CAD in light optics and electron optics

K. J. van Oostrum

In 1854 J. C. Maxwell devised his fish-eye lens — a sphere with a radially diminishing refractive index. The optical characteristics of this lens, which can produce a perfect image of an object, as the computer plot above suggests, are fairly easy to calculate analytically. This is not generally the case with optical elements of varying refractive index as used today for data transmission. The refractive index distribution for light-optical elements can be more or less freely chosen, but for electron-optical elements the Laplace equation or Maxwell’s equations have to be taken into account when determining the corresponding distribution. In the absence of analytical solutions it is necessary to resort to numerical methods of approximation. Light optics and electron optics have developed quite independently, so it is somewhat surprising to find that the same software is now being used at Philips Research Laboratories in Eindhoven for solving problems in both fields. Different designs can thus be optimized in a short time by CAD (Computer-Aided Design). Hypothetical experiments can be performed on the monitor screen in the first stage of development, so that there is no need for the time-consuming and expensive construction of experimental models.

Introduction

The Laplace equation is of great use in various branches of physics: in hydrodynamics, mechanics and more especially in electrical field theory. In a rectangular coordinate system the Laplace equation for an electrostatic field with no space charge may be written:

\[
\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0
\]

or briefly

\[
\Delta V = 0,
\]

where \( V \) is the electrical potential and \( \Delta \) the Laplacian operator. The solution of this important equation was long a scientific challenge. Since an analytical solution only exists for simple cases, various methods of approxi-
approximation have been devised, some of them showing considerable ingenuity. A few of these methods will be briefly reviewed here, in the same chronological sequence in which they appeared in the pages of this journal [1].

The form of a stretched membrane (such as a soap film) is also described by the Laplace equation. The vertical deflection of the membrane is equivalent to the potential $V$ in equation (1). The boundary conditions, which are given by direct voltages on electrodes, can be taken into account by local deflections of the membrane. In experiments the membrane is usually a rubber sheet; steel balls rolling on it simulate the electrons and the force of gravity simulates the electrostatic force. Only two-dimensional problems can be solved with this model. The familiar electrolytic tank, illustrated in fig. 1a, can be used to solve two-dimensional problems and three-dimensional problems with rotational symmetry. The solutions obtained are equipotential surfaces; the electron trajectories have to be constructed from equipotential lines. The use of the electrolytic tank for two-dimensional problems is roughly comparable with the use of conducting paper, which was widely used for such problems.

Resistance networks, as a method of solving two-dimensional problems, give a transition to the numerical methods of solution used today; see fig. 1b. The correspondence with the numerical methods is that a network divides the area under investigation into elements; the nodes of the network are connected by resistances. The equipotential lines are found by measuring the voltages at the nodes. The method is improved by combining the resistance network with an analog computer, which can simulate differential equations by means of circuits including operational amplifiers. The resistance network then gives the equipotential lines, and from these the analog computer calculates the electron trajectories, which are traced on an XY recorder.

An advantage of these classical methods of solving the Laplace equation and determining the electron trajectories is that the geometry and the electrode voltages can fairly easily be changed. These classical methods do however have the disadvantage that their accuracy is limited. Another disadvantage is that it is extremely difficult or impossible to take account of the space-charge effects of the electrons in their trajectories. The presence of space charge requires a correction of the Laplace equation, transforming it to the Poisson equation:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{\varrho}{\varepsilon_0},$$

where $\varrho$ is the charge density (in C/m$^3$) and $\varepsilon_0$ is the permittivity of free space. Using the Laplace equation for the calculation thus introduces an additional error in the electron trajectories.

It was largely due to C. Weber that the foundations for the ELOP (electron optics) software package for calculating electron trajectories in electrostatic fields [2] were laid at Philips Research Laboratories in Eindhoven in the sixties. The first such calculations were made at the time on the PASCAL computer [3] developed at Philips, which succeeded an experimental computer built in 1954 at Philips Research Laboratories. The mathematical foundation of Weber's program for determining potentials at the nodes of a network is formed by the iterative solution of a set of difference equations that approximate to the Laplace equation. The convergence of the iteration is improved by applying the method known as successive over-relaxation. The associated program for calculating the electron trajectories is based on the solution of the equations of motion of an electron, for a series of successive
points in time, and makes use of a Runge-Kutta third-order integration process proposed by J. A. Zonneveld [4]. The special feature here is that the program determines the size of each 'time interval' in such a way that the residual error in position and velocity per calculation step always has the same (predetermined) value. The interval length thus increases with increasing radius of curvature of the trajectory.

The ELOP software package can be used to calculate two-dimensional, rotationally-symmetrical three-dimensional and general three-dimensional potential distributions, as well as the associated electron trajectories. Space charge can be taken into account if there is rotational symmetry. The increase in the storage capacity and speed of computers through the years has resulted in greater accuracy and shorter computer times. Until recently, however, the difficulty with these programs was that they were not so convenient to use as the classical methods of solution. Varying the problem parameters took a relatively long time, as the data had to be fed into the computer on punched cards. Nor was the presentation of the results very convenient: the output was first obtained as series of numbers and later in the form of diagrams traced by a plotter.

The addition of the GELOP ('graphics for electron optics') software package has made the programs much easier to use. The user can now input the data to the system from his own terminal; see fig. 2. ‘Menus’ displayed on the alphanumeric screen enable the user to make his wishes known to the computer system. The results of the calculations are shown on the graphic display in the form of equipotential lines and electron trajectories. Successive programs in the package can be called as required. The user can move crossed wires over the graphic display to indicate the electrodes whose geometry or potential he wishes to change. If required, the modulation transfer functions, intensity distributions in the image plane or geometrical aberrations can be calculated. The user thus performs hypothetical experiments, which would be very difficult to carry out

Fig. 2. The GELOP terminal. Left, the graphic display, right, the alphanumeric screen. ‘Menus’ displayed on the alphanumeric screen enable the user to make his wishes known to the computer system. The results of the calculations are displayed on the graphic display.

calculation step always has the same (predetermined) value. The interval length thus increases with increasing radius of curvature of the trajectory.

The ELOP software package can be used to calculate two-dimensional, rotationally-symmetrical three-dimensional and general three-dimensional potential distributions, as well as the associated electron trajectories. Space charge can be taken into account if there is rotational symmetry. The increase in the storage capacity and speed of computers through the years has resulted in greater accuracy and shorter computer times. Until recently, however, the difficulty with these programs was that they were not so convenient to use as the classical methods of solution. Varying the problem parameters took a relatively long time, as the data had to be fed into the computer on punched cards. Nor was the presentation of the results very convenient: the output was first obtained as series of numbers and later in the form of diagrams traced by a plotter.

The addition of the GELOP ('graphics for electron optics') software package has made the programs much easier to use. The user can now input the data to the system from his own terminal; see fig. 2. ‘Menus’ presented on the alphanumeric screen enable the user to make his wishes known to the computer system, and the results are shown on a graphic display in the form of equipotential lines and electron trajectories. Successive programs in the package can be called as required. The user can move crossed wires over the graphic display to indicate the electrodes whose geometry or potential he wishes to change. If required, the modulation transfer functions, intensity distributions in the image plane or geometrical aberrations can be calculated. The user thus performs hypothetical experiments, which would be very difficult to carry out

G. Hepp, Measurements of potential by means of the electrolytic tank, Philips Tech. Rev. 4, 223-230, 1939;
J. C. Francken, The resistance network, a simple and accurate aid to the solution of potential problems, Philips Tech. Rev. 21, 10-23, 1959/60;


C.-E. Fröberg, Introduction to numerical analysis, Addison-Wesley, Reading, MA, 1966;
or perhaps impossible in an actual experiment in the real world. Once reasonably definitive values for the design parameters have been decided on, the user can obtain a definitive plot of the geometry on the graphic display, with the equipotential lines and the electron beams. With all these facilities available an electron-optical design can be optimized in an hour or two. This used to take many months, because of the time-consuming construction of experimental models, and the results were far less reliable.

The practical scope of the ELOP and GELOP software packages has recently been considerably extended. The facilities for solving general three-dimensional problems have been increased and extra software packages are now available for calculating magnetic fields and for solving problems of light optics. The calculation of electron trajectories in a combined electrostatic and magnetic field has special relevance to the design of Plumbicon camera tubes and television picture tubes. (In both types of tube the electron beam is deflected horizontally and vertically by magnetic fields.) The program for light-optical problems can calculate the path of light rays in media of varying refractive index. This is especially important for the design of data transmission lines with graded-index optical fibres.

In the following we shall first look at the theoretical background of the programs for the potential calculations, the trajectory calculations and the light-optical calculations. We shall then consider the software and a number of applications.

Theoretical background

Calculating the potential distribution

The space for which we wish to calculate the potential distribution is divided into elements by a rectangular grid or network. The solution of the Laplace equation (1) is based on the solution of as many difference equations as there are nodes in the grid. For each node a difference equation gives the relation between the potential at the node and the potential at the surrounding nodes. A unique solution of the set of difference equations is possible if the potential at the contours of the space considered is specified in the form of boundary conditions. As the number of nodes increases, i.e. as the grid becomes finer, the set of difference equations approximates more closely to the Laplace equation, though the computer time required becomes longer.

We shall derive the difference equation for one point for the two-dimensional case ($\partial^2V/\partial x^2 = 0$). The equation gives the relation between the potential $V(P_0)$ at the point $P_0$ and the potential at the surrounding points $P_1$ to point $P_4$; see fig. 3a. We expand the potentials $V(P_1)$ to $V(P_4)$ in a Taylor series, neglecting terms of third and higher order:

$$V(P_1) = V(P_0) + h_x \left( \frac{\partial V}{\partial x} \right)_{P_0} + \frac{1}{2} h_x^2 \left( \frac{\partial^2 V}{\partial x^2} \right)_{P_0} ,$$

$$V(P_2) = V(P_0) - h_x \left( \frac{\partial V}{\partial x} \right)_{P_0} + \frac{1}{2} h_x^2 \left( \frac{\partial^2 V}{\partial x^2} \right)_{P_0} ,$$

$$V(P_3) = V(P_0) + h_y \left( \frac{\partial V}{\partial y} \right)_{P_0} + \frac{1}{2} h_y^2 \left( \frac{\partial^2 V}{\partial y^2} \right)_{P_0} ,$$

$$V(P_4) = V(P_0) - h_y \left( \frac{\partial V}{\partial y} \right)_{P_0} + \frac{1}{2} h_y^2 \left( \frac{\partial^2 V}{\partial y^2} \right)_{P_0} .$$

From these equations we derive the expressions for $(\partial^2 V/\partial x^2)_{P_0}$ and $(\partial^2 V/\partial y^2)_{P_0}$. The Laplace equation gives a relation between these two derivatives, from which in turn the required difference equation for the point $P_0$ can be obtained:

$$V(P_0) = \frac{h_x^2 [V(P_1) + V(P_2)] + h_y^2 [V(P_3) + V(P_4)]}{2(h_x^2 + h_y^2)} .$$

For $N$ nodes of the grid there are thus $N$ difference equations. At a number of the nodes the potentials are known because of the boundary conditions.

![Fig. 3. The rectangular grid for the numerical solution of the Laplace differential equation.](image)

Proceeding from the boundary conditions and a first estimate of the potential at the points inside the boundaries, we calculate a new value for the potentials with the aid of the difference equations. With these values and the boundary conditions we can repeat the
calculation. In the resulting iterative process the calculated values give a steadily improving approximation to the true solution as the number of iterations increases. During the process we thus have to keep replacing the potentials at all the points by new potential values. This means that for every two successive computing steps the potentials at all the points should be stored in the computer memory. In fact half this amount of storage is sufficient if we adopt a procedure in which the potential at one point is always replaced by a better approximation, and all the points are then worked through in succession. D. Young has shown that the order in which this is done for the various points of the grid is very important [6]. He demonstrated that the nodes should be grouped in such a way that for calculating the potentials in one group it is only necessary to have the potentials of the previous group. A distribution of points that satisfies this requirement is known as the checker-board distribution; see fig. 3b.

In the procedure described the potential at a point gradually converges towards the true value. The convergence of the iterative process can be improved by extrapolating in the direction of the change every time a potential is replaced. Thus, in step \( k + 1 \) we do not replace the potential \( V_k \) by \( V_{k+1} \) but by \( V_k + \omega(V_{k+1} - V_k) \), where \( \omega \) is the over-relaxation factor. This procedure is known as the method of successive over-relaxation. Convergence is only obtained when \( 0 < \omega < 2 \). The convergence is improved when the range for the over-relaxation factor is \( 1 < \omega < 2 \). Our program calculates the optimum value of \( \omega \) for every step in the iteration process [4], after the user has entered an initial value, which is usually 1.5.

### Calculating the electron trajectories

The three equations that describe the motion of an electron in a combined electrostatic and magnetic field for the general three-dimensional case are:

\[
\frac{d^2 x}{dt^2} = \frac{e}{m} \left\{ \frac{\partial V}{\partial x} + \mu_0 \left( \frac{dz}{dt} H_y - \frac{dy}{dt} H_z \right) \right\},
\]

\[
\frac{d^2 y}{dt^2} = \frac{e}{m} \left\{ \frac{\partial V}{\partial y} + \mu_0 \left( \frac{dx}{dt} H_z - \frac{dz}{dt} H_x \right) \right\},
\]

\[
\frac{d^2 z}{dt^2} = \frac{e}{m} \left\{ \frac{\partial V}{\partial z} + \mu_0 \left( \frac{dy}{dt} H_z - \frac{dx}{dt} H_y \right) \right\},
\]

where \( H_x, H_y \) and \( H_z \) are the three components of the magnetic field-strength, \( t \) is the time, \( \mu_0 \) is the magnetic permeability of free space, and \( e \) and \( m \) are the charge and mass of an electron. In this article the treatment will be limited to calculations of electron trajectories in purely electrostatic fields, although the software package can also be combined with other programs to allow calculation of trajectories in a combination of electrostatic and magnetic fields.

The numerical method of approximation used, based on a Runge-Kutta integration process of the third order, will be described here for the one-dimensional equation of motion:

\[
\ddot{x} = f(x),
\]

where

\[
f(x) = \frac{e}{m} \frac{\partial V}{\partial x}.
\]

The initial values for \( x \) and \( dx/dt \) at the starting time \( t = t_0 \) are:

\[
x(t_0) = x_0,
\]

\[
\dot{x}(t_0) = \dot{x}_0,
\]

where \( x_0 \) and \( \dot{x}_0 \) represent given numbers.

The successive phases in the algorithm will be given, resulting in estimates \( \tilde{x}(t_0 + t) \) and \( \tilde{x}(t_0 + t) \) for the position and the velocity of the electron at the time \( t = t_0 + t \), where \( t \) represents a small time interval. It will then be shown that the error introduced is of the order of magnitude of \( t^4 \). It must therefore be shown that:

\[
\tilde{x}(t_0 + t) - x(t_0 + t) = O(t^4),
\]

and

\[
\dot{x}(t_0 + t) - \dot{x}(t_0) = O(t^4).
\]

First of all, equation (6) can be used to calculate two ancillary quantities \( k_0 \) and \( k_1 \) from the data (7a,b) and the time interval \( t \). These quantities \( k_0 \) and \( k_1 \) are defined as:

\[
k_0 = tf(x_0),
\]

and

\[
k_1 = tf(x_0 + \frac{3}{4} t \dot{x}_0 + \frac{3}{4} \tau k_0).
\]

The algorithm gives estimates for the position and the velocity that depend on the initial values and the ancillary quantities in the following way:

\[
\tilde{x}(t_0 + t) = x_0 + t \dot{x}_0 + \frac{3}{4} k_0 + \frac{1}{4} k_1,
\]

and

\[
\dot{x}(t_0 + t) = \dot{x}_0 + \frac{3}{4} \dot{x}_0 \tau + \frac{3}{4} \tau k_1.
\]

To prove the relations (8a,b) equation (6) is differentiated twice with respect to time:

\[
\ddot{x} = \ddot{x} f'(x) + 2 \dot{x} \ddot{x}' f'(x).
\]

From this, using (6) expressions can be obtained for the second, third and fourth derivatives of position at time \( t_0 \):

\[
\ddot{x}(t_0) = f(x_0),
\]

\[
\dddot{x}(t_0) = \dot{x}_0 f'(x_0),
\]

and

\[
\ddddot{x}(t_0) = f(x_0) \ddot{x}' f'(x_0) + \dot{x}_0 \dddot{x}' f'(x_0).
\]

Expanding (9b) in a Taylor series, we have:

\[
k_1 = tf(x_0) + \frac{3}{4} \tau \dot{x}_0 f'(x_0) + \frac{3}{4} \tau^2 k_0 f'(x_0) + \frac{3}{4} \tau^3 \ddot{x}_0 f''(x_0) + O(t^4),
\]

in which the residual term is of the order of magnitude of \( t^4 \).
Substituting equations (11a,b,c) in this equation and in (9a) gives the result:

\[ k_0 = \tau \tilde{x}(t_0), \quad (12a) \]
and

\[ k_1 = \tau \tilde{x}(t_0) + \frac{3}{2} \tau^2 \tilde{x}(t_0) + \frac{3}{2} \tau^3 \tilde{x}(t_0) + O(\tau^4). \quad (12b) \]

The ancillary quantity \( k_1 \) is thus an approximation for \( \tau \tilde{x}(t_0 + \frac{3}{2} \tau) \). Substituting (12a,b) in (10a,b) now gives the equations (8a,b), which we wished to prove.

When terms in \( \tau^4 \) and higher are neglected the term in \( \tau^2 \), just found in the substitution in the equations for the estimates \( \tilde{x}(t_0 + \tau) \) and \( \tilde{x}(t_0 + \tau) \), gives some idea of the errors made in a single step. The errors \( \epsilon_1 \) in position and \( \epsilon_2 \) in velocity are then found to be approximately equal to:

\[ \epsilon_1 = \frac{1}{2} \tau x(t_0), \]
and

\[ \epsilon_2 = \frac{1}{2} \tau^2 \tilde{x}(t_0). \]

The quantities \( \epsilon_1 \) and \( \epsilon_2 \) can be calculated for each step with the aid of a third ancillary quantity \( k_2 \), defined as:

\[ k_2 = \tau \tilde{x}(t_0 + \tau). \]

Estimates of the errors in each step can be expressed as follows in terms of the ancillary quantities:

\[ \epsilon_1 = -\frac{1}{4} \tau k_0 + \frac{1}{4} \tau k_1, \]
and

\[ \epsilon_2 = \frac{1}{4} k_0 - \frac{3}{8} k_1 + k_2. \]

These expressions can be derived in the same way as equations (8a,b).

The calculation procedure is as follows. Limiting values, which must not be exceeded in any iteration step, are set for the residual errors of position and velocity. In addition an initial value is chosen for the time interval \( \tau \). From the equations given above in the small print, estimates are made of the position and velocity at the end of the interval. The residual errors of position and velocity are also determined, and a check is then made to see whether both residual errors are smaller than the specified values. If they are, a new step in the calculation is started, with the calculated position and velocity as the initial conditions. The interval is now increased such that the estimate for the error in position amounts to 0.9 of the specified limiting value. In this way, using variable interval lengths, estimates are calculated for the position and velocity of the entire trajectory until the boundaries of the region for the electrostatic field are passed.

**Light-optical calculations**

The program used for calculating electron trajectories can also be used for light-optical calculations. Since the Laplace equation does not have to be taken into account in light-optical problems we usually have a refractive-index variation that can be chosen more or less freely. The freedom of choice is limited only by the ability to make optical components with such a varying refractive index.

A particular distribution of the optical refractive index \( n \) can be converted into a hypothetical electrostatic field by means of the equation

\[ n = \sqrt{\frac{V}{V_0}}. \quad (13) \]

In this equation the constant voltage \( V_0 \) is set equal to 1 V, so that light rays in free space are equivalent to electrons with a kinetic energy of 1 eV. The path of the light rays is thus equivalent to the paths of electrons in the hypothetical field, and we have assumed that the electrons are at rest when \( V = 0 \). As noted earlier, we are no longer concerned with the Laplace equation in converting the refractive index into an electrostatic field, so that a completely different program is required for the conversion.

Electron optics differs from light optics in that the variation of the ‘refractive index’ is much greater. A potential difference of 100 kV (which is not exceptional) results in a 300-fold variation in the refractive index. In light optics, on the other hand, a refractive index of 2 is exceptionally high. Another difference is that in conventional optics the refractive index in the various components is assumed to be constant, so that the index makes discontinuous transitions from component to component. In electron optics, on the other hand the refractive index is usually continuously variable. A system of two parallel mesh electrodes, situated at an infinitely small distance apart and carrying different direct voltages, constitutes the electron-optical analogue of a discontinuous transition in the refractive index.

Equation (13) thus forms the basis for light-optical calculations with the software package for the calculation of electron trajectories. For calculations on graded-index optical fibres it is important to have a good understanding of the velocity of propagation of light-optical signals. We shall therefore take a closer look at the significance of the refractive index concept in connection with the velocity of electrons and light waves.

In a medium of variable refractive index the refractive index can be defined as a scalar field quantity, which is a function of the position vector \( r \). From Snell’s law (a special case of Fermat’s principle) the path of an electron or of a light quantum can in general be described by

\[ \frac{d}{ds} \left( \frac{n}{dr} \right) = \nabla n, \quad (14) \]

where \( s \) is a variable that gives the position along the path.
With the aid of fig. 4 it will be shown that equation (14) is the general formulation of Snell’s Law. An electron or light wavefront moves from point 1 to point 2, which are separated by a distance ds. The change in the refractive index is represented as a discontinuous change from \(n_1\) to \(n_2\), while \(e_n\) is the unit vector perpendicular to a hypothetical boundary plane between the spaces with refractive indices \(n_1\) and \(n_2\). The vectors \((dr/ds)_1\) and \((dr/ds)_2\) are unit vectors that are tangential to the path at the points 1 and 2.

**Equation (14) has a general validity in electron optics and in light optics.** In the first case the refractive index is proportional to the velocity of the electrons; in the second case the refractive index is proportional to the reciprocal of the velocity of the light wavefronts. It must be borne in mind here that the velocity of the electrons should be regarded as the group velocity of the wave packet that according to De Broglie is equivalent to an electron. The velocity of light should be regarded as the phase velocity of the wavefronts, after Huygens. The product of the phase velocity and the group velocity of a wave packet is equal to the square of the velocity of light \([8]\). To calculate the propagation velocity of light wavefronts with the program packages for electron-optical calculations we use equation (13) and the following two equations:

\[
eV = \frac{1}{2} m v_e^2, \tag{17}
\]

and

\[
\text{or}\n
\[
\frac{c}{v_1}, \tag{18}
\]

where \(c\) is the velocity of light, \(v_e\) the velocity of an electron and \(v_1\) the phase velocity of a light wavefront.

To convert electron-optical to light-optical calculations a path trajectory \(v_1\tau_1\) must be set equal to a path trajectory \(v_e\tau_e\), where \(\tau_e\) and \(\tau_1\) are the time intervals for an electron and a light wavefront respectively. From (13), (17) and (18) it follows that:

\[
\tau_1 = \frac{\tau_e v_1}{c} \sqrt{\frac{2eV_0^2}{mV_0}}.
\]

When \(V_0 = 1\ V\) this equation becomes:

\[
\tau_1 = \frac{\tau_e v_1}{c} \sqrt{\frac{2e}{m}}. \tag{19}
\]

Equation (19) permits an interval \(\tau_e\) used in the Runge-Kutta procedure to be converted into an interval \(\tau_1\) for a light wavefront, so that the time of transit of light signals can be calculated.

**The software**

*The ELOP software package*

ELOP contains the individual programs for the actual calculations, which are based on the mathematical and physical principles described earlier. The procedures for the calculations are largely comparable with those described earlier by C. Weber in this journal and elsewhere \([21]\). The present software package ELOP contains the following programs:

• RELA. This calculates the voltages at the nodes of a rectangular grid by solving a large number of difference equations.

• BAAN. This program calculates the trajectories of electrons in an electrostatic field, described by data supplied by RELA. The BAAN program can at the same time make use of data provided by programs for calculating magnetic fields [9].

• CODE. This program calculates the position of the boundary of the electrostatic field region with respect to the nodes of the rectangular grid.

The GELOP software package

GELOP is used to input the data for a problem to the ELOP software package. After a computation cycle GELOP presents the results to the user via various output peripherals. If required, several CAD cycles can be run one after the other (CAD = Computer-Aided Design) see fig. 5. The user ‘talks’ to GELOP by using 30 menus, each offering him 13 choices. Data input and job control in the various programs are performed by means of EXEC commands in the VM/CMS operating system used in the IBM 3081 computer. GELOP contains the following programs, which we shall discuss in more detail shortly.

• CONTOUR. This program is used to give a geometrical description of the boundaries enclosing the space for which the problem is to be solved. The user also states the boundary conditions, i.e. the voltages for the different parts of the boundary. The program also defines the grid for the electrostatic field.

• BEAM GENERATION. The user uses this program to input initial conditions for the individual electrons of the beam.

• DIAGNOST. This program shows beam cross-sections at locations selected by the user. It also supplies him with information about aberrations.

• PLOT. This program operates a digital plotter to draw the boundaries, the grid, a number of equipotential lines and the electron trajectories.

The GRINOP software package

GRINOP consists of somewhat modified versions of ELOP and GELOP and is designed for solving light-optical problems in media of varying refractive index. In the RELA program the subroutines for solving the Laplace equation are replaced by subroutines for calculating a hypothetical voltage from the refractive index from equation (13).

The CONTOUR program

When defining the boundaries of the problem a distinction must be made between geometries with two-dimensional symmetry and rotational symmetry on the one hand and three-dimensional geometries on the other. In the following the rotationally symmetric and two-dimensional cases will be lumped together, since there is no difference between them in practice.

The boundary of the problem region forms a barrier for the moving electrons (no account is taken of effects such as reflection or secondary emission). The voltage distribution at the boundary is also specified as a boundary condition for the Laplace equation. Finally, the boundary limits the number of nodes for which the voltage must be calculated. For two-dimensional problems the maximum number of nodes is 20,000, for three-dimensional problems it is 30,000.

For two-dimensional cases a boundary consists of a number of closed curves that must not intersect. The closed curves represent intersections of the problem plane with surfaces in which the voltage distribution is given as a boundary condition. For each subregion within the outer boundary it is necessary to state whether it is ‘full’ or ‘empty’, the statement ‘empty’ meaning that electrons can move freely. By definition the region outside the outer boundary is ‘full’. The program is designed so that every crossing of a closed curve represents a transition from ‘full’ to ‘empty’ or vice versa. Each closed curve is built up from straight or circular line segments. The user must state the radius of curvature and end-point for each circular line segment. He must also indicate whether the curvature is concave or convex on the outside. In this manner any two-dimensional curve can be defined unambiguously.

For three-dimensional cases the problem region must be built up from a number of volume elements. There are five basic shapes for these volume elements: cylinder, cone, spherical sector, rectangular block and irregular hexahedron. To each of these volume ele-
ments the user assigns the statement 'full' or 'empty'. Also, for regions in which two or more basic shapes have been defined, the user must indicate by priority numbers 1, 2 etc. which basic form is applicable.

The user also uses the CONTOUR program to relate the grid for the potential distribution to the problem region. The grid can be refined locally if required. The user specifies symmetry planes if possible, thus reducing the number of nodes and the computer time. Fig. 6 gives an example of a problem — a calculation procedure the user can vary the design parameters even more simply than was possible with the classical simulation methods.

The BEAM GENERATION program

In GELOP the position and velocity of a moving electron at a specified time are determined by the quantities \( X, Y, Z, E, QX/Q \) and \( QY/Q \). \( X, Y \) and \( Z \) are the coordinates of position; \( E \) is the kinetic energy of the electron; \( Q \) is the length of the velocity vector; \( QX \) and \( QY \) are the components of this vector in the \( X \)- and \( Y \)-directions. The magnitude of the velocity vector thus follows from the value of \( E \). The direction of the velocity vector is completely determined by \( QX/Q \) and \( QY/Q \), since the cosine of the angle \( \gamma \) between the velocity vector and the \( Z \)-coordinate is given by

\[
\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1,
\]

where \( \cos \alpha = QX/Q \) and \( \cos \beta = QY/Q \).

For  the reference numbers to the various parts of the boundary surfaces of the problem space. During the various design runs the reference numbers can be assigned to a direct voltage, a voltage distribution (linear or logarithmic) or a value for the first derivative of the voltage (the field-strength). With this product the user can vary the design parameters even more simply than was possible with the classical simulation methods.

Fig. 6. Statement of the geometry of the problem by means of the CONTOUR program. The problem represents the calculation of a set of deflection plates. The program gives the projections of the problem region in the \( ZX \)-, \( ZY \)- and \( XY \)-planes. It also produces a 'wire-frame model' in oblique projection. In this case there is one plane of symmetry, so that the user only has to define a mesh network for half of the problem region. All boundaries are planes. The actual problem region has the shape of a rectangular block and is 'empty'. The regions defined inside it are 'full' and form the two electrodes.

Fig. 6g gives an example of a problem — a calculation on a set of deflection plates — with one plane of symmetry. The figure also shows that the CONTOUR program presents the problem boundaries in three orthogonal projections perpendicular to the \( X \)-, \( Y \)- and \( Z \)-axes and also as a 'wire-frame model' shown in oblique projection. The \( Z \)-axis usually corresponds to the direction in which the electrons are moving, so that in general the object plane and the image plane are perpendicular to the \( Z \)-axis.

The user also uses CONTOUR to assign electrode reference numbers to the various parts of the boundary surfaces of the problem space. During the various design runs the reference numbers can be assigned to a direct voltage, a voltage distribution (linear or logarithmic) or a value for the first derivative of the voltage (the field-strength). With this product the user can vary the design parameters even more simply than was possible with the classical simulation methods.

The BEAM GENERATION program

In GELOP the position and velocity of a moving electron at a specified time are determined by the quantities \( X, Y, Z, E, QX/Q \) and \( QY/Q \). \( X, Y \) and \( Z \) are the coordinates of position; \( E \) is the kinetic energy of the electron; \( Q \) is the length of the velocity vector; \( QX \) and \( QY \) are the components of this vector in the \( X \)- and \( Y \)-directions. The magnitude of the velocity vector thus follows from the value of \( E \). The direction of the velocity vector is completely determined by \( QX/Q \) and \( QY/Q \), since the cosine of the angle \( \gamma \) between the velocity vector and the \( Z \)-coordinate is given by

\[
\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1,
\]

where \( \cos \alpha = QX/Q \) and \( \cos \beta = QY/Q \).

For  the reference numbers to the various parts of the boundary surfaces of the problem space. During the various design runs the reference numbers can be assigned to a direct voltage, a voltage distribution (linear or logarithmic) or a value for the first derivative of the voltage (the field-strength). With this product the user can vary the design parameters even more simply than was possible with the classical simulation methods.

The BEAM GENERATION program

In GELOP the position and velocity of a moving electron at a specified time are determined by the quantities \( X, Y, Z, E, QX/Q \) and \( QY/Q \). \( X, Y \) and \( Z \) are the coordinates of position; \( E \) is the kinetic energy of the electron; \( Q \) is the length of the velocity vector; \( QX \) and \( QY \) are the components of this vector in the \( X \)- and \( Y \)-directions. The magnitude of the velocity vector thus follows from the value of \( E \). The direction of the velocity vector is completely determined by \( QX/Q \) and \( QY/Q \), since the cosine of the angle \( \gamma \) between the velocity vector and the \( Z \)-coordinate is given by

\[
\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1,
\]

where \( \cos \alpha = QX/Q \) and \( \cos \beta = QY/Q \).

For  the reference numbers to the various parts of the boundary surfaces of the problem space. During the various design runs the reference numbers can be assigned to a direct voltage, a voltage distribution (linear or logarithmic) or a value for the first derivative of the voltage (the field-strength). With this product the user can vary the design parameters even more simply than was possible with the classical simulation methods.

The BEAM GENERATION program

In GELOP the position and velocity of a moving electron at a specified time are determined by the quantities \( X, Y, Z, E, QX/Q \) and \( QY/Q \). \( X, Y \) and \( Z \) are the coordinates of position; \( E \) is the kinetic energy of the electron; \( Q \) is the length of the velocity vector; \( QX \) and \( QY \) are the components of this vector in the \( X \)- and \( Y \)-directions. The magnitude of the velocity vector thus follows from the value of \( E \). The direction of the velocity vector is completely determined by \( QX/Q \) and \( QY/Q \), since the cosine of the angle \( \gamma \) between the velocity vector and the \( Z \)-coordinate is given by

\[
\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1,
\]

where \( \cos \alpha = QX/Q \) and \( \cos \beta = QY/Q \).
In the BEAM GENERATION program moving electrons are represented for a particular value of the Z-coordinate by points in the four coordinate systems \((X,Y), (X,QY/Q),(Y,QY/Q)\) and \((QX/Q,QY/Q)\). The user also gives the initial conditions for an electron beam as points in these coordinate systems. The points in the coordinate system \((X,Y)\) form an 'electron cloud'. The propagation of this electron cloud as a function of time and hence of the Z-coordinate determines the dimensions of an electron beam.

The initial conditions for an electron cloud may be a uniform distribution of position, velocity and direction, or a random distribution of these quantities. The user may decide to choose the same starting point for all electrons (indicating it on the graphic display by means of the crossed wires). The end-points of the velocity vectors may be situated on a hemisphere, for example, with the end-points of the projections \(\sqrt{QX^2 + QY^2}\) of the velocities at the nodes of a rectangular grid; see fig. 7. If a random distribution is chosen for the various quantities, the program operates with uniform distribution functions. The sets of initial conditions for position, direction and velocity can if required be multiplied by one another. At each of the various starting points the electrons then start their movement from different angles. Different velocities are then associated with each angle and with each starting point. In addition to these facilities it is also possible to assign to each electron trajectory a relative intensity number with a value between 0 and 1. This number gives the probability that an electron with the specified initial conditions will actually be generated.

**The DIAGNOST program**

The DIAGNOST program can be used to determine the extent of the geometrical aberrations and the parts of the electrostatic field that are responsible for them. DIAGNOST thus provides the user with various ways of evaluating the characteristics of an electron beam, and hence of the imaging system.

In the first place DIAGNOST can present the position and velocity of individual electrons in a beam at various locations in the object under investigation, either for different values of the Z-coordinate, or for different equipotential surfaces (i.e. for different voltage values). In addition the program can calculate the modulation transfer functions of the imaging system for the \(X-\) and \(Y-\)directions. (The quality of an imaging system can be characterized more satisfactorily by the modulation transfer function than by a value for the resolution \([10]\).)

The DIAGNOST program can also display intensity distributions in the image plane. The program supplies these distributions by integrating the number of electrons for each mesh of a grid to be defined in the image plane. It does this on the basis of the stated relative intensity distribution of the electrons generated in the object plane. The program presents the intensity dis-

---

**Fig. 7.** Example of the statement of the initial conditions for an electron cloud, using the BEAM GENERATION program. The same starting point and the same velocity have been chosen for all the electrons. The end-points of the projections in the \(XY\)-plane of the velocity vectors are situated at the nodes of a rectangular grid, defined by the user. a) Schematic representation of the initial conditions. \(Q\) velocity vector with components \(QX, QY\) and \(QZ\) along the \(X, Y\) and \(Z\)-coordinates. \(\alpha, \beta\) and \(\gamma\) angles made by the velocity vector with these projections. \(Qc\) velocity vector for the central trajectory. The arcs are the cross-sections of the spherical surface on which the end-points of the velocity vectors are located with the three planes of the coordinate system. b) Presentation of the initial conditions by the program. The starting points of the velocity vectors are represented as points (here at the origin) in the coordinate system \((X,Y)\), the directions of the velocity vectors are represented as points in the coordinate systems \((X,QX/Q),(Y,QY/Q)\) and \((QX/Q,QY/Q)\).
Fig. 8. Some of the facilities offered by the DIAGNOST program. 

a) Presentation of an electron-intensity distribution. The lines connect points that represent the number of electrons, integrated for each mesh of a grid defined by the user. 

b) Schematic representation of the construction of intermediate images. 

c) Presentation of the corrected dimensions in the X-direction of the intermediate images as a function of a sequential number for the equipotential planes. The computer also prints out a list of voltages corresponding to the numbers of the equipotential planes. The values along the vertical axis are a measure of the relative aberration in the X-direction. If the imaging system produces no aberration, regions P with increasing aberration must be compensated completely by regions N with decreasing aberration.

b) Schematic representation of the construction of intermediate images. a, b electron trajectories, c central electron trajectory. The optical properties of the refracting surface \( R \) follow from the change in the intermediate images associated with the neighbouring equipotential surfaces \( V_1 \) and \( V_2 \). The intermediate image \( I_1 \) is obtained from the extrapolation of a and b at the points \( A_1 \) and \( B_1 \). Similarly, \( I_2 \) is obtained from the extrapolation of a and b at \( A_2 \) and \( B_2 \). The position of the intermediate images is obtained from the point of intersection of the extrapolation of c with the extrapolation of a reference trajectory (not shown) of an electron that leaves the object point at an angle to c. The magnifications of the intermediate images follow from a second reference trajectory (not shown) of an electron that leaves the object plane at some distance from the object point. The dimensions \( d_1 \) and \( d_2 \) of the intermediate images in the X- or Y-directions are corrected by the magnifications thus determined.

c) Presentation of the corrected dimensions in the X-direction of the intermediate images as a function of a sequential number for the equipotential planes. The computer also prints out a list of voltages corresponding to the numbers of the equipotential planes. The values along the vertical axis are a measure of the relative aberration in the X-direction. If the imaging system produces no aberration, regions P with increasing aberration must be compensated completely by regions N with decreasing aberration.

DISTRIBUTIONS AS 'CONTOUR LINES' FOR FIXED INTENSITY VALUES, AS A TWO-DIMENSIONAL CURVE ALONG A LINE IN THE X- OR Y-DIRECTION, OR AS A THREE-DIMENSIONAL DISTRIBUTION; SEE FIG. 8A.

DIAGNOST can also transform the position coordinates of an electron cloud, without any change in the original velocities, by displacing it through a specified distance to another plane perpendicular to the Z-axis. In cases with rotational symmetry an electron cloud can be moved to the Z-coordinate where the beam cross-section is at a minimum. In this way beam cross-sections at different positions can be compared and the origin of aberrations such as astigmatism and coma can be established.

The principal facility that DIAGNOST offers is the construction of a virtual intermediate image for a stated potential surface by extrapolation of the electron trajectories of the beam; see fig. 8b. The figure caption explains how the program determines the position of the intermediate images and how it calculates the magnification of the intermediate images. Each intermediate image is corrected by the calculated magnification factor, so that the various intermediate images are comparable with one another. DIAGNOST can display the aberrations in the X- and Y-directions, obtained from the dimensions of the corrected intermediate images, as a function of a number corresponding to the sequence of the equipotential surfaces; see fig. 8c. At the same time the program prints out a list of the potentials associated with these numbers. In fig. 8c the region of the field in which the aberrations increase is denoted by \( P \); the region in which the aberrations decrease is denoted by \( N \). The resulting aberration is zero when the con-
Fig. 9. Modification of the characteristics of a triode electron gun, e.g. of a picture tube. 

a) Geometry of the original gun. The horizontal axis, which is the Z-axis, is the axis of rotational symmetry. The asterisks and points indicate nodes of the grid defined by the CONTOUR program. The points lie in the region stated to be 'empty'; the asterisks lie in the region stated to be 'full'. The nodes denoted by '+' lie exactly on the boundary of the problem region; the nodes denoted by '0' are separated from the boundary by less than one mesh width. 

b) A number of equipotential lines and electron trajectories for the original gun. The electrodes are indicated by thickening of the lines traced by the plotter. The cathode is on the left, the grid is at the centre and the anode is on the right. (In reality the anode has a centre hole aligned with the Z-axis.) The potentials on the electrodes are 0, - 200 and 4000 V; between the electrodes the voltage changes linearly along the boundary. The beam crossover — which is also the object point for the picture tube — is on the right outside the diagram. The problem was to shift the crossover towards the cathode. 

c) Geometry of the gun with modified grid-electrode dimensions. 

d) The corresponding equipotential lines and electron trajectories (with the same electrode voltages). It can be seen that the equipotential lines are more curved, so that the gun converges the electron trajectories more strongly. The crossover is therefore about half-way between the cathode and the anode.
Fig. 10. Calculations on a lip lens. 
(a) The geometry specified with CONTOUR, shown in a plane through the axis of symmetry. For clarity the material cross-sections are shown hatched and the lines indicating the boundaries of the material are shown thickened. The thin lines and the thick lines correspond to the volume elements that together form the two electrodes. The volume elements in this case are a rectangular block and five cylinders, four of them with a common symmetry axis. The symmetry axis of the fifth cylinder is perpendicular to the plane of the drawing. 
(b) Perspective drawing of the lip lens. The ‘lips’ of the inner electrode are clearly visible. The outer electrode has completely rotational symmetry. 
(c) Equipotential lines and electron trajectories in the ZX-plane. The Z-axis coincides with the axis of symmetry. The electrons start from the plane \(Z = 0\); the direction of the initial velocity is parallel to the Z-axis. The initial kinetic energy of the electrons is equal to the product of their charge and the voltage on the inner electrode. The voltage on the inner electrode is 500 V and the voltage on the outer electrode is 16000 V. 
(d) Equipotential lines and electron trajectories in the ZY-plane. Comparison of (c) and (d) confirms that the beam is asymmetrical. 
(e) Characteristics of the electron cloud in the vicinity of the smallest cross-section of the beam. It can be seen from the XY graph that this cross-section resembles a line focus.
Fig. 11. Calculations on an X-ray image intensifier with large input format. a) Geometry of the image intensifier with five electrodes; the electrodes are indicated by thick lines. The X-ray screen (object plane), which acts as the cathode, is on the left; the output screen (image plane), which acts as the anode, is on the right. b) The equipotential lines and the trajectories of electrons that start from a point situated well off-centre in the object plane. c) XY plot of the cross-section of the beam in (b) near to the smallest cross-section. This cross-section is close to the output screen. The asymmetry of the beam shows that the aberrations of astigmatism and coma are present; they are relatively minor, however. d) Computer-calculated radial dimensions of the intermediate images, constructed by extrapolation and subsequently corrected, as a function of the sequential number of the equipotential surfaces. The increase in the aberrations at the location of the two inner electrodes is compensated by a decrease in the vicinity of the anode so that the resultant aberration is relatively small.
Fig. 12. Calculations on an optical fibre with the GRINOP software package. 

a) The refractive index profile: the refractive index as a function of the ratio $\rho$, equal to the distance from the symmetry axis divided by the outer radius of the fibre. An unusual refractive index profile has been chosen.

b) Paths of the light rays in the fibre, seen in the direction of its length. The rays are numbered from 2 to 21. All the rays start at the position where $\rho = 0.6$; all the light rays have the same initial velocity. The radial component of the velocity is zero for all rays; the tangential component, starting from zero for ray 2, increases in proportion to the number of the light ray; the axial component thus decreases correspondingly. Ray 2 propagates in a plane through the axis of symmetry. The rays 2 to 16 travel more or less along the circumference of the fibre, but because of total internal reflection they do not leave the 'skin' of $\rho = 0.6$ to $\rho = 1.0$. The tangential component of the velocity of the rays 17 to 21 has become so large that these rays are not totally internally reflected, and therefore leave the fibre.
tributions of regions $P$ and $N$ compensate one another in the image plane.

**Examples of applications**

*Fig. 9* shows calculations on an electron gun, e.g. for a television picture tube. From left to right the gun consists of a cathode, a grid and an anode, at constant potentials of 0, $-200$ and 4000 V. We wanted to find how to change the position of the beam 'crossover' without altering the voltages on the electrodes. (Changing the position of the crossover changes the magnitude of the spot on the tube screen.)

*Fig. 10* shows calculations on a 'lens'. A lens is a classical example of a line-focus lens. By using an approximation to a line-focus electron source astigmatism in an imaging system can be compensated.

*Fig. 11* gives an idea of the calculations for the aberrations (they are only small) of the large-format X-ray image intensifier described in an earlier article in this journal. For a beam issuing from a point on the X-ray screen the DIAGNOST program calculates the shape of the beam cross-section close to the output screen. It also calculates the magnitude of the extrapolated and corrected intermediate images as a function of the sequential number of the equipotential planes, thus indicating where the aberrations arise.

---