INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONES: 224281/2/3/4/5/6
CABLE: CENTRATOM - TELEX 460392 - I

SMR.201/4

SECOND WORKSHOP ON MATHEMATICS IN INDUSTRY

(2 - 27 February 1987)

"SOME MATHEMATICAL MODELLING PROBLEMS FOR THE MICROELECTRONICS INDUSTRY."

R. SMITH
Department of Mathematics
University of Technology
Loughborough, Leics. LE11 3TU
U.K.

SOME MATHEMATICAL MODELLING PROBLEMS FOR THE MICROELECTRONICS INDUSTRY

by

R. SMITH

Department of Mathematics.
Loughborough University of Technology.
Loughborough.
Leicestershire.
LE11 3TU

1. INTRODUCTION

Microelectronics has been one of the worlds growth industries at a time when in the west generally and in Britain in particular, traditional manufacturing industry has been in serious decline. Just as the post-war industrial boom in Western Europe gave rise to, for example, the manufacture of motor cars with its associated other industries - the machine tool industry to make the parts for the cars, the road building and leisure industries to allow full use of the newly available consumer product, so the manufacture of microelectronics devices has its own associated industries; in particular, the manufacture of equipment to make the devices and the software industry to make full use of the manufactured computer products. Mathematicians have normally regarded their contribution to Information Technology in terms of the writing of software packages and it may not be generally recognised that the applied mathematician has an important role to play in developing models that can be used to explain certain aspects of the microfabrication process itself and to assist in modelling fabrication and also the design of equipment used in the process. This equipment - the machine tools - of the microelectronics industry, itself is often a highly sophisticated electronic device. It can involve the use of charged particle beams to delineate patterns and operate under high vacuum conditions. V.L.S.I. (very large scale integration) devices which

have circuits with high density and progressively smaller features are invariably manufactured using such equipment, which has many advantages over chemical etching techniques.

Microelectronics circuit fabrication is based on the ability to remove and add material selectively from and to the surface of a suitable substrate, e.g., a semiconductor. However, new devices such as gallium arsenide integrated circuits demand that materials difficult to etch chemically be patterned down to submicron ($< 10^{-6}$ m) dimensions.

The patterns can be manufactured by using a beam of energetic particles incident on the surface to remove material. This process of the removal of atoms by an incident ion beam - akin to sand blasting on the atomic scale - is known as sputtering. Masks can be used with broad beams, so that only certain parts of the surface are eroded, corresponding to the required surface pattern. Alternatively if the incident particles are of a very small energy then the incident material can stick to the substrate and build-up a thin film. This is often of a different material to the substrate so that patterns can be etched in materials which have different electronic properties to the substrate. Sometimes material can be added and etched back to build bi or even tri-level electronic structures, in this way.

The technique outlined above is essentially non-directional but because the incident particles are often charged, electric and magnetic fields can be used to steer and focus the beams. Direct write of patterns can therefore be achieved, a process known as lithography and a question that naturally arises is 'what are the fundamental size-limiting factors on the production of devices'? Is present technology near the limits of smallness of size in terms of microelectronic device production? Electron and ion beam lithography are capable of producing even smaller devices that those producible by masking but how viable is this economically? It will be

shown that the mathematician has a role to play in the answer to this question.

In both lithographic and masking techniques it is not only the production of a pattern by atomic sand blasting techniques which is of interest. In addition to pattern delineation, ions from the beam can be implanted in the material in order to change its electronic properties in some subsurface region, with the surface damaged induced being a subsiduary factor. Mathematicians have a role to play in the description of this process also.

This paper will concentrate on two mathematical problems associated with processes described briefly above. First, the erosion/deposition phenomena on the surfaces of materials will be modelled by a non-linear wave equation. Secondly, electrostatic optical systems for focussing charged particle beams will be computed. This requires the solution of Poisson's equation for complicated geometries[2]. This is not a complete list but one used to illustrate some of the mathematical complexities that can arise in modelling real life physical problems.

2. PATTERN DELINEATION: BASIC MATHEMATICAL TECHNIQUES

2.1 The erosion of two-dimensional amorphous sufaces

Method of Characteristics

Consider the simplified model of a two-dimensional section of a surface, subject to bombardment by a beam of particles incident in the y-direction. This is shown schematically in figure 2.1. In the simple case of a uniform beam, the rate of erosion of the surface is a function only of the angle of incidence θ, so that

$$\left.\frac{\partial y}{\partial t}\right|_x = -\frac{J}{N}\,Y(\theta) \qquad (2.1)$$

where J is the particle flux, N the atomic density of the target and $Y(\theta)$ is called the sputtering yield and its units are atoms/incident particle. Differentiating equation (2.1) with respect to x gives

$$\frac{\partial^2 y}{\partial x \partial t} = -\frac{J}{N}\,Y'(\theta)\,\frac{\partial \theta}{\partial x} \qquad (2.2)$$

However,

$$\left.\frac{\partial y}{\partial x}\right|_t = \tan\theta$$

and differentiating this equation with respect to t, and substituting into equation (2.2) gives

$$\frac{\partial \theta}{\partial t} + \frac{J}{N}\,Y'(\theta)\,\frac{\partial \theta}{\partial x} = 0 \qquad (2.3)$$

Similarly, from figure 2.1, since $\delta x = -\cot\theta\,\delta y$:

$$\left.\frac{\partial x}{\partial t}\right|_y = -\cot\theta\left.\frac{\partial y}{\partial t}\right|_x = \cot\theta\,\frac{J}{N}\,Y(\theta)$$

so

$$\frac{\partial^2 x}{\partial t \partial y} = \frac{J}{N}\left[\,Y'(\theta)\cot\theta - Y(\theta)\,\mathrm{cosec}^2\theta\right]\frac{\partial \theta}{\partial y} \qquad (2.4)$$

but

$$\left.\frac{\partial x}{\partial y}\right|_t = \cot\theta$$

and differentiating this equation with respect to t and substituting into (2.4) gives

$$\frac{\partial \theta}{\partial t} + \frac{J}{N}\left[\,Y(\theta) - Y'(\theta)\sin\theta\cos\theta\right]\frac{\partial \theta}{\partial y} = 0 \qquad (2.5)$$

Equations (2.3) and (2.5) form the basis of a scheme to calculate changes in surface shape under particle bombardment once the sputtering yield $Y(\theta)$ is known. This can be measured experimentally or deduced using computer models.

Equations (2.3) and (2.5) are both first order non-linear partial differential equations (P.D.E.'s) of a type first studied by Lagrange. The solution is by the method of characteristics. This method effectively reduces the P.D.E. to a series of ordinary differential equations (O.D.E.'s) along the characteristic lines.

For a general equation of the form

$$P(\theta,x,t)\ \frac{\partial\theta}{\partial x} + Q(\theta,x,t)\ \frac{\partial\theta}{\partial y} = R(\theta,x,t) \qquad (2.6)$$

the characteristic lines are lines in the $x - t$ plane defined by $dx/dt = P/Q$, since along these lines (2.6) reduces to

$$\frac{dx}{dt}\frac{\partial\theta}{\partial x} + \frac{\partial\theta}{\partial t} = \frac{R}{Q} \quad \text{i.e.} \quad \frac{\partial\theta}{\partial x}dx + \frac{\partial\theta}{\partial t}dt = \frac{R}{Q}dt$$

i.e.

$$\frac{d\theta}{dt} = \frac{R}{Q} \qquad (2.7)$$

This is an O.D.E for θ.

If these general results are applied to equations (2.3) and (2.5), the characteristics are lines of constant surface orientation (θ = const) defined by.

$$\frac{dx}{dt} = \frac{J}{N}Y'\cos^2\theta\ , \qquad \frac{dy}{dt} = \frac{J}{N}\left[Y - Y'\sin\theta\cos\theta\right] \qquad (2.7)$$

Thus the characteristics are also straight lines.

If however, the flux is spatially non-uniform $J = J(x)$ then the characteristic equations remain the same, but θ is no longer constant along them. Equation (2.7) becomes

$$\frac{d\theta}{dt} = -\frac{J'(x)}{N}Y\cos^2\theta \qquad (2.8)$$

and although JY/N is constant along the characteristics, these are no longer straight lines.

Computational Method for Determining Surface Shape

At first sight, the method of computing the surface shape appears very simple. However, the non-linearity of the process reveals a number of hidden difficulties. First define the initial surface shape by a series of points (x_i, y_i) $i = 1,\ldots,n$. The surface at time t, for the uniform flux case, is (x_i', y_i') $i = 1,\ldots,n$ where

$$x_i' = x_i + \frac{J}{N}Y'(\theta_i)\cos^2\theta_i t$$

$$y_i' = y_i + \frac{J}{N}\left[Y - Y'\sin\theta_i\cos\theta_i\right]t \qquad (2.9)$$

However, if this scheme is applied, in the simple form as given, to surfaces with initial surface gradient discontinuities, gaps can appear in computed surface at time t, due to divergence of the characteristics (an expansion fan). Similarly characteristics from different portions of a smooth surface can intersect (shock waves) to produce a surface at time t which would consist of cusps and folds. These are unphysical and the real surface consists of gradient discontinuities at these points.

For the case of a non-uniform beam, then the simulatenous O.D.E.'s (2.7) and (2.8) must be solved at each point on the initial surface, usually this has been done using a Runge-Kutta fourth order

method. This increases the computing time and the same problems of shock and expansion waves arise due to the non-linearity of the process as with the case of uniform characteristics.

Examples of the Erosion of Smooth Surfaces

Consider the erosion of sections of a circular contour with a sputtering yield dependence on incidence angle that due to 1 keV Ar^+ ions on Si. This dependence which has been measured experimentally[3] is shown in figure 2.2. The ion current i.e. flux is assumed to vary spatially of the form

$$J = J_0 \, e^{-x^2/\rho^2}$$

where J_0 and ρ are constant.

Figures 2.3. a-d. indicate quite clearly, the formation of edges from a smooth initial surface. For the case of a uniform beam (figure 2.3 a), an edge first form at a point near the top of the circle. This edge then moves inwards, to intersect with a similar edge formed symmetrically on the opposite side of the circle, eventually producing a sharp pointed structure. In figures 2.3. b-d, the characteristics are curved and a sharp rimmed crater is formed in figure 2.3 d, due to the intersection of curved characteristics. The formation of sharp-pointed structures such as cones is a well observed experimental facet of surfaces undergoing ion bombardment.

The Formation and Motion of Edges

An edge first forms when neighbouring characteristics intersect (see figure 2.4). If the velocity of erosion along the characteristics (given by equations 2.7) is $v = (v_x, v_y)$, then neighbouring characteristics intersect at a point (x, y) where

$$x = x_0 + v_x t = x_0 + \delta x_0 + v_x(\theta_0 + \delta\theta_0)t \qquad (2.10)$$

The quantities in this equation are defined in figure 2.4. with (x_0, y_0) being the point on the initial surface when $t = 0$ whose gradient is θ_0. Thus from equation (2.10), the edge occurs

$$t = \frac{dx_0}{dv_x} = -\frac{dx_0}{d\theta} \Big/ \frac{dv_x}{d\theta} \qquad (2.11)$$

However, this is only the condition for any two neighbouring characteristics to intersect. The edge first forms minimum value of t as a function of θ_0. i.e. when $dt/d\theta_0 = 0$. This yields the further condition

$$\dot{v}_x(\theta)\dot{x}_0(\theta) - \dot{x}_0(\theta)\dot{v}_x(\theta) = 0 \qquad (2.12)$$

After its initiation, the edge is produced by the intersection of the characteristics from points of orientation θ_1 and θ_2 at (x_1, y_1), (x_2, y_2) i.e.

$$\frac{x_1(\theta_1) - x_2(\theta_2)}{v_x(\theta_1) - v_x(\theta_2)} = t = \frac{y_1(\theta_1) - y_2(\theta_2)}{v_y(\theta_1) - v_y(\theta_2)} \qquad (2.13)$$

Example

Consider the edge formation on a symmetric surface protrusion

$$y = -x^m \quad ; \quad x > 0 \qquad m > 1$$
$$y = -(-x)^m \quad ; \quad x < 0$$

The surface is continuous at $x = 0$, with continuous derivative. The equation of the surface can also be written as

$$x = -\left[\frac{1}{m}\tan\theta\right]^{1/m-1}$$

so $\dfrac{dx}{d\theta} = 0$ when $\theta = 0$ if $m < 2$. if $1 < m < 2$

an edge forms immediately at the maximum point on the surface. If $m = 2$, the edge forms at a time t

$$t = \frac{\sec^2\theta}{2\dfrac{dv_x}{d\theta}}$$

where θ is given by

$$\frac{d^2v_x}{d\theta^2} = 2\tan\theta\,\frac{dv_x}{d\theta}$$

so for $m > 2$, the edge forms at symmetric points away from the maximum point for finite $t > 0$.

The Expansion of Corners

Consider now the erosion of a point P at a surface gradient discontinuity. Because θ is discontinuous at P, the point (x_P, y_P) is not give straighforwardly by equations (2.9). It is however, possible to investigate the erosion of P by considering, first, a value of θ at P corresponding to the left-hand value θ_L and, secondly, a value corresponding to the right-hand plane θ_R. see figure 2.5. If the characteristics from P are such that the planes intersect at a point θ as in Figure 2.5a, then the locus of the edge P as erosion continues is easy to calculate. However, Figures 2.5b and 2.5c indicate cases where the eroded left and right-hand sections from P do not intersect. In this case the point P is treated by assuming that all angles between θ_L and θ_R are potentially present at P and the locus of these characteristics is drawn. The shortest portion of this locus curve then joins the two portions of the surface. The normalized characteristic locus curve (assuming $Jt/N = 1$) is shown in figure 2.6 for the sputtering yield of figure 2.2 and two important examples in microelectronic device production are illustrated in figure 2.7.

These are the erosion at normal incidence of a convex and concave 90^o corner. The rounding of the convex corner which occurs in figure 2.7a is a well known feature of patterns which are etched using ion beams but for many applications it is important to maintain square-wave structures. The expansion of the concave corner, 2.7 b, is a less important problem since the expansion only marginally alters the perpendicular wall, which at such a concave corner is in any case subject to the influence of secondary effects such as redeposition or particle reflection.

Application to Device Production

In the production of patterns on microelectronic devices, material is often added to a substrate by low energy sputter deposition, in addition to being removed by etching. Thus for example, a thin metallic film can be laid down on a semiconductor substrate and then a mask overlaid and the pattern etched using a broad ion beam. Sometimes it is not solely a single structure that is manufactured. It is sometimes required to build bi or even tri level structures by a combination of (sputter) deposition and sputter erosion. In this case it is necessary to hide the structures beneath flat layers of material before the next structure is laid down. This technique is know as planarisation.

The deposition process can often occur isotropically. By this we mean that material is added to the surface in such a way that the surface moves a fixed amount in its normal direction. Chemical erosion, where ballistic effects are unimportant can also be an isotropic process. In this case the altered morphology can be calculated by a Huyghen's wavefront construction from the original surface where circular arcs of a fixed radius are drawn from each point on the surface. The envelope of these circular arcs is the new suface after erosion/deposition.

Examples for the square wave structure are illustrated in figure 2.8.

The cusps and expansions of corners which occur are just a more simple form of these previous descrived bya general Y(θ) dependence. These are well known in optics - being called caustic curves.

An example of the use of masks is shown in figure 2.9. In figure 2.9a. the erosion is due to a ballistic process with a sputtering yield as shown in figure 2.2. The walls are never straight but take the shape of the characteristic locus curve due to the flux discontinuity at the edge of the mask and the "Snell's Law" requirement to match the speeds parallel to the interface. In figure 2.9b. the situation where a reactive gas component is introduced into the system causing undercutting of the mask is illustrated. In this example the area under the mask is subjected to an isotropic erosion rate, whereas the area between the masks is subjected to both a chemical and ballistic component. The shape of the walls under the mask consists of an arc of a circle directly underneath the mask joined smoothly to a straight line whose gradient can be calculated as shown in the figure caption.

Figure 2.8 shows that planarisation of square wave structures can be achieved by isotropic deposition on a square wave structure. In the industrial manufacturing process it is required (1) to make the planarised surface as flat as possible while (2) maintaining the submerged structure near to the surface. This is effectively a non-linear optimisation problem. The parameters which can be adjusted are combinations of the erosion/deposition rate. In all cases it was found that planarisation could be best achieved by first deposition and then etching back the surface after deposition. A combination of the two processes is always inferior.

2.2 Erosion in Three Dimensions

General Non-linear First Order Equations

Consider a function φ of n variable $x_1, \ldots, x_n$ whose derivative $p_i = \partial\varphi/\partial x_i$. If φ satisfies a differential equation of the form

$$H(p, \varphi, x) = 0 \tag{2.14}$$

then characteristics can be defined which reduce equation (2.14) to a series of ordinary differential equations as in two dimensions. Consider a curve in n-space C defined by $x = x(\lambda)$ along C

$$\frac{d\varphi}{d\lambda} = \sum_j p_j \frac{dx_j}{d\lambda} \tag{2.15}$$

$$\frac{dp_i}{d\lambda} = \sum_j \frac{\partial^2\varphi}{\partial x_i \partial x_j} \frac{dx_j}{d\lambda} \tag{2.16}$$

Differentiating (2.14) with respect to x_i yields

$$\sum_j \frac{\partial^2\varphi}{\partial x_i \partial x_j} \frac{\partial H}{\partial p_j} + \frac{\partial H}{\partial \varphi} \frac{\partial \varphi}{\partial x_i} + \frac{\partial H}{\partial x_i} = 0 \tag{2.17}$$

If the curve C is now chosen to be defined by

$$\frac{dx_i}{d\lambda} = \frac{\partial H}{\partial p_i} \tag{2.18}$$

then (2.16) becomes

$$\frac{dp_i}{d\lambda} = - p_i \frac{\partial H}{\partial \varphi} - \frac{\partial H}{\partial x_i} \tag{2.19}$$

and (2.15) becomes

$$\frac{d\varphi}{d\lambda} = p_j \frac{\partial H}{\partial p_j} \tag{2.20}$$

Thus along curves C defined by

$$\frac{dx_i}{dx} = \frac{\partial H}{\partial p_i}$$

the function φ and its derivatives can be determined by solving the O.D.E.'s. (2.19) and (2.20).

Application to Three Dimensional Surface Erosion

Consider a surface the equation of which at time t is $S(r,t) = 0$, where r is the position vector of a point on the surface, see figure 2.10. If an infinitesimal surface element moves along its normal direction with a speed c to the point $r + \delta r$ at time $t + \delta t$, then expansion to first order of $S(r+\delta r, t+\delta t) = 0$ gives

$$S_t \pm (c\cdot\nabla)S = 0 \qquad (2.21)$$

where c is the velocity of the surface in the normal direction and subscripts denote differentiation. If equation of the surface is written in the alternative form

$$S(r,t) = \sigma(r) - t = 0 \qquad (2.22)$$

then use of (2.21) shows that

$$\sum_{i=1}^{3} \sigma_{x_i}^2 = c^{-2} . \qquad (2.23)$$

where $r = (x_1, x_2, x_3)$, n is the unit normal to the surface and t_1, the unit tangent in the direction of n x k. Hence

$$n = \left[\sigma_{x_1}, \sigma_{x_2}, \sigma_{x_3}\right] / \left[\sum_{i=1}^{3} \sigma_{x_i}^2\right]^{\frac{1}{2}} \qquad (2.24)$$

$$t_1 = \left[\sigma_{x_3}, -\sigma_{x_1}, 0\right] / \left[\sum_{i=1}^{2} \sigma_{x_i}^2\right]^{\frac{1}{2}} . \qquad (2.25)$$

$$\cos\theta = -\sigma_{x_3} / \left[\sum_{i=1}^{3} \sigma_{x_i}^2\right]^{\frac{1}{2}} . \qquad (2.26)$$

$$\cos\psi = \sigma_{x_2} / \left[\sum_{i=1}^{2} \sigma_{x_i}^2\right]^{\frac{1}{2}} . \qquad (2.27)$$

Equations (2.21) and (2.23) are for many conditions of the normal speed c first-order non-linear equations: a method of solution exists by using (2.23) and (2.26) to define a Hamiltonian

$$H = 1 - \frac{c^2\sigma_{x_3}^2}{\cos^2\theta} = 0 . \qquad (2.28)$$

Equation (2.28) can also be given in the form

$$H = 1 - g^2\sigma_{x_3}^2 = 0 . \qquad (2.29)$$

where g is the effective wavefront velocity in the k direction. Attention will now be restricted to the case where

$$g = g(x_1, x_2, x_3, t, \theta, \psi) .$$

The physical interpretation of this assumption is that if g is a function of x_1, x_2 or x_3 then the wavefront propagation is appropriate to an inhomogeneous system: if g is a function of time then the propagation is in a time dependent system: if g is a function of θ or ψ the system is anisotropic.

Following the characteristic method, the equations of the characteristics curves are given by

$$dx_1/dt = - g_\theta \cos\theta \sin\psi - g_\psi \cot\theta \cos\psi$$

$$dx_2/dt = - g_\theta \cos\theta \cos\psi - g_\psi \cot\theta \sin\psi \qquad (2.30)$$

$$dx_3/dt = \cos\theta \sin\theta \, g_\theta - g$$

and the invariant relations holding along the characteristics

$$d\sigma_{x_i}/dt = (1/g)(\sigma_{x_i} g_t + g_{x_i}) \quad . \quad i = 1,2,3 \; . \qquad (2.31)$$

Without being specific about the processes which lead to surface motion some general results can be derived from (2.30) and (2.31).

First, if g is not a function of x_i or t then $g = g(\theta,\psi)$ and (2.31) shows that the σ_{x_i} are constants along the characteristics. Thus from equations (2.26), (2.27) and (2.30) the characteristics are straight lines and relate to constant surface orientation values. The characteristics are also lines of constant surface orientation if g is time dependent but spatially invariant. In this case (2.31) integrates to give

$$\sigma_{x_i} = k_i g \quad . \quad i = 1,2,3 \; .$$

where k_i are constants: from (2.26) and (2.27) the characteristics are still lines of constant θ and ψ. However, since t appears explicitly on the right hand side of (2.30) the characteristics are no longer straight lines. In the most general case, the characteristics are lines of constant surface orientation provided

$$\sigma_{x_i} g_t + g_{x_i} = 0 \quad . \quad i = 1,2,3 \; .$$

If g is temporally invariant, then provided that $g_{x_j} = 0$ for some j, σ_{x_j} = constant. In particular, if $g_{x_3} = 0$, then (2.29) identifies the characteristics of those lines of constant g. Finally, if g has a stationary value with respect to θ and ψ, then (2.30) show that the characteristics

propagate in the x_3 direction.

It should be noted that when the σ_{x_i} are constants, and the characteristics are those of constant orientations θ and ψ, then only values of θ and ψ between the lower (θ_L, ψ_L) and upper (θ_u, ψ_u) bounds presents on the initial surface are present at some time later. If the σ_{x_i} vary with time, this is not the case and new surface orientations are generated.

Equations (2.30) and (2.31) in general, form a set of six non-linear ordinary differential equations which can be solved to track the motion of a surface if the function g is completely specified. A given initial surface is prescribed digitally in terms of its coordinates in space (x_1, x_2, x_3) and its surface orientations $(\sigma_{x_1}, \sigma_{x_2}, \sigma_{x_3})$. The method is equivalent to that of ray tracing in geometric optics. Care must be taken when solving these equations, because an initially continuous surface can develop surface gradient discontinuities (shocks) due to the intersection of the characteristics or rays and the generation of a multivalued solution. This aspect is considered in more detail in the next section.

Development and Motion of Points and Edges

When rays intersect, the surface σ remains continuous but orientation or slope discontinuities occur at the point of intersection. The angles θ and ψ which bound the edge can change continuously as the surface continues to evolve. The formation of such a slope discontinuity is shown schematically in figure 2.11a, b. It is possible that a two dimensional edge or an apex (point) could form depending on the precise conditions on the initial surface.

Consider the time t at which edge formation first occurs from an initially continuous surface, as a result of the intersection of characteristics from the neighbouring points $r_0, r_0 + \delta r_0$. These points transpose to the

same point in space r_f at time t. If the characteristics velocities are denoted from 2.30 by $v = (v_1, v_2, v_3)$, where

$$v_i(r,t) = dx_i/dt \ . \quad i = 1,2,3 \ .$$

then

$$r_0 + \int_0^t v(r,t')dt' = r_0 + \delta r_0 + \int_0^t v(r+\delta r, t')dt' \ .$$

i.e.
$$\delta r_0 = \int_0^t (\delta r(t') \cdot v(r,t')dt' \ . \qquad (2.32)$$

Here $\delta r(t')$ is the vector separation of these points on the surface at time t', and

$$\delta r(t') = \delta r_0 + \int_0^{t'} (\delta t(t'') \cdot \nabla) v(r,t'')dt'' \ . \qquad (2.33)$$

In principle, (2.32) and (2.33) can be solved to give the time at which a surface gradient discontinuity initially forms, provided the point r_0 is known. This point must also be determined by seeking the minimum value of t from (2.32) and (2.33) as r_0 ranges over all the values on the initial surface. After the edge has formed, the intersection of characteristics from neighbouring points no longer defines the edge position. The edge propogates as a result of the intersection of characteristics from non-neighbouring points collapsing into the edge.

In practice, (2.32) and (2.33) are rarely used to determine edge formation. This is most easily accomplished by numerical solution of (2.30) and (2.31) and by direct ray tracing.

Linear Characteristics

Equations (2.32) and (2.33) can be simplified in the important case of linear characteristics where $g = g(\theta, \psi)$ only, giving

$$\delta r_0 = -t(\delta r_0 \cdot \nabla) v$$

$$\delta r_0 = (\delta x_{10}, \delta x_{20}, \delta x_{30}) \qquad (2.34)$$

or letting

$$\sum_{j=1}^{3} \left[\delta_{ij} + t \frac{\partial v_i}{\partial x_j}\right] \delta x_{j0} = 0 \ . \quad i = 1,2,3 \ . \qquad (2.35)$$

Equations (2.34) indicates that the edge formation occurs when

$$t = -\left[dv_i/dx_{i0}\right]^{-1} \ . \quad i = 1,2,3 \ . \qquad (2.36)$$

However, the characteristics velocities depend on g, which is defined in terms of θ and ψ, and so (2.34) can be written

$$\delta x_{i0} = t\left[v_{i\theta}\delta\theta + v_{i\psi}\delta\psi\right] \ . \quad i = 1,2,3 \ . \qquad (2.37)$$

Since δx_{30} is explicitly defined in terms of δx_{10} and δx_{20} on the initial surface, the third component of (2.37) is not independent of the first two components. By using the result

$$\delta x_{i0} = x_{i\theta}\delta\theta + x_{i\psi}\delta\psi \ . \quad i = 1,2,3 \ .$$

where $x_{i\theta}, x_{i\psi}$ are the partial differential coefficients of x_i on the initial surface with respect to θ and ψ, the first two components of equations (2.37) are given as

$$\left[\begin{bmatrix} x_{1\theta} & x_{1\psi} \\ x_{2\theta} & x_{2\psi} \end{bmatrix} - t\begin{bmatrix} v_{1\theta} & v_{1\psi} \\ v_{2\theta} & v_{2\psi} \end{bmatrix}\right]\begin{bmatrix} \delta\theta \\ \delta\psi \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} . \qquad (2.38)$$

This has a solution for t given by

$$\det[X - tV] = 0 \ . \qquad (2.39)$$

where

$$X = \begin{bmatrix} x_{1\theta} & x_{1\psi} \\ x_{2\theta} & x_{2\psi} \end{bmatrix} , \quad V = \begin{bmatrix} v_{1\theta} & v_{1\psi} \\ v_{2\theta} & v_{2\psi} \end{bmatrix} .$$

This is a quadratic equation for t. The edge first forms at the minimum non-negative value of t given by (2.39), as θ and ψ range over all the values on the initial surface. For a smooth surface, t is given from (2.39) by $t = t(\theta,\psi)$ and the additional conditions

$$\partial t/\partial\theta = \partial t/\partial\psi = 0 \qquad (2.40)$$

are required to determine the values θ_0, ψ_0 of θ and ψ at which the edge first forms. Equation (2.39) also gives the orientations of the line elements on the initial surface which collapse to the edge. Having determined t, the line elements on the initial surface which collapse to the edge are given by

$$\frac{d\theta}{d\psi} = - \frac{(x_{1\psi} - tv_{1\psi})}{x_{1\theta} - tv_{1\theta}} \left| = \frac{x_{2\psi} - tv_{2\psi}}{x_{2\theta} - tv_{2\theta}} \right| . \qquad (2.41)$$

If all the elements of the matrix X - tV are zero, then line elements of all orientations near $x_{10}(\theta_0,\psi_0)$ collapse to the edge and a vertex rather than a finite edge, forms. Some computed erosion profiles of symmetric and elliptic hummocks are shown in figure (2.12)

3. THE ELECTROSTATIC GUIDANCE OF CHARGED PARTICLE BEAMS

The beams of particles which are used for pattern delineation need to be guided towards the etched material. In the case of ion beam lithography, the beams also need to be focussed to a point. Both magnetic and electrostatic forces are used for the general guidance of charged particle beams but for ion beam lithography electrostatic focussing is the only practical method.

This section describes the numerical techniques necessary to solve the equations of motion of the charged particles. An optimisation problem will be described whereby the spot size of a microfocussed ion beam can be minimised by optimisation of the chromatic aberration of the electrostatic lenses. Thus, the possibility exists that the size of VLSI structures can be reduced even further than present limits.

The Mathematical Problem

A beam guidance system consists of a number of metal electrodes (lenses) held at different potentials. In order the calculate the electric forces on the charged particles, Laplace's (or sometimes Poisson's) equation is solved. The particle trajectories are then calculated by solving Newton's laws of motion.

Thus the mathematical statement of the problem is the solution of

$$\nabla^2\varphi = 0$$

$$m \frac{d^2 r}{dt^2} = - e\nabla\varphi \qquad (3.1)$$

where φ is the electric potential, m is the mass and e the charge of a particle, and r(t) is the vector position of the particle at time t.

For the case of dense beams where the charge of the particles in the beam, contributes significantly the electric field, Poisson's equation

replaces Laplace's equation and a charge conservation equation is added.

The system to be solved becomes

$$\nabla^2\varphi = \rho$$

$$m\frac{d^2r}{dt^2} = -e\nabla\varphi \qquad (3.2)$$

$$\mathrm{div}\left[\rho\frac{dr}{dt}\right] = 0$$

where ρ is the density of the space charge.

The Numerical Solution of Laplace's Equation for Electrostatic Lenses

The problem is usually stated with Dirichlet (φ specified) boundary conditions, where the electrodes are held at certain specified potentials (which can be adjusted). Because the geometry is extremely complicated, a numerical method of solution must be used which can deal with complicated boundary shapes. It was found convenient to do this using a variable mesh finite difference scheme (although finite element methods are often used for similar problems with complicated boundaries). However, a data structure was devised, which allowed the finite difference grid to be constructed of arbitrally sized sub-grids which are defined by a special code as part of the input file (in algol 68). This approach allows fine and course grids to be placed in regions of high and low potential difference respectively, allowing easy modification of the input file with no restructuring of previously input nodal ponts, when new ones are added. This enables numerical experiments to be carried out by altering the grid sizes in order that the potentials and fields may be accurately determined in an iterative way. Each grid point is automatically assigned a unique address and is related to its neighbouring points by the computational modelcule shown in

figure 3.1. With this form, a Tayor series expansion about φ_p yields

$$\varphi_1 = \varphi_p + h\frac{\partial\varphi_p}{\partial r} + \frac{h^2}{2!}\frac{\partial^2\varphi_p}{\partial r^2} + \frac{h^3}{3!}\frac{\partial^3\varphi_p}{\partial r^3} + \ldots$$

and

$$\varphi_3 = \varphi_p - h\theta_1\frac{\partial\varphi_p}{\partial r} + \frac{h^2\theta_1^2}{2!}\frac{\partial^2\varphi_p}{\partial r^2} - \frac{\theta_1^3h^3}{3!}\frac{\partial^2\varphi_p}{\partial r^3} + \ldots$$

This yields

$$\frac{\partial\varphi_p}{\partial r} = \frac{\varphi_1\theta_1^2 - \varphi_3 - (1-\theta_1^2)\varphi_p}{h\theta_1(1+\theta_1)} + O(h^2)$$

$$\frac{\partial^2\varphi_p}{\partial r^2} = \frac{\varphi_1\theta_1 + \varphi_3 - (1+\theta_1)\varphi_p}{\frac{1}{2}h^2\theta_1(1+\theta_1)} + O(h) \qquad (3.3)$$

Similarly

$$\frac{\partial^2\varphi_p}{\partial z^2} = \frac{\varphi_2\theta_2 + \varphi_4 - (1+\theta_2)\varphi_p}{\frac{1}{2}k^2\theta_2(1+\theta_2)} + O(h)$$

These numerical representations of the derivatives are then substituted directly into Laplace's equation in cylindrical polar coordinates (for axially symmetric systems) given by

$$\frac{\partial^2\varphi}{\partial r^2} + \frac{1}{r}\frac{\partial\varphi}{\partial r} + \frac{\partial^2\varphi}{\partial z^2} = 0 \qquad (3.4)$$

Substituting, the numerical representation of the derivatives (3.3) into equation (3.4) yields,

$$\varphi_p = (\varphi_3(1-W_1) + \varphi_1\theta_1(1+\theta_1W_1) + (\varphi_2\theta_2+\varphi_4)/W_2)/D \qquad (3.5)$$

where $W_1 = \frac{h}{2r_p}$; $W_2 = k^2\varphi_2(1+\theta_2)/h^2\theta_1(1+\theta_1)$

and $$D = (1+\theta_1)(1-W_1(1+\theta_1)) + (1+\theta_2)/W_2 \ .$$

For n interior grid points, a set of n linear equations is obtained which forms a sparse n x n matrix, and the solution for the potentials at the grid points is consequently the most time consuming part of the programme. The real problem requires obtaining the solution in a minimum time while conserving computer storage space.

The finite difference form for the partial derivatives means that the value of φ at any given node is linked to the φ values at the four neighbouring nodes, only. This means that the set of linear equations for the values of φ at the nodes is 'sparse'. In that the matrix for these equations has many zeros as elements.

Initially, a sparse matrix solver based on a standard Gaussian elimination was used with an n x 5 matrix and proved fast, but required an overall matrix size declaration of approximately n x (n/6)(n > 100) to allow for the 'filling-in' of rows in the matrix at the elimination stage. Since only a few rows towards the centre of the final elimination matrix grow to this size, the method results in wasted computer storage. However, the use of sparse matrix methods to reduce what would be effectively an n x n problem allows a considerable saving to be made, and as a result many more points can be used than would be otherwise possible on a small machine. In order to overcome the problem of the 'filling-in' of rows, each row was then represented as a flexible array. This allowed each individual row to expand or contract exactly according to the insertion of a new element or the elimination of an existing one, but required data transportation to and from a heap storage while the arrays were manipulated. Minimal storage was attained by this method but the vast increase in computing time consequently outweighed this advantage. Several different block methods were also considered, which involved using the

sparse matrix solver on multiple segments, each of a fixed array size. In theory, this put no limit on the computer store available, but again proved slow in, the external linking of the individual segments.

The most economic method tested, using a Honeywell Multics computer, was to off-line the set of equations to a data file and then use a Fortran version of the sparse matrix solver, utilizing a 'very large array' compiler option to handle the large arrays with speed and precision. The calculated potentials were then fed back to the main programme to complete the trajectory analysis. It was also noted that the performance of a sparse matrix solver can be improved considerably by reducing the number of row interchanges and ensuring diagonal dominance, but this requires a certain sophistication by the user of the programme when setting up the original grid.

Example numerical solution of Laplace's equation on a square grid

Consider a square region, where two adjacent sides of the square are earthed, and the other sides are held at a potential of 100 V. In order to solve the problem by hand, the interior region is divided by a regular grid, with nine internal nodes, see figure 3.2. In two dimensions, Laplace's equation is

$$\frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} = 0 \ .$$

With a regular grid equation 3.5 becomes

$$\varphi_p = \frac{1}{4}\left[\varphi_3 + \varphi_1 + \varphi_2 + \varphi_4\right] \ .$$

and a system of nine linear equations results for the internal values of φ

$$
\begin{aligned}
0 + 100 + \varphi_{12} + \varphi_{21} - 4\varphi_{11} &= 0 \\
\varphi_{11} + 100 + \varphi_{13} + \varphi_{22} - 4\varphi_{12} &= 0 \\
\varphi_{12} + 100 + 100 + \varphi_{23} - 4\varphi_{13} &= 0 \\
0 + \varphi_{11} + \varphi_{22} + \varphi_{31} - 4\varphi_{21} &= 0 \\
\varphi_{21} + \varphi_{12} + \varphi_{23} + \varphi_{32} - 4\varphi_{22} &= 0 \\
\varphi_{22} + \varphi_{13} + 100 + \varphi_{33} - 4\varphi_{23} &= 0 \\
0 + \varphi_{21} + \varphi_{32} + 0 - 4\varphi_{31} &= 0 \\
\varphi_{31} + \varphi_{22} + \varphi_{33} + 0 - 4\varphi_{32} &= 0 \\
\varphi_{32} + \varphi_{23} + 100 + 0 - 4\varphi_{33} &= 0 \,.
\end{aligned}
$$

Note that with this computational molecule, the corner points where φ is non-uniquely defined does not enter into the calculation. Note symmetry implies $\varphi_{12} = \varphi_{23}$, $\varphi_{11} = \varphi_{33}$ and $\varphi_{21} = \varphi_{32}$ so the number of equations can be reduced to six. The equations can be written

$$
\begin{bmatrix}
-4 & 1 & 0 & 1 & 0 & 0 \\
1 & -4 & 1 & 0 & 1 & 0 \\
0 & 2 & -4 & 0 & 0 & 0 \\
1 & 0 & 0 & -4 & 1 & 1 \\
0 & 2 & 0 & 2 & -4 & 0 \\
0 & 0 & 0 & 2 & 0 & -4
\end{bmatrix}
\begin{bmatrix}
\varphi_{11} \\ \varphi_{12} \\ \varphi_{13} \\ \varphi_{22} \\ \varphi_{22} \\ \varphi_{31}
\end{bmatrix}
=
\begin{bmatrix}
-100 \\ -200 \\ -200 \\ 0 \\ 0 \\ 0
\end{bmatrix}
$$

and solution by Gaussian elimination yields

$$
\begin{aligned}
\varphi_{11} &= \varphi_{33} = 50.00 \\
\varphi_{12} &= \varphi_{23} = 71.43 \\
\varphi_{13} &= 85.72 \\
\varphi_{21} &= \varphi_{32} = 28.57 \\
\varphi_{22} &= 50.00 \\
\varphi_{31} &= 14.29 \,.
\end{aligned}
$$

This example illustrates many features of the more complex calculations carried out in the real problem.

Trajectory analysis and analytical comparisons

The paths of charged particles can be traced using the equation of motion in either time dependent of time independent form. For time dependence we solve the equations

$$\frac{dr}{dt} = q$$

and

$$\frac{dq}{dt} = \frac{e}{m} \text{ grad } \varphi \qquad (3.6)$$

where q is the particle velocity. These differential equations are solved by a Kutta–Merson routine with automatic step adjustment to tally with the localized grid spacing, and yield the values r, z and u and v, the components of velocity in the r, z directions respectively. In general, the particle trajectories are fitted locally to a five-point quadratic model of the form

$$ar^2 + bz^2 + cr + dz + e = \varphi \qquad (3.7)$$

where the five coefficients are determined by substituting into the r, z and φ values of the nearest grid point and its four corresponding neighbours to obtain five linear equations. A special form of equation (3.5) is used near the axis of symmetry where the boundary condition $\partial\varphi/\partial r = 0$ means that the nearest grid point does not possess a full set of neighbouring points. The use of a fine axial strip in this region ensures that the time spent within this narrow area is kept to a minimum, since the approximation for φ is least accurate here.

By considering the motion along a trajectory it is possible to obtain a modification to equation (3.4) in time independent form. Equation (3.6) integrates to give an energy conservation equation

$$\frac{1}{2}q^2 - \frac{e}{m}\left[\varphi - \bar{\varphi}_0 - E_0\right] = 0 \qquad (3.8)$$

where E_0 is the initial energy of the particle when the potential is $\bar{\varphi}_0$. If the trajectory of the particle is given by $r = r(t)$ then

$$\frac{dr}{dt} = \frac{dr}{dz}\cdot\frac{dz}{dt} \quad \text{and} \quad \frac{d^2r}{dt^2} = \frac{d^2r}{dz^2}\cdot\left[\frac{dz}{dt}\right]^2 + \frac{dr}{dt}\cdot\frac{d^2z}{dt^2} \qquad (3.9)$$

combining equations (3.6), (3.8) and (3.9) gives

$$\frac{d^2r}{dz^2}\left[\varphi - \bar{\varphi}_0 - E_0\right] + \frac{1}{2}\left[1 + \left[\frac{dr}{dz}\right]^2\right]\left[\frac{\partial\varphi}{\partial z}\cdot\frac{dr}{dz} - \frac{\partial\varphi}{\partial r}\right] = 0 \qquad (3.10)$$

Equation (3.8) requires the solution of only two first order simultaneous ordinary differential equations to give r, z and the gradient dr/dz. By the conservation of energy, and using the trajectory gradient, it is then possible to calculate the components of velocity in the r and z directions.

By either method, a full range of particle paths are obtained and these are then off-lined along with the grid coordinates to an external plotting programme. This then allows particle paths with differing initial energies and emission angles to be compared directly for the same electric field and on the same grid plot.

In order to illustrate the method on a simple example for which an analytic solution exists, a variable grid is given below for two spherically symmetric electrodes (Figure 3.3) with the centre sphere at a potential V_1 and the outer sphere earthed.

For this spherical system, the field lines are radial and the variable rectangular grid is not the most physical way of representing in finite difference form, Laplace's equation for the potential. However, even for this problem, with a coarse mesh, the method proves accurate to within 1% when calculated φ values are compared with the exact ones given by

$$\varphi = \frac{k_1}{r} + k_2 \qquad (3.11)$$

where k_1, k_2 are constants.

Furthermore, the accuracy of the tracking for this model can be checked by solving the equation of motion in spherical polar coordinates analytically to yield

$$\theta = 2\tan^{-1}\left[\frac{(F^2+E^2-G^2)^{\frac{1}{2}} + F}{G + E}\right] \qquad (3.12)$$

where

$$F = -\frac{\alpha V_1^2}{2}\sin 2\psi \,, \qquad E = \alpha V_1^2 \sin\psi^2 + \beta \,,$$

$$G = \frac{\alpha^2 V_1^2 \sin\psi^2 + \beta}{R}$$

and

$$\beta = \frac{e}{m} V_1 \frac{\alpha R}{R - \alpha}$$

where R is the radius of the outer sphere, α the radius of the inner sphere, V_1 the initial velocity, ψ the initial angle and e/m the usual charge-to-mass ratio.

The values of θ, the angle subtended at the centre of the inner sphere by particles hitting the outer sphere were to within 1% of the exact results when calculated using equation (3.12).

Design Applications

Individual Lenses

The important characteristics of an electrostatic lens are its magnification, it focal length and the magnitude of its spherical and chromatic aberrations. The spherical aberration arises because charged particle trajectories at difference emission angles are focussed at different points. The chromatic aberration occurs because of an energy spread in the beam, which focusses particles with the same initial trajectory, but with different energies at different points.

The parameters involved in the lens calculations are described in Figure 3.4 (a). For this system where V_2/V_1 = 20, A/D = 0.5 and T/D = 0.05, the focal length F/D is calculated to be in good agreement with the results of Harting and Read[4]. For this problem a fine mesh was constructed between the central and outer elements of the lens where the field is most rapidly changing, with a coarser mesh away from these regions.

Figure 3.4 (b) shows the effect of both the spherical and chromatic aberration of this lens.

The spherical aberration is manifested by rays of different initial angles focusing at different positions. The chromatic aberration is displayed by tracking rays with initial energies of V_1 and $V_1 + \epsilon$, but with the same initial emission angles. Figure 3.4 (b) indicates virtually no difference in

the image side focal length f with ϵ = 500 V but a large shift is indicated in the location of the gaussian image plane, by the distance δ indicated in the diagram.

It is thus possible to compute the performance of individual lenses before they are used in a column, thus saving the engineer considerable time and effort.

Lithography

The possibility of using low energy scanning ion beams of high resolution for either direct write lithography or mask repair has attracted much attention over recent years. The motion of the beam is controlled by electrostatic deflector plates but the principal criterion to be achieved is the production of the smallest focussed ion beam spot size, consistent with an acceptable current density required for the application. The size of these microfocussed beams is limited by the chromatic aberration of the system and the column designer must insert a blanking aperture into the column at an appropriate point and at an appropriate diameter in order that the aberration discs be reduced and only ions which are emitted with a small extraction angle from the sourse, pass through the system. For such a column only paraxial rays need be considered (ion trajectories close to the axis) and some analytical approximations are possible in this case.

Near the axis it is assumed that the potential φ can be expanded in a power series in r^2

$$\varphi(r,z) = \varphi_0(z) + \varphi_1(z)r^2 + \varphi_2(z)r^4 + \ldots$$

so that

$$\frac{\partial^2\varphi}{\partial r^2} + \frac{1}{r}\frac{\partial\varphi}{\partial r} = 4\varphi_1 + 16\varphi_2 r^2 + \ldots \; ; \; \frac{\partial^2\varphi}{\partial z^2} = \varphi_0'' + \varphi_1'' r^2 + \varphi_2'' r^4 - \ldots \qquad (3.13)$$

where $\varphi_0(z)$ is the potential along the axis. Substituting into equation (3.4), and equating coefficients of r, gives

$$\varphi(r,z) = \varphi_0(z) - \varphi_0''(z)\,\frac{r^2}{2^2} + \varphi_0^{iv}(z)\,\frac{r^4}{2^2 4^2} - \ldots \qquad (3.13)$$

Substituting equation (3.13) into (3.10) and assuming dt/dz is small gives

$$\frac{d^2r}{dz^2} + \frac{dr}{dz}\left[\frac{\varphi_0'}{2\varphi_0}\right] + \frac{r}{4}\,\frac{\varphi_0''}{\varphi_0} = 0 \qquad (3.14)$$

where the notation for φ_0 has slightly changed being the quantity $\varphi - \varphi_0 - E_0$, evaluated on the axis.

The focal point of the lens systems corresponds to the first value of z_F for which r is zero. The chromatic aberration is calculated by the effect on z_F of a small change in φ_0. For a perturbation where φ_0 increases by a small amount ϵ, equation (3.14) becomes

$$r'' + \frac{1}{2}\,r'\left[\frac{\varphi_0'}{\varphi_0 + \epsilon}\right] + \frac{1}{4}\,r\left[\frac{\varphi_0''}{\varphi_0 + \epsilon}\right] = 0$$

expanding this equation in powers of ϵ gives to first order

$$r'' + \frac{1}{2}\,r'\left[\frac{\varphi_0'}{\varphi_0}\right] + \frac{1}{4}\,r\left[\frac{\varphi_0''}{\varphi_0}\right] = \epsilon\left[\frac{1}{2}\,\frac{\varphi_0'}{\varphi_0^2}\,r' + \frac{1}{4}\,\frac{\varphi_0''}{\varphi_0^2}\,r\right].$$

If the trajectory is of the form

$$r = r_0 + \epsilon\, r_1 + \epsilon^2\, r_2 + \ldots$$

then equating powers of ϵ gives

$$r_1'' + \frac{1}{2}\,r_1'\left[\frac{\varphi_0'}{\varphi_0}\right] + \frac{1}{4}\,r_1\left[\frac{\varphi_0''}{\varphi_0}\right] = \left[\frac{1}{2}\,\frac{\varphi_0'}{\varphi_0^2}\,r_0' + \frac{1}{4}\,\frac{\varphi_0''}{\varphi_0^2}\,r_0\right]$$

This equation can be solved by the method of variation of parameters. Consider first the solution of equation (3.14). The general solution can be written as a linear combination of two independent solutions X, Y passing through the point z_0, with the conditions, $X(z_0) = 0$, $X'(z_0) = 1$, $Y(z_0) = 1$, $Y'(z_0) = 0$. The focal point z_F is defined by $X(z_F) = 0$.

Writing

$$r_1 = \alpha(z)\,X(z) + \beta(z)\,Y(z)\ .$$

then the conditions for the variation of parameters method gives

$$\alpha'(z)\,X(z) + \beta'(z)\,Y(z) = 0$$

$$\alpha'(z)\,X'(z) + \beta'(z)\,Y(z) = \gamma\left[\frac{1}{2}\,\frac{\varphi_0'}{\varphi_0^2}\,X' + \frac{1}{4}\,\frac{\varphi_0''}{\varphi_0}\,X\right]. \qquad (3.15)$$

Here, the ray r_0 is given by $\gamma X(z)$, see figure 3.5.

The radius of the aberration disc when $z = z_I$ is just

$$r_1 = \alpha(z_F)X(z_F) + \beta(z_F)Y(z_F) = \beta(z_F)Y(z_F)\ . \qquad (3.16)$$

Solving for β' from equation (3.15) gives

$$\beta'(z) = \gamma\left[\frac{1}{2}\,\frac{\varphi_0'}{\varphi_0^2}\,X' + \frac{1}{4}\,\frac{\varphi_0''}{\varphi_0}\,X\right]X/(X'Y - XY') \qquad (3.17)$$

However

$$X'' + \frac{1}{2}\,\frac{\varphi_0'}{\varphi_0}\,X' + \frac{\varphi_0''}{4\varphi}\,X = 0 \qquad (3.18)$$

$$Y'' + \frac{1}{2}\,\frac{\varphi_0'}{\varphi_0}\,Y' + \frac{\varphi_0''}{4\varphi}\,Y = 0 \qquad (3.19)$$

Multiplying (3.18) by Y (3.19) by X, subtracting gives

$$Y \frac{d}{dz}\left[\varphi_0^{\frac{1}{2}} X'\right] - X \frac{d}{dz}\left[\varphi_0^{\frac{1}{2}} Y'\right] = 0 .$$

i.e. $$\frac{d}{dz}\left[\varphi^{\frac{1}{2}}\left[YX' - XY'\right]\right] = 0 .$$

Integrating between z_0 and z gives

$$YX' - XY' = \sqrt{\frac{\varphi_0(z_0)}{\varphi_0}}$$

Thus equation (3.17) becomes

$$\beta'(z) = \gamma\left[\frac{1}{2}\frac{\varphi_0'}{\varphi_0^2}\frac{X'}{X} + \frac{1}{4}\frac{\varphi_0'}{\varphi_0}\right] X^2 \sqrt{\frac{\varphi_0}{\varphi_0(z_0)}}$$

and hence from equation (3.16), the radius of the aberration disc is given by

$$r_c = \frac{\gamma\epsilon\, Y(z_F)}{\varphi_0^{\frac{1}{2}}(z_0)} \int_{z_0}^{z_F} \left[\frac{1}{2}\frac{\varphi_0'}{\varphi_0^2}\frac{X'}{X} + \frac{1}{4}\frac{\varphi_0'}{\varphi_0^2}\right] X^2 \varphi_0^{\frac{1}{2}}\, dz . \qquad (3.20)$$

A similar formula can be established for the radius of the spherical aberration disc. This is

$$r_s = \gamma^3 Y(z_F)\varphi_0^{\frac{1}{2}}(z_0)/64 \int_{z_0}^{z_F} \varphi_0^{\frac{1}{2}}\left[X^4(3T^4 - 5T'^2 - \frac{11}{2}T^2T' + 4X^3X'TT'\right]dz \qquad (3.21)$$

where

$$T = \frac{\varphi_0'}{\varphi} .$$

Equations (3.20) and (3.21) indicated that whatever the values of ϵ, it is alway possible to choose γ small enough so that $r_s < r_c$.

A typical lithography column is shown schematically in figure 3.6. It consists of a liquid metal ion source, an extractor at a potential difference

of 5 kV and two lenses, the first controls the spread of the beam and the second lens focuses the beam to a point. The chromatic aberration of this column has been calculated in various modes of operation but typically for V_1 = 15 kV, V_2 = 30 kV, V_3 = 15 kV, then the focus occurs at z_F = 457 mm, with a column demagnification of .98, C_s = 2772 mm and C_c = 32.4 mm. The spherical aberration coefficient C_s and the chromatic aberration coefficient C_c are defined by

$$r_c = C_c\gamma \frac{\epsilon}{\varphi_0(z_0)}$$

$$r_s = Y(z_F)C_s\,\gamma^3$$

The value of X,Y at the aperture are given by X = 34.7 mm Y = 1.15 mm. Further calculations show that for a spot size limited by the size of the chromatic aberration disc, of 10^{-8} m at the focal point, the aperture radius must be 2.6×10^{-6} m and for a typical ion source which gives a current density of 20 μA/st.rad., the beam current at the point of impact with the sample is $.16 \times 10^{-12}$ A.

The corresponding radius of the spherical aberration disc is $= 10^{-12}$ m, and only becomes of comparable importance with the chromatic aberration disc when microfocussed spots of the order of 10^{-7} m or larger are required.

The currents generated by this technique of microfocussing are generally insufficient for large scale production of VLSI devices, since the etching of the patterns would take too long. However, the technique has potential for the manufacture or repair of masks or specialist extremely small chip production.

Since the chromatic aberration of the column is the principle factor which limits the spot size, a constrained optimisation problem was formulated in order to determine the optimum condition for as small a spot

as possible consistent with as large a current as possible.

Minimisation of Chromatic Aberration Coefficients

First equation (3.14) is transformed using the substitution $R = r\varphi_0^{1/4}$, giving

$$\frac{d^2R}{dz^2} + \frac{3}{16}\left[\frac{\varphi_0'}{\varphi}\right]^2 R = 0 \qquad (3.22)$$

Equations (3.18) and (3.19) can then be written as the set

$$\frac{dx_1}{dt} = x_2 \qquad (a)$$

$$\frac{dx_2}{dt} = -\frac{3}{16}\left[\frac{\dot{u}}{u}\right]^2 x_1 \qquad (b)$$

$$\frac{dx_3}{dt} = x_4 \qquad (c) \qquad (3.23)$$

$$\frac{dx_4}{dt} = -\frac{3}{16}\left[\frac{\dot{u}}{u}\right]^2 x_3 \qquad (d)$$

and $C_c = x_5(T)$. where

$$\frac{dx_5}{dt} = \frac{3\sqrt{5}}{16\sqrt[4]{60}}\, x_3(T) x_1^2 u^{-\frac{1}{4}}(T)\left[\frac{\dot{u}}{u}\right]^2 \qquad (e)$$

with the initial conditions

$$u(0) = 55,\ x_1(0) = 0,\ x_2(0) = u^{\frac{1}{4}}(0) = x_3(0);\ x_4(0) = \frac{\dot{u}(0)}{4u^{\frac{1}{4}}(0)}$$

and boundary conditions

$$x_5(0) = 0$$

and constraint $0 < u(t) < 55$.

In the set of equations (3.23). the variable z is replaced by t and φ_0 by u. The focal point t = T is prescribed to be 480 mm. The optimisation problem is to minimise $x_5(T)$ (i.e. C_c) subject to certain constraints on the axial potential (u(t)). The constraints on the axial potential were formulated so that the column was consistent with 1) an initial beam accelerator, 2) an objective lens, 3) a projection lens and also that the potential difference in the system was less than 60 kV so that insulation breakdown did not occur.

The shape for the axial potential is given in figure 3.7. together with a definition of the adjustable parameters. The shape for the axial potential u(t). is assumed to consist of straight line, quadratic and cubic segments. One of the constraints $x_1(T) = 0$ was added to the objective function as a penalty function. This was because most of the optimisation methods tested, do not handle non-linear equality constraints. The objective function eventually chosen was

$$F = \left\{x_5(T)\right\}^2 + W\left\{x_1(T)\right\}^2$$

where W was a suitable weight, taken as 10.000 in this problem.

The problem thus became the minimisation of F, subject to

$$x_1 - x_0 \geq 0;\ x_2 - x_1 \geq 0;\ x_3 - x_2 \geq 0;\ x_4 - x_3 \geq 0;$$

$$0 \leq h_1 \leq 55,\ 0 \leq h_2 \leq 55,\ x_0, x_1, x_2, x_3, x_4 \geq 0$$

A Runge-Kutta method with fixed step length was used to integrate the equations.

Results

Three different optimisation methods were used. Two were direct search methods due to Box and Rosenbrock. The third was a quasi-Newton method. The results were all calculated with starting values of $x_0 = 10.0$, $x_1 = 20.0$, $x_2 = 60.0$, $x_3 = 410.0$, $x_4 = 450.0$, $h_1 = 40.0$, $h_2 = 35.0$. Other starting values were also tried, but gave no drastic improvement to the results.

The results presented in the following tables are based on values of x_5 (and hence F) where the factor of $\sqrt{5}$ in 3.23 (a) has been omitted.

1. Quasi Newton

No. of Function Evaluations	100	200	300	400	500
X1	0.0038	0.0001	-0.00004	0.0001	0.0001
X5	-4.7432	-4.7522	-4.75113	-4.7507	-4.7507
F	23.9473	22.5847	22.5734	22.5706	22.5706

This method used 14 function evaluation to approximate the gradient vector by finite-differences.

The best value of F = 22.5706 was attained when

$x_0 = 10.0$, $x_1 = 19.8623$, $x_2 = 60.0$, $x_3 = 421.50$

$x_4 = 449.88501$, $h_1 = 39.94308$, $h_2 = 34.9136$

2. Rosenbrock's Method of Rotating Coordinates

No. of Function Evaluations	100	200	300	400	500
X1	-0.0008	-0.00021	0.00002	0.00005	-0.00065
X5	-6.69076	-6.67047	-6.665818	-6.66407	-6.613035
F	44.83807	44.4997	44.43318	44.410189	43.7750862

Number of orthogonalisations for 500 function evaluations were 6.

For the best result of F = 43.7750862 at

$x_0 = 19.98$, $x_1 = 20.37$, $x_2 = 59.78$, $x_3 = 411.29$,

$x_4 = 449.24$, $h_1 = 40.38$, $h_2 = 35.48$

3. The Method of Box

No. of function evaluations	100	200	300	400	500	600
Xi	1.8929	0.0278	0.0278	0.0028	0.0102	0.0120
X5	-9.3964	-10.6215	-10.6215	-10.6668	-10.3387	-9.8260
F	35917.93	120.53	120.53	113.86	107.92	97.99

The best point obtained by this method after 600 function evaluations occurred when F = 97.99 at

$x_0 = 28.41$, $x_1 = 49.87$, $x_2 = 111.30$, $x_3 = 361.50$,

$x_4 = 419.55$, $h_1 = 37.70$, $h_2 = 33.42$

It can be seen that the quasi-Newton method outperforms the direct methods on this problem.

The conclusion to be drawn from this is that the chromatic aberration of the lenses is minimised when the column operates in the non-beam-crossover mode, where the first lens produces as near a parallel beam as possible.

References

1. R. Smith, G. Carter and M.J. Nobes, "The theory of surface erosion by ion bombardment" Proc. Roy. Soc. London A (1986). In press.

2. R.J. Amos, G.A. Evans and R. Smith, "High accuracy ion optics computing" Vacuum (1986). In press.

3. J.P. Ducommun, M. Cantagrel and M. Moulin (1975). J. Mater. Sci. 70, 42-62.

4. E. Harting and F.H. Read, "Electrostatic Lenses" 1976, Elsevier, Amsterdam.

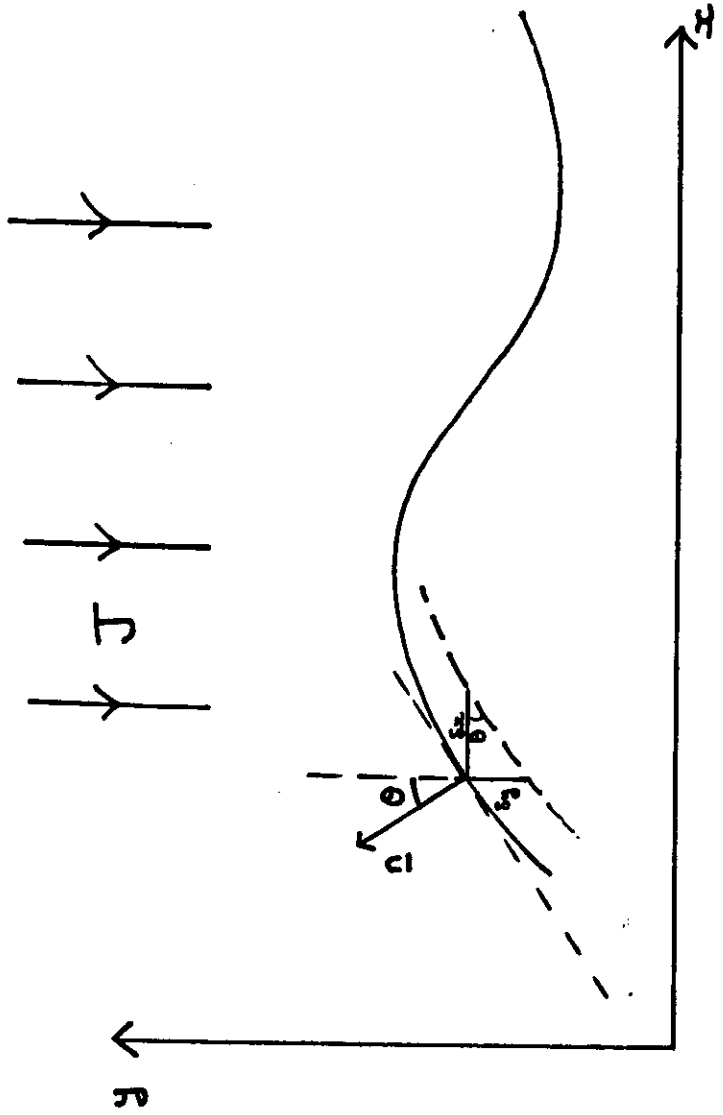


Figure 2.1. A schematic diagram illustrating the erosion of a two dimensional surface contour.

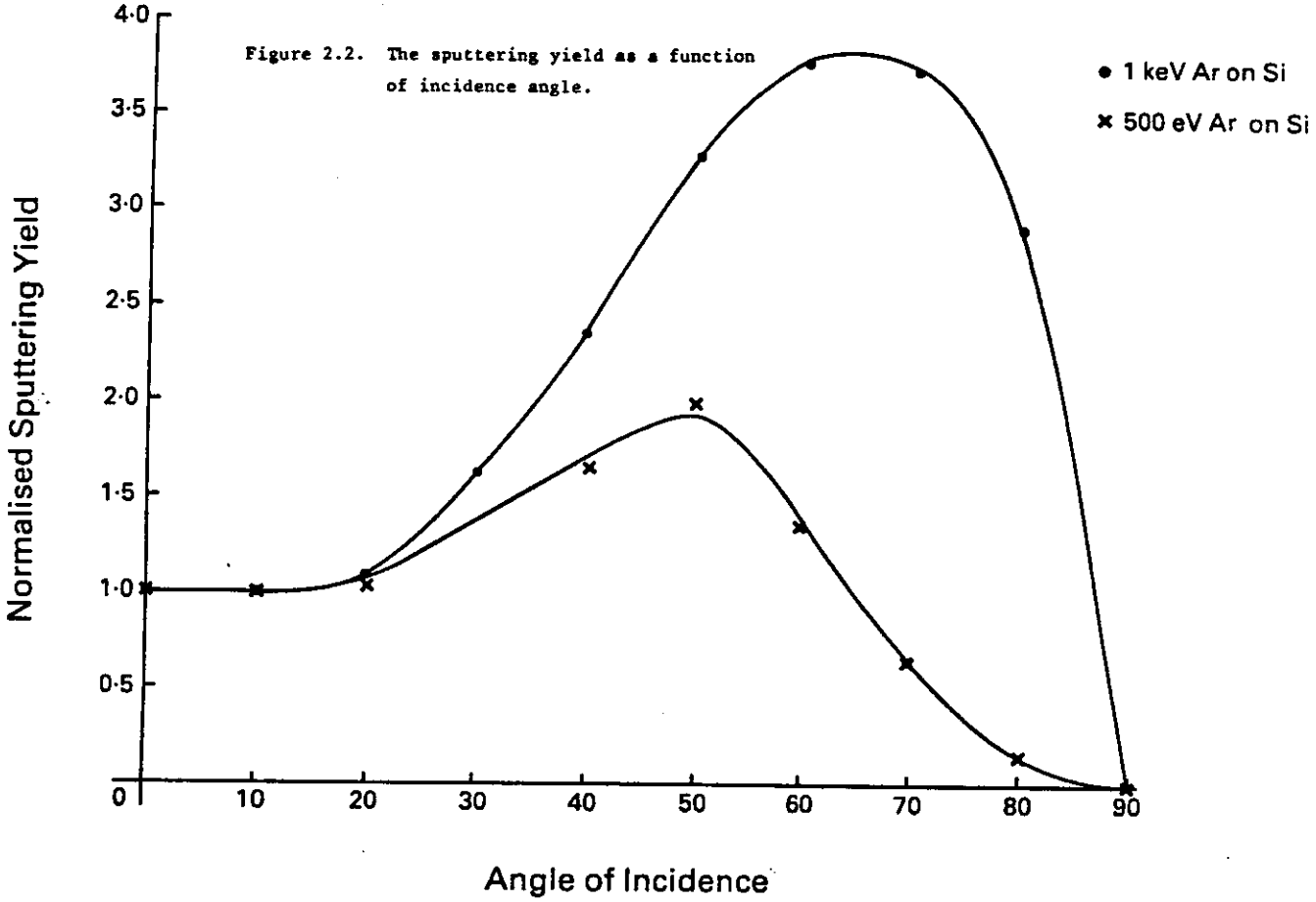
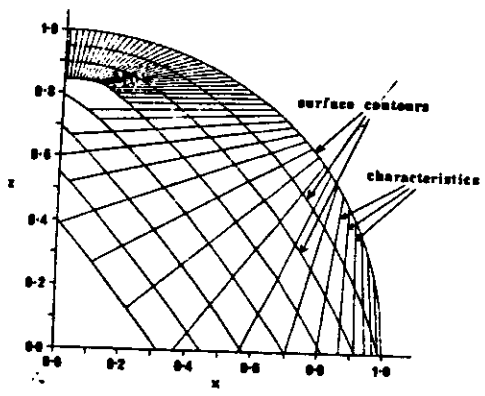
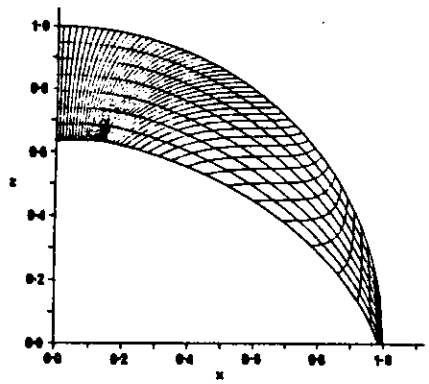


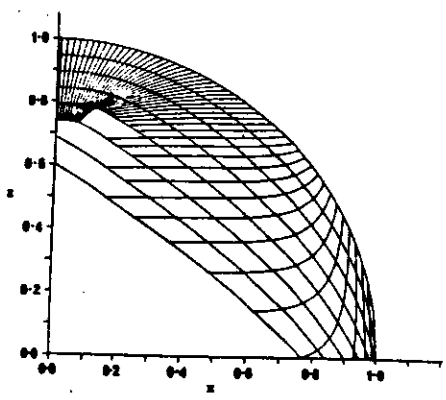
Figure 2.2. The sputtering yield as a function of incidence angle.



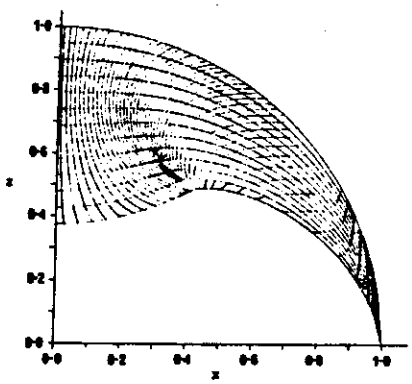
(a)



(c)



(b)



(d)

Figure 2.3. Successive erosion contours of the cylinder $x^2+z^2=1$, with characteristic lines. (a) Uniform beam $\rho=\infty$, (b) Gaussian beam $\rho=1$, (c) Gaussian beam $\rho=1/\sqrt{2}$ and (d) Gaussian beam $\rho=1/\sqrt{3}$.

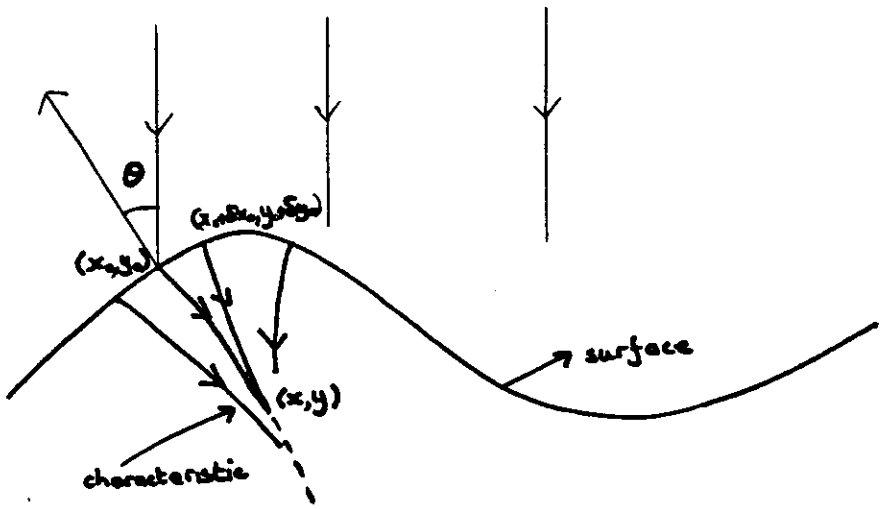


Figure 2.4. The formation of an edge from an initially smooth surface due to characteristic intersections.

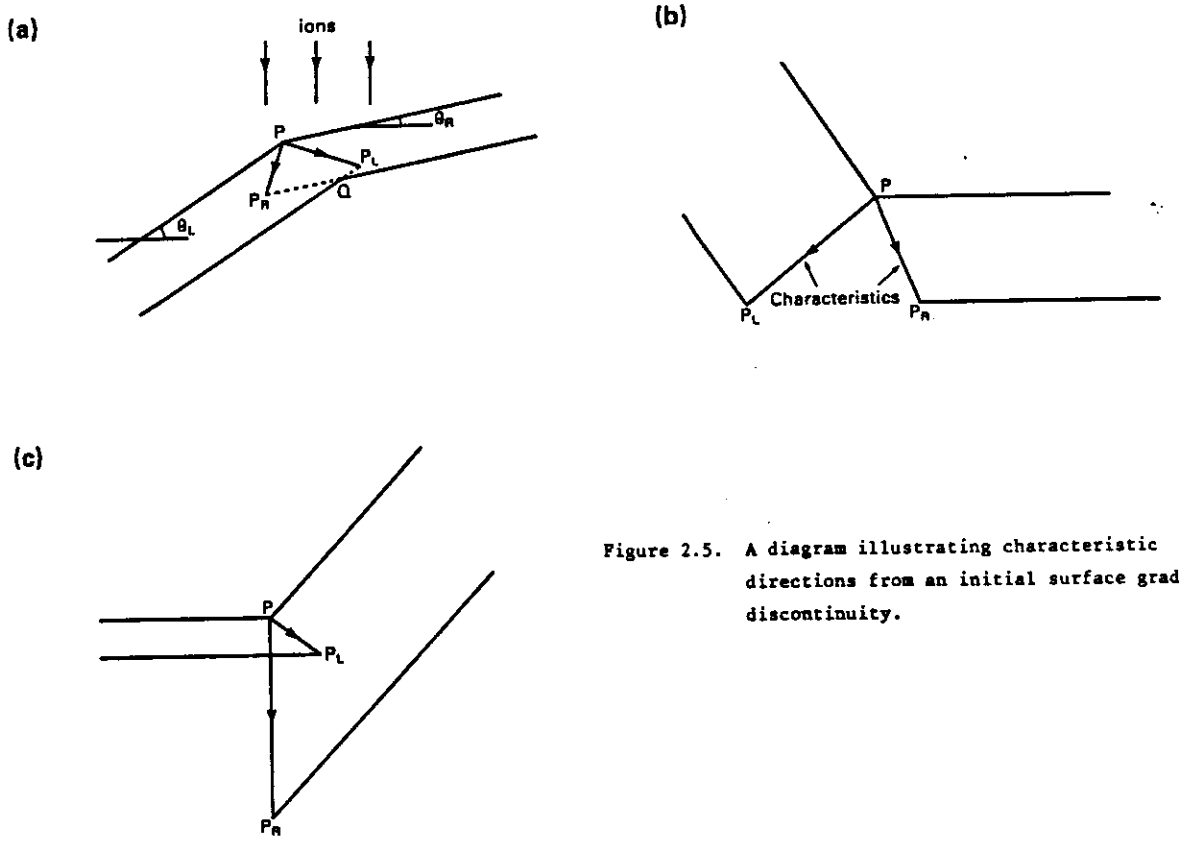


Figure 2.5. A diagram illustrating characteristic directions from an initial surface gradient discontinuity.

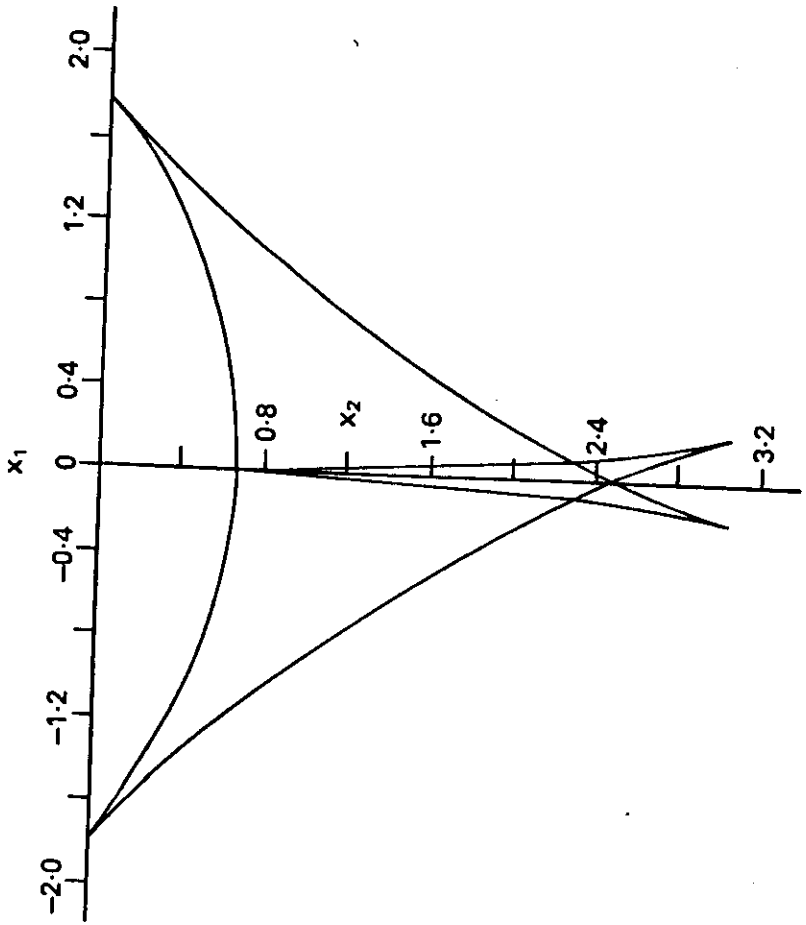


Figure 2.6. The characteristic locus curve for angles of all orientations from 0.

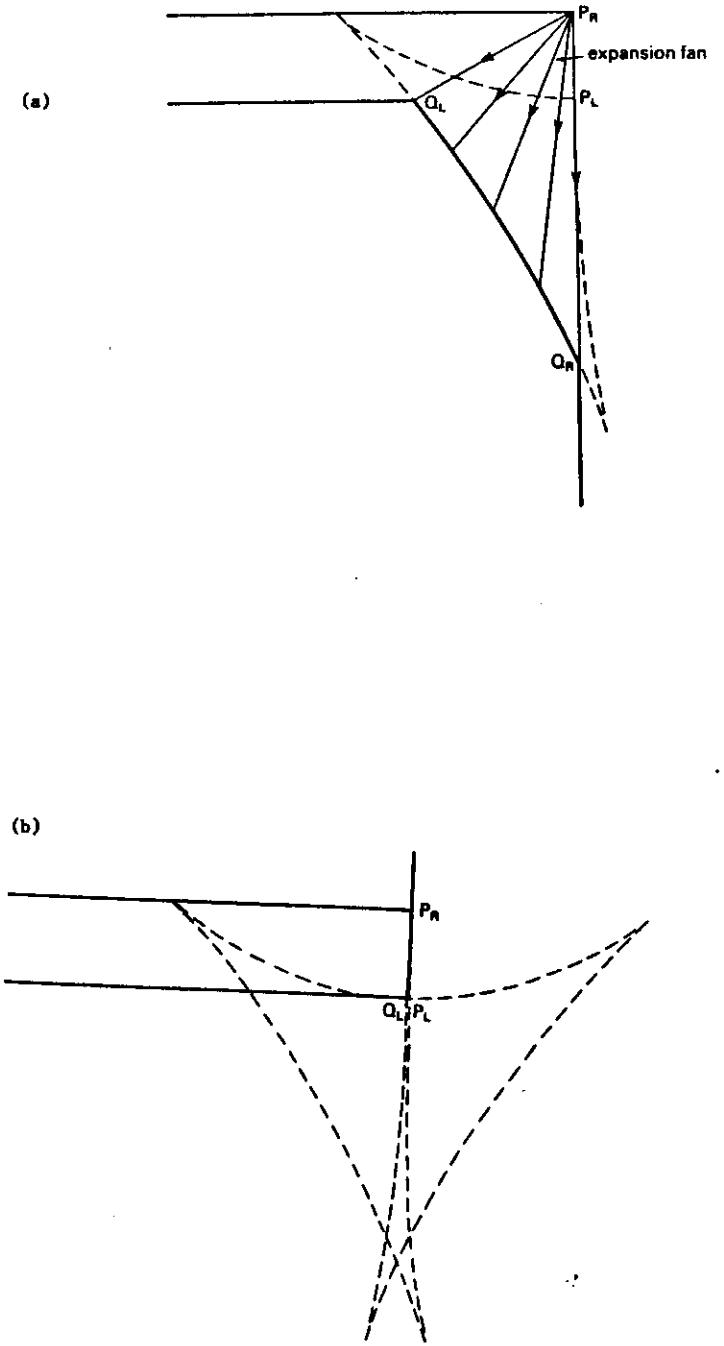


Figure 2.7. The expansion of 90° corners on an ion bombarded surface (a) Convex corner. (b) Concave corner

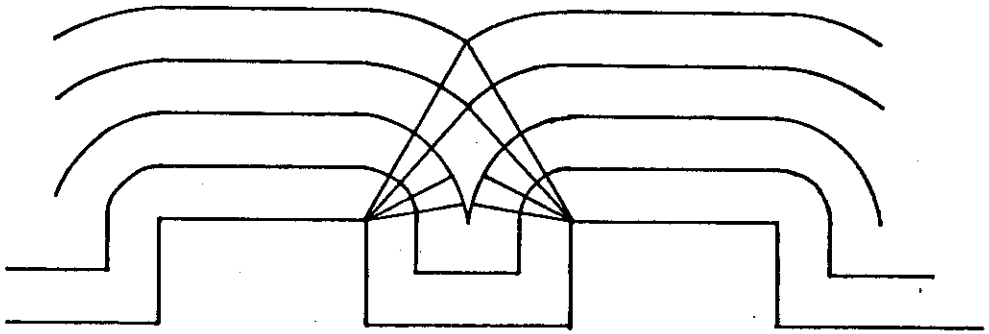


Figure 2.8. The development of an initially square wave profile due to isotropic deposition.

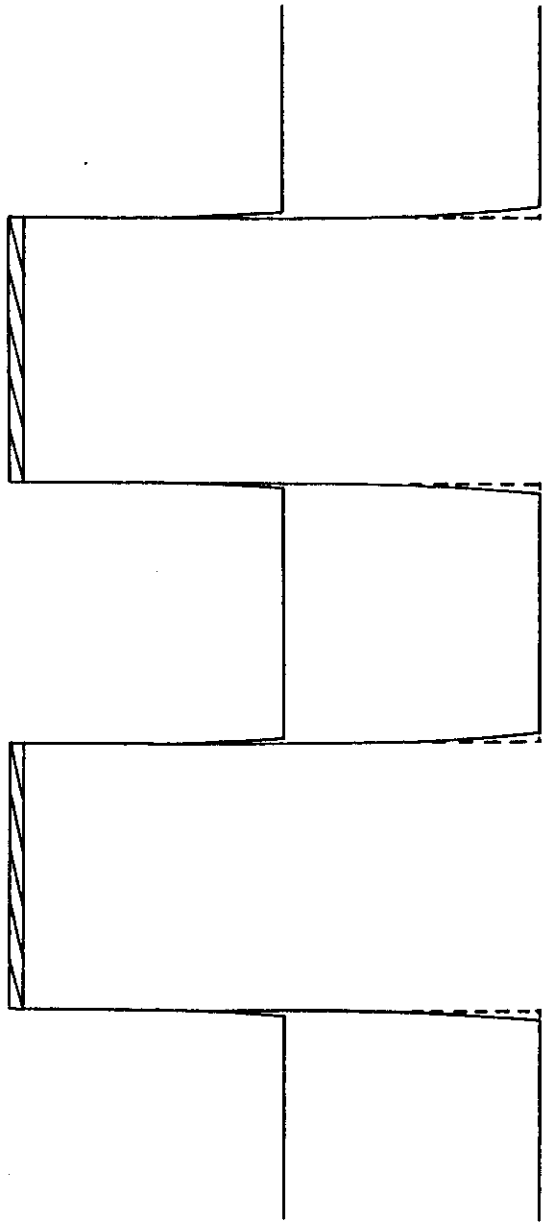


Figure 2.9a. The erosion of a square wave profile defined by a non erodible mask.

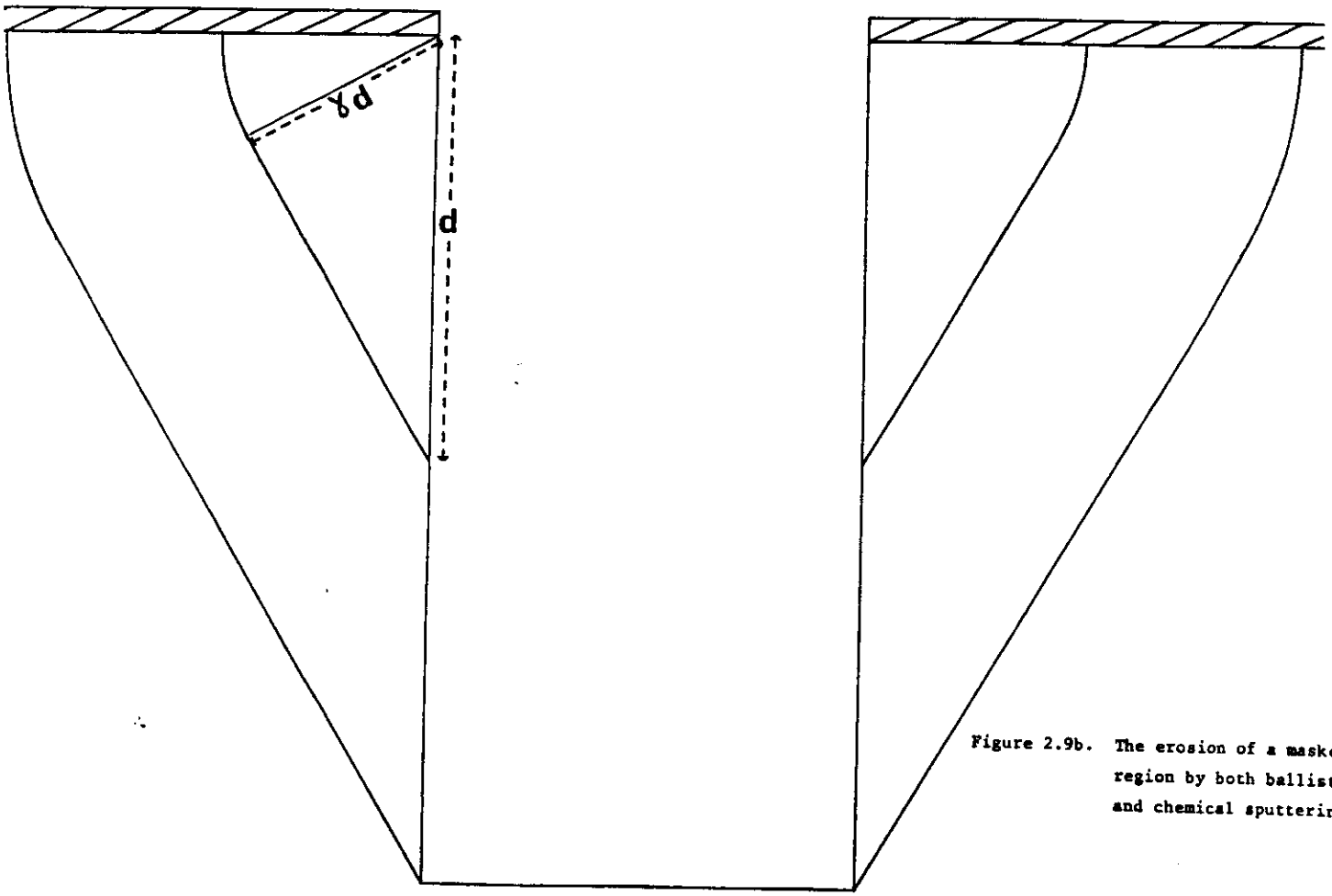


Figure 2.9b. The erosion of a masked region by both ballistic and chemical sputtering.

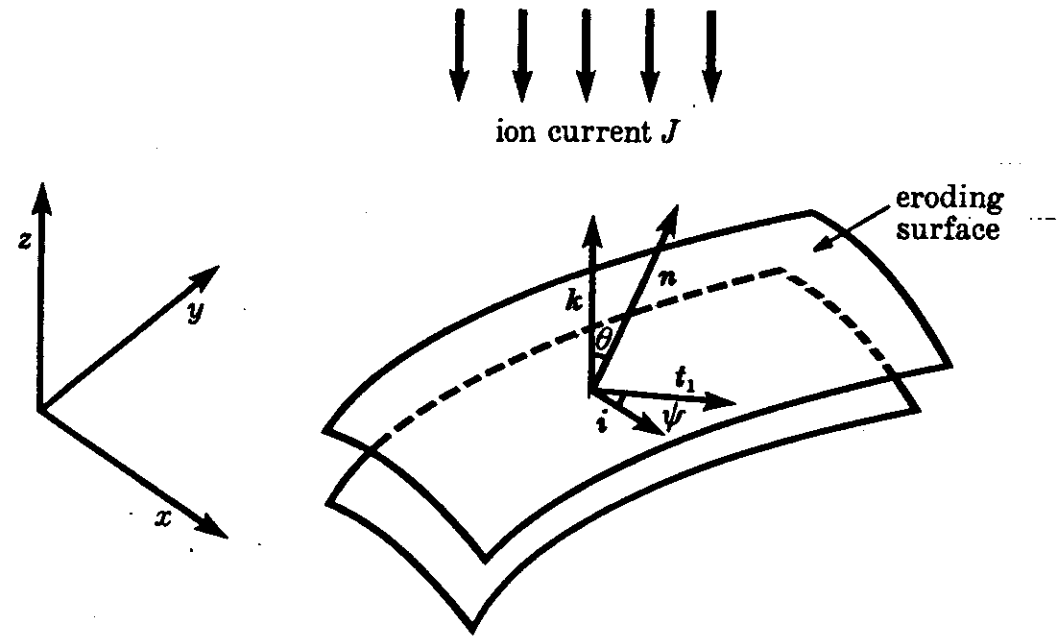


Figure 2.10. A schematic diagram illustrating surface erosion in three dimensions.

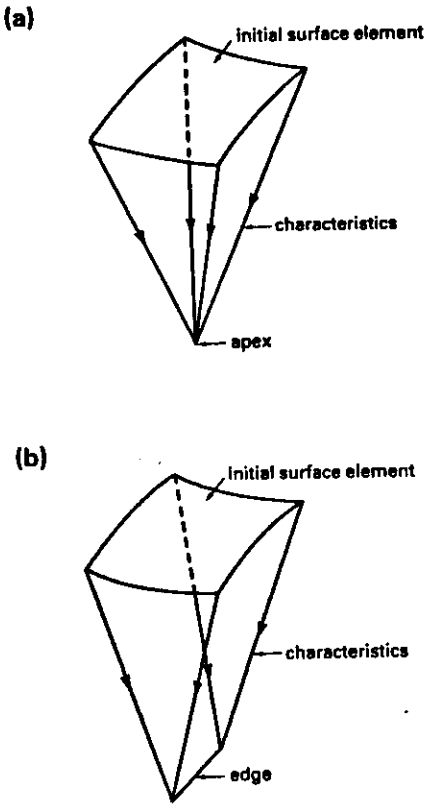


Figure 2.11. A diagram illustrating the formation of a) an apex, (b) an edge by characteristic instersections.

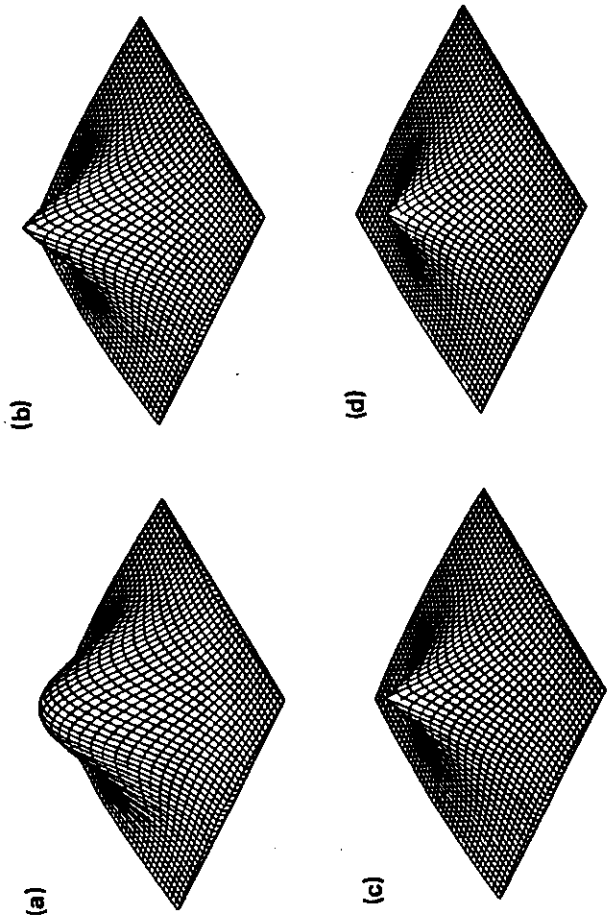


Figure 2.12(a)-(d). Various stages in the erosion of a symmetric hummock.

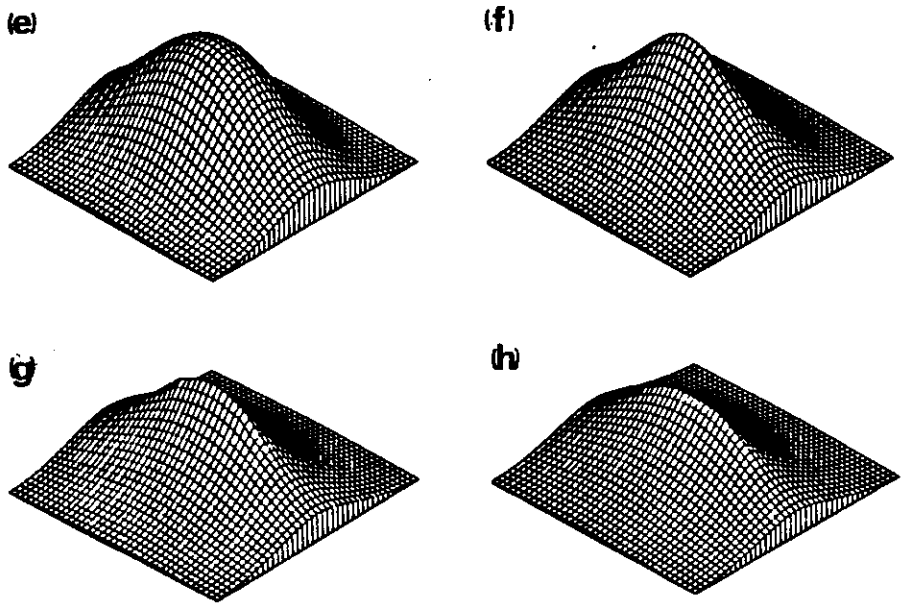


Figure 2.12(e)-(h). Various stages in the erosion of an elliptical hummock.

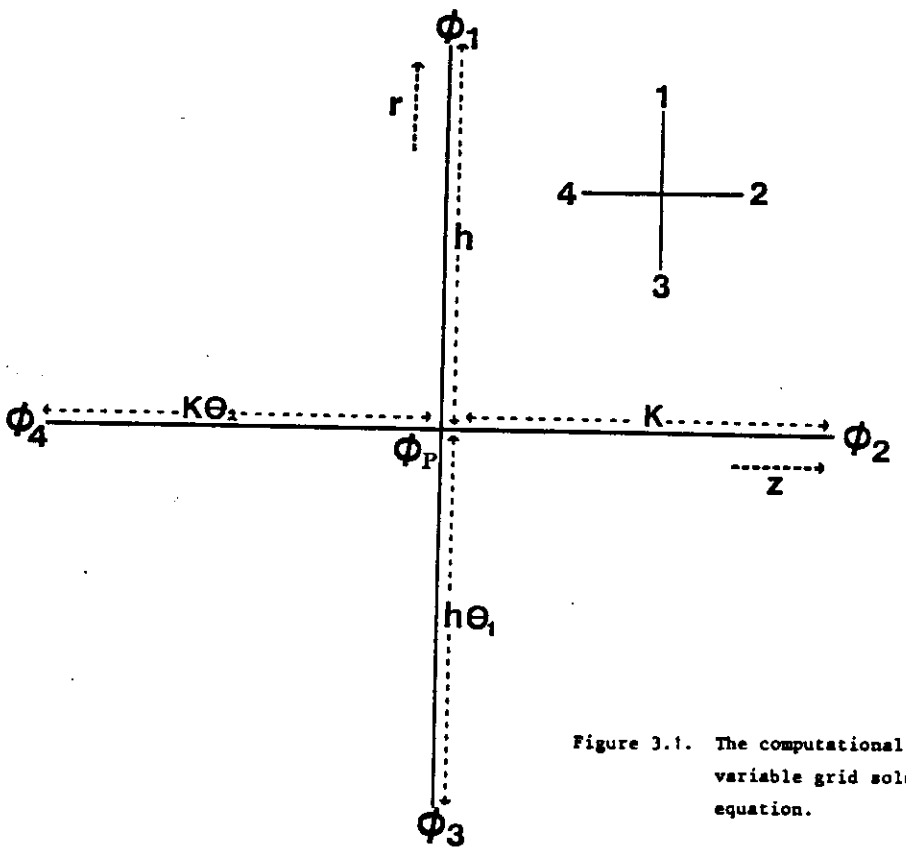


Figure 3.1. The computational molecule used for the variable grid solution of Laplace's equation.

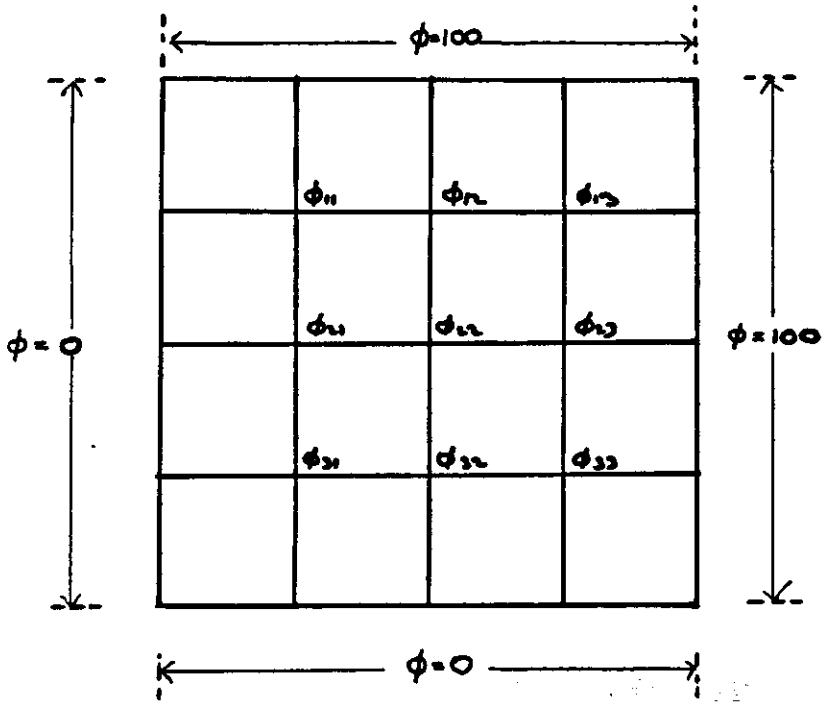


Figure 3.2. The regular grid used in the model problem.

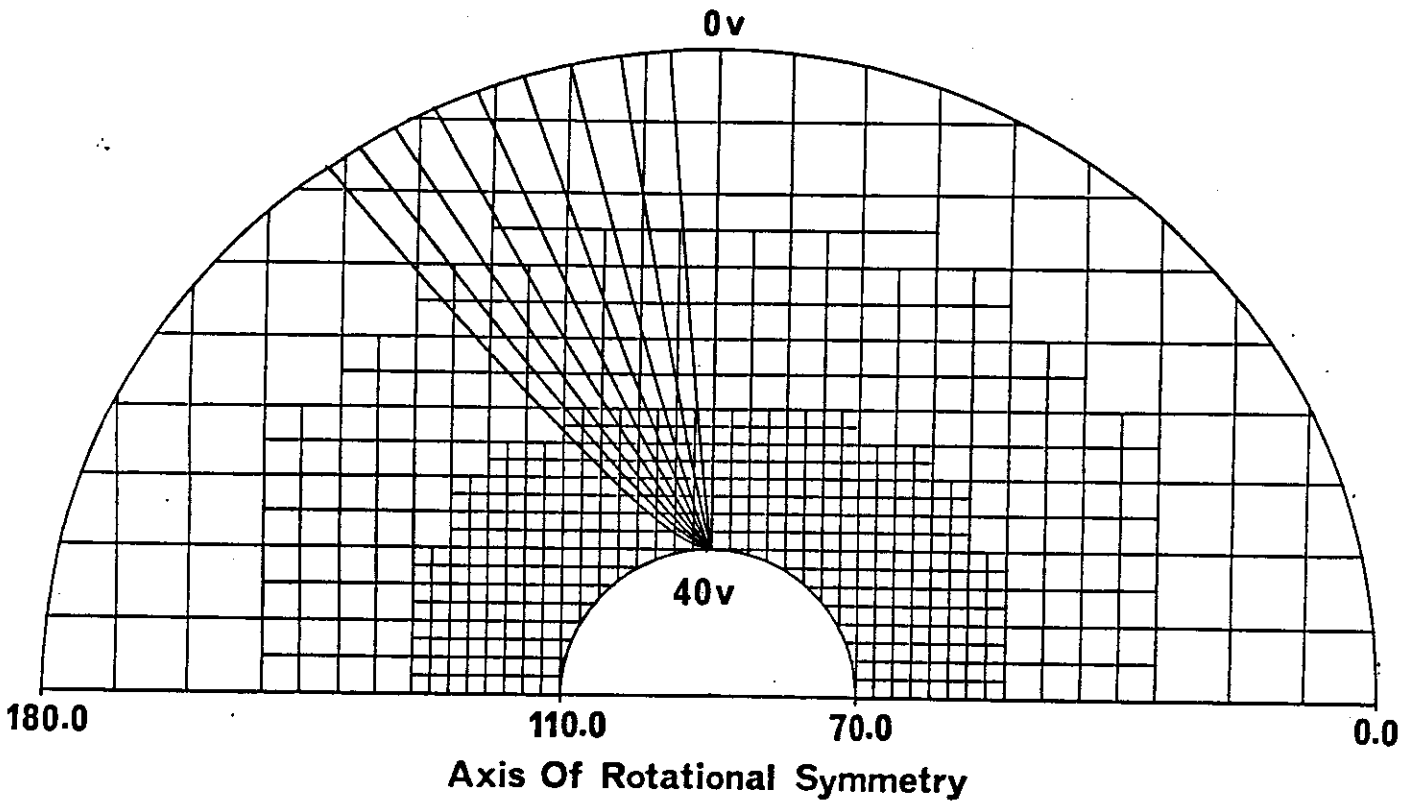


Figure 3.3. The variable grid used for the spherical electrode problem.

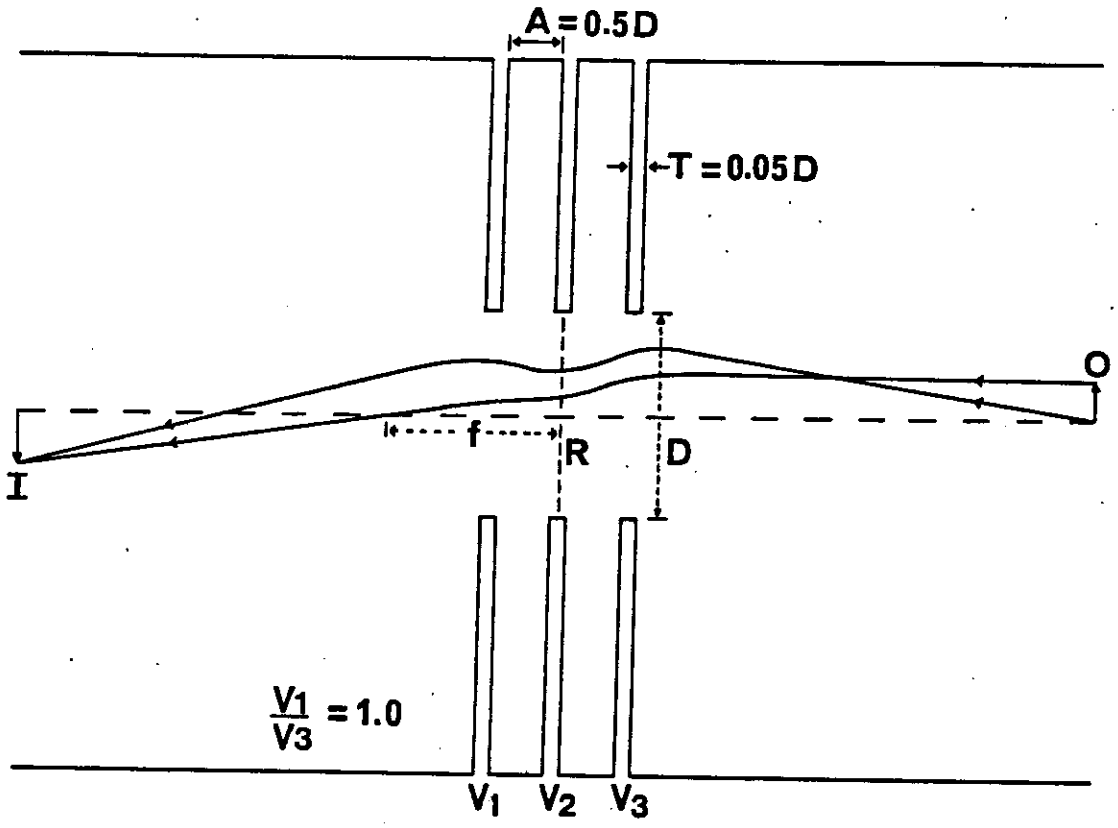


Figure 3.4a. A schematic diagram illustrating the important quantities in a three aperture electrostatic lens.

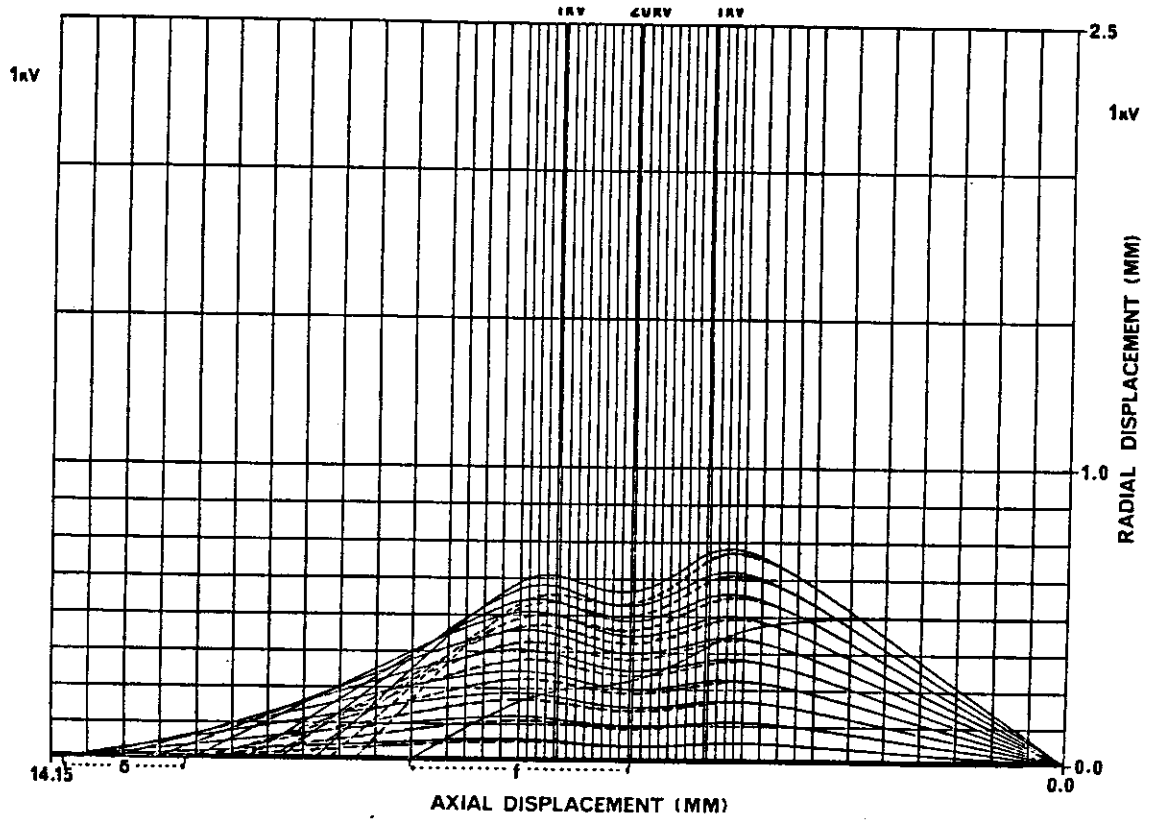


Figure 3.4b. Ion trajectory calculations for a three aperture lens.

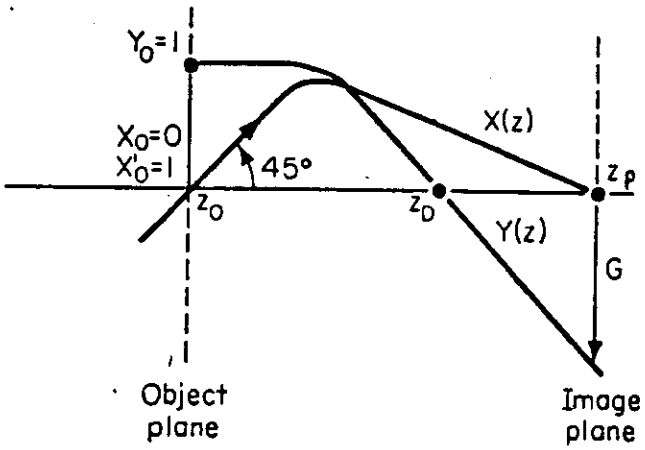


Figure 3.5. The principle trajectories use in the paraxial ray approximation.

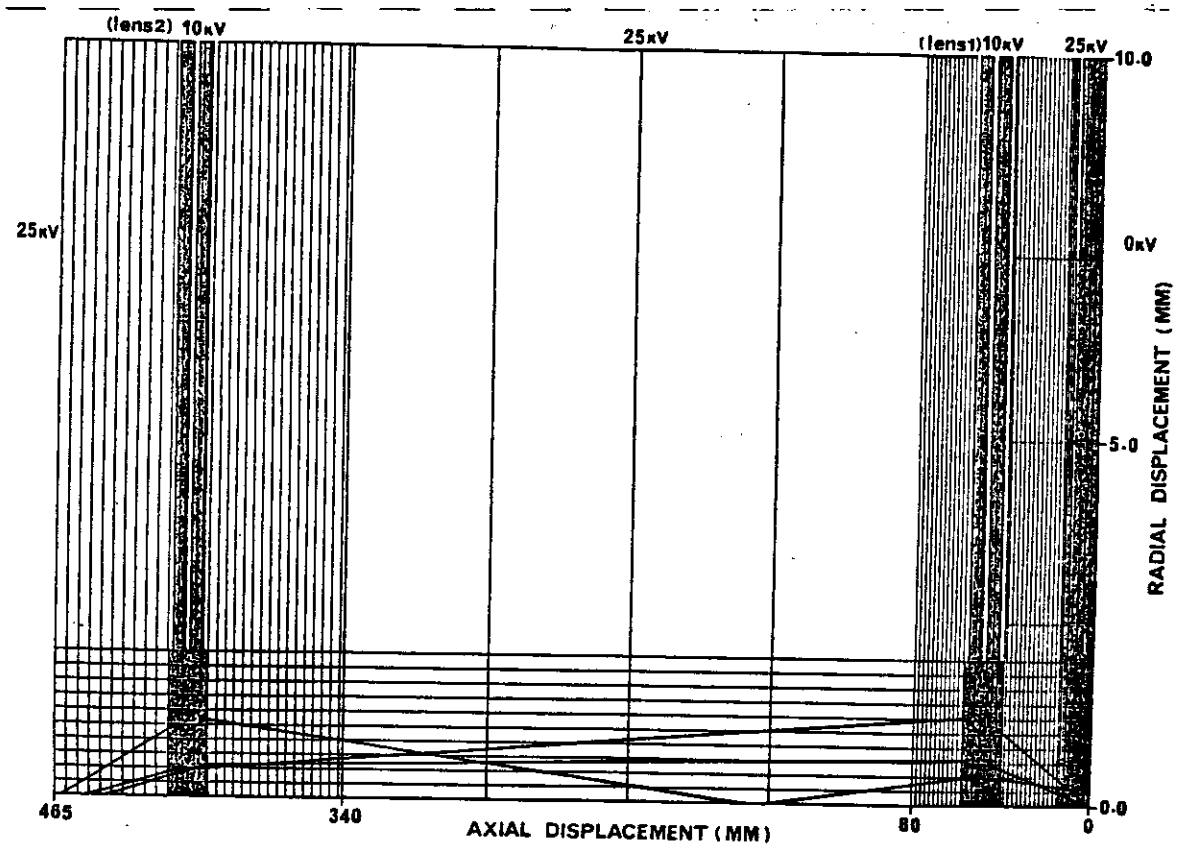


Figure 3.6. Ion trajectory calculations for a possible lithography column.

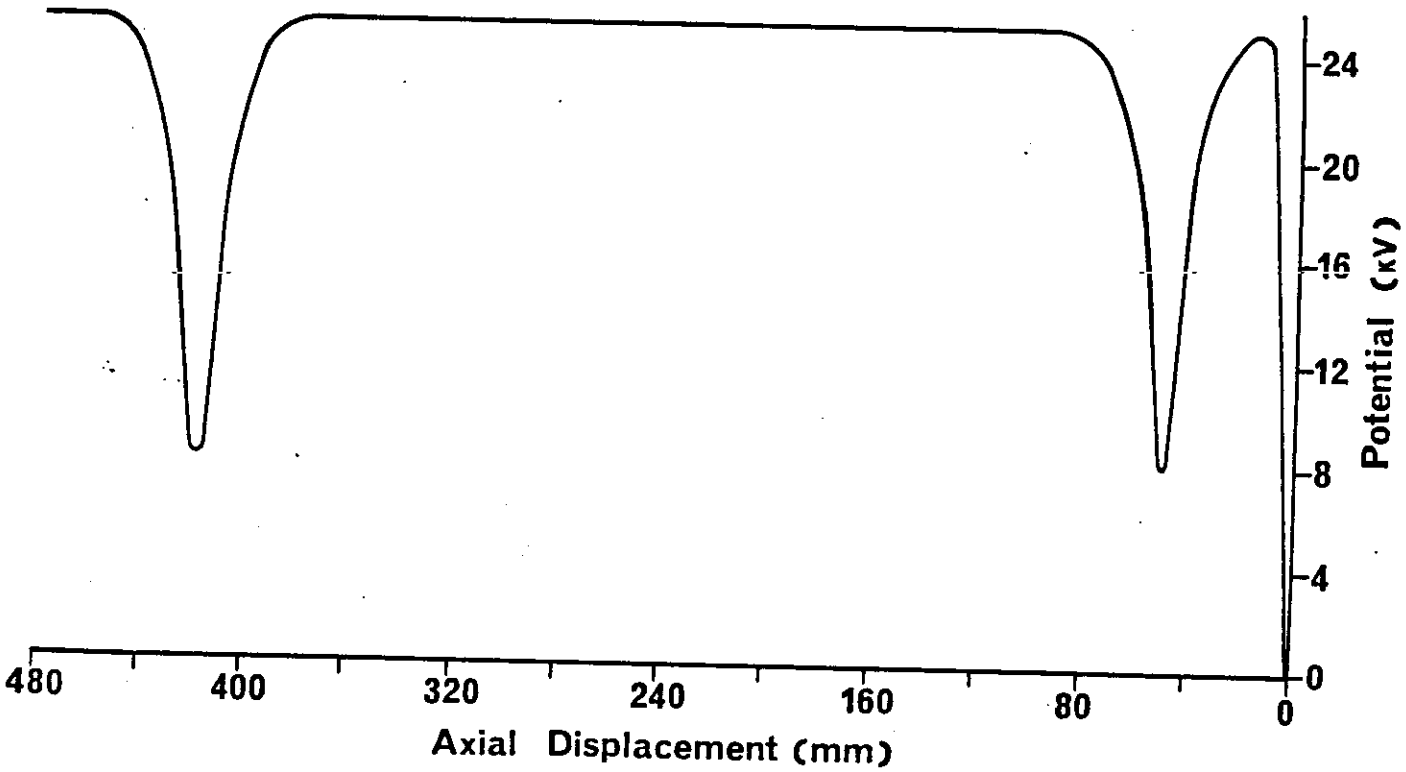


Figure 3.7. The axial potential variation for the column of figure 3.6.