadvantage is that the matrix referred to above is fully populated and therefore can take a comparatively long time to invert, although methods have been suggested for alleviating this (see for example [12,13]), which are not used at present in the CPO programs. In fact the maximum time taken to invert the matrix with a 32-bit 2.2 GHz PC is approximately 5 h when the number N of segments is 8000 (the present maximum) and the dependence on N is approximately $N^{3.5}$ (the exponent here is larger than the theoretical value 3).

3. Computational details

Although the potential distributions of uniformly charged 3D triangles and rectangles are known analytically the formulations are complicated and so take a comparatively long time to compute. To give faster computing times the CPO programs use approximations where appropriate. As a part of this technique the user is asked to specify the inaccuracy ε that is desired for the calculation of potentials, which might typically be 10^{-3} in the initial stages of a new simulation or 10^{-7} for the final results. As an example of the technique we consider the calculation of the potential at a distance s from a triangular segment that has a representative length w (which we take as the maximum distance from the center of gravity to any corner). If the program judges that w/s is sufficiently small it will use the simplest approximation and treat the segment as a point charge. This happens when $w/s < f(\varepsilon)$, where the dependence of f on ε is built into the program, which was previously determined analytically by looking at the expansion of the potential in powers of w/s. In fact $f(\varepsilon)$ also depends slightly on the shape and orientation of the triangle, but a global average is used. When w/s is not small enough for this level of approximation the program next considers adding the first-order term in the expansion of the potential in powers of w/s. The dependence of this term on the shape and orientation of the triangle is also built into the program, which is therefore able to calculate the relevant parameters of each segment and store them. If this approximation is inappropriate then the program next considers the use of quadrature, using 7 quadrature points for a triangle (9 for a rectangle), the positions of which are pre-calculated and stored. The corrected Coulomb approximation described above is used for each of the quadrature points. Finally, if the potential is required at a point that is very near or on the segment then one of the exact expressions (in-plane or out-of-plane) is used for the potential, where again the necessary parameters of the triangle are pre-calculated and stored. In this way the program calculates the potential due to each triangle in the shortest time for the desired inaccuracy. This general technique is also used for 3D rectangles and 2D circular hoops, with additional types of approximation available for the hoops, namely the use of multipole moments (dipole, quadrupole and octopole, all pre-calculated) and near-axis approximations. Electrostatic fields are also calculated in an analogous way or are obtained by numerical differentiation, using the potentials at the 4 corners of a tetrahedron (in 3D systems) that surrounds the point. Extra care is taken when calculating the field at points that are near to electrode surfaces, if necessary dividing segments into smaller parts (which takes a longer computing time, but fortunately particles usually spend only a small fraction of the flight time in such situations). The criteria for selecting the appropriate approximations have usually been derived analytically but all of them have of course been carefully tested in practice.

As mentioned above, the errors due to the segments being flat and uniformly charged can be substantially reduced. Dealing first with the errors caused by the segments being flat, consider a 3D cylinder that has been subdivided into flat rectangular segments, all of which touch the cylinder at their outer edges but are otherwise inside the cylinder. The average distance s of the surface of the segments from the axis is therefore less than the radius r of the cylinder, which thus gives rise to an 'inscribing' error. To correct this the effective radius of the cylinder is increased before subdividing it by a factor that depends only on the angle subtended at the axis by a rectangle. Similarly for a spherical surface the factor depends on the average solid angle subtended by the triangles into which the sphere is subdivided. These factors were derived analytically but then slightly modified after practical tests.

The more interesting error is that due to the charge uniformity of the segments (see also below for a recent development). Clearly this error decreases when the number of segments is increased but the maximum number of segments is limited in practice by the available computer memory and the computing time. An extrapolation technique is therefore recommended, an illustration of which is given by the parameters of the two-tube lens treated in the CPO 'example' file xmpl2d01. Increase in the total number N of segments from 200 to 800 causes, for example, the object focal length f_1 of the lens to decrease from 0.799716 to 0.799684. More interestingly, a plot of f_1 versus $1/N^2$ is approximately straight, as shown in Fig. 1, giving an intercept at $f_1 = 0.799682$, which is therefore effectively the value for $N = \infty$. Analogous linear dependences are frequently found, sometimes with different exponents, for parameters such as potentials, fields, lens parameters or capacitances [14] (see also the CPO inbuilt help). This technique of 'extrapolating to infinity' is essential in some studies, such as the determination of the capacitance of the unit square with an estimated fractional inaccuracy of only 2×10^{-6} [4].

Although trajectory integration is very accurate, because the potential and field distributions are continuous, it can be slow, because all the segment charges are involved. To alleviate this a 'mesh' method is available in addition to the direct method. In the mesh method a grid of mesh points is created in the volume traced out by the beam. These points are created only as required and the potentials and fields at them are stored for use when other later trajectories pass near these points. The field at points between mesh points are obtained by bilinear interpolation using the fields at the corners of the relevant box (the mesh spacings are usually