6. A new model of the steady-state distribution of metastable particles in a discharge with cylindrical symmetry.

Previous chapters discuss a number of methods of determining transport parameters by time resolved measurements of the concentration of excited gas particles. Useful information can also be obtained from steady state measurements. For example, the flux of metastable atoms into the cathode can be determined by measuring $\partial n_m/\partial z$ in the vicinity of the cathode; comparison of the total flux with the total current passed by the discharge enables one to deduce the value (averaged over the whole cathode surface) of γ_m, the efficiency of ejection of secondary electrons by metastable atoms. Although this quantity has been measured previously for a number of combinations of excited state and substrate material (see table 2.2 for examples), there is a wide variation in the values obtained by different experimenters for the same combination of materials. In the case of metastable argon atoms incident upon 'chemically clean' copper, the two available values differ by an order of magnitude (Dunning and Smith 1971, Schohl *et al* 1992). In addition, nearly all the published data have been measured using beams of metastable particles in a high vacuum. The substrates under these conditions are often much cleaner than it is possible to achieve when a substantial quantity of gas is present. It was therefore thought desirable to attempt an accurate determination of the secondary ejection efficiency γ_m under the conditions prevailing in an electrical discharge.

In order to determine the total flux of excited states, measurements of $\partial n_m/\partial z$ must be made across the entire cathode radius, these samples then being integrated with the appropriate weights to achieve the desired result. This laborious measurement of the profile of the concentration of metastables might be avoided if the shape of the profile could be deduced *a priori*; a single measurement would then suffice to normalise the profile and thus allow the flux to be determined. It is the purpose of the present chapter to show how this shape can be estimated.

The analysis developed in this chapter can also be used to examine the effect that variations in the value of γ_m across the cathode have on the spatial growth of current. Ernest (1991) suggested that such a non-constant γ_m might explain the anomalies in the shapes of $\overline{\omega}(d)$ curves (the parameter $\overline{\omega}(d)$ is defined in chapter 2.2.2) which are observed under certain circumstances in nitrogen. In section 6.6, these observations are shown to be consistent with a step-function variation in $\gamma_m(r)$.

In the present chapter, the diffusion equations for electrons and metastable particles are solved within a discharge geometry that is similar to that used in chapter 5 (see figure 3.1, page 23). Infinite parallel plane electrodes are assumed, with the primary current restricted to a circular region on the cathode of radius R_w. The origin of coordinates is again placed at the centre of this 'window'. The single departure from the model of chapter 5 is the restriction of the secondary current to a larger circular area, concentric with the first and of radius R_E. This arrangement represents an attempt to model the finite radius of the actual electrode arrangement without imposing a barrier to the diffusive escape of metastable particles from the discharge volume. The assumption of infinite planar electrodes probably leads to a slight overestimate of this loss flux, due to the continued 'quenching' of metastable particles by the model electrode planes at $r > R_E$.

6.1. Principal equations and methods of solution.

The inhomogeneous diffusion equation for metastable particles, in the absence of significant coupling between states, is given by equation (2.15). In cylindrical polar coordinates the steady state form of this can be written

$$\frac{D_m}{r}\frac{\partial}{\partial r}\left(r\frac{\partial n_m}{\partial r}\right)-G_m n_m=-w\alpha_m(\mathbf{r})n_e(\mathbf{r}). \qquad (6.1)$$

where D_m, G_m and α_m are, respectively, the diffusion, volume quenching and excitation coefficients for the metastable particle, n_m and n_e are the concentrations, respectively, of metastable atoms and electrons and w is the electron drift velocity. As discussed in chapter 5, the electron concentration n_e can be approximated by solving equation (5.12), which is reproduced here for convenience of reference:

$$D\left(\frac{\partial^2 n_e}{\partial r^2}+\frac{1}{r}\frac{\partial n_e}{\partial r}\right)+D_L\frac{\partial^2 n_e}{\partial z^2}-w\frac{\partial n_e}{\partial z}+w\alpha_i n_e=0. \qquad (6.2)$$

The quantities D and D_L are the radial and longitudinal diffusion coefficients of the electrons under the prevailing gas pressure and α_i is the primary ionisation coefficient. Equations (6.1) and (6.2) are coupled through the cathode boundary condition (see equations (5.14) and (5.15))

$$\frac{j_z(r,0)}{e} = w n_e(r,0) - D_L \left.\frac{\partial n_e}{\partial z}\right|_{z=0} = q\left[\frac{j_{prim}(r)}{e} + \gamma_i(r)\phi_i(r,0) + \gamma_m(r)\phi_m(r,0)\right]. \qquad (6.3)$$

Here q is the fraction of electrons that escape back-diffusion, e is the electronic charge, j_{prim} is the primary current density and the quantities γ and $\phi(r,0)$ are, respectively, the secondary ejection efficiencies and cathodic flux densities of ionised or metastable atoms. As mentioned in the introduction, the secondary coefficients γ are taken to be zero outside the radius $r = R_E$.

Equation (6.2) is solved in chapter 5 by the use of the Hankel transform. Although this transform can also be used to simplify equation (6.1), the transform of equation (6.3) creates a problem, because $\mathscr{H}[\gamma(r)\phi(r,0)]$, the transform of the product $\gamma\phi$, cannot easily be expressed in the form $f(\gamma)\mathscr{H}[\phi]$, for some function f. This difficulty prevents the simple treatment of the equations within the Hankel transform space. The Hankel method has no advantages under these circumstances.

Although several alternative approaches are explored in the present chapter, the best of these in terms of flexibility, rate of convergence and computation time involves the expansion of the concentration of electrons $n_e(\mathbf{r})$ in a Fourier-Bessel series. In order for this to be possible, the infinite-parallel-plane model of chapter 5 must be modified so as to include a boundary at $r = R_E$. This boundary should be perfectly reflective to electrons, but transparent to the motion of metastable particles. This model simulates the real discharge conditions better than that of chapter 5. This is because, in a real discharge between finite parallel electrodes, the edge fields tend to prevent the diffusive escape of charged particles. However, the differences between this model and that of chapter 5 are minor for most of the discharge regimes employed in the present experiments (see, for example, figure 6.1).

As mentioned above, the Fourier-Bessel approach was found, on several counts, to be the most favourable method of solution of the diffusion equations. One disadvantage of the method is that complicated expressions arise in the course of the analysis. Because of this, it is desirable to insure to some extent against errors by comparing these results against those of other methods whenever possible. Two such checks of the Fourier-Bessel approach are performed in the present chapter. In the first case, the diffusive motions of the electrons are neglected. In this limit the electron number density n_e may be approximated by

$$n_e(r,z) = \frac{S(d)\,j_z(r,0)}{ew} \begin{cases} \exp(\alpha_i \Delta z), & z > d_0 \\ \\ 1, & z < d_0 \end{cases} \tag{6.4}$$

where $\Delta z = z - d_0$ and

$$j_z(r,0) = \begin{cases} j_{\text{prim}} + e\gamma_m \phi_m(r,0), & r < R_w \\ e\gamma_m \phi_m(r,0), & R_w < r < R_E \\ 0, & R_E < r. \end{cases} \tag{6.5}$$

The Townsend amplification factor $S(d)$ is given by

$$S(d) = \left\{1 - \gamma_m \left[\exp(\alpha_i \Delta d) - 1\right]\right\}^{-1}. \tag{6.6}$$

Note that $\alpha_i(z)$ has been approximated by the step function described in equation (2.4). This approximation is used throughout the present chapter. Because the diffusive motion of the metastable particles tends to smooth out any fluctuations in the generating function, the flux density of metastable particles across the cathode ought to be relatively insensitive to the exact form of $\alpha_i(z)$.) This form of the function $n_e(\mathbf{r})$ may be inserted directly into equation (6.1). Two ways of solving the resulting equation are explored in the next two sections, and compared in figure 6.2 with the Fourier-Bessel result obtained in the limit of negligible electron diffusion.

The second check on the Fourier-Bessel treatment is performed by neglecting the contribution of metastable particles. In this case, the solution of equation (6.2) can be compared directly with the analysis of chapter 5. An example of such a comparison is given in figure 6.1.

6.2. Finite-difference approximation.

The method of finite differences (see, for example, Smith 1974) can be a useful way of approximating the solutions $y(\mathbf{r})$ to equations of the type

$$\nabla^2 y - \mu^2 y = f(\mathbf{r}). \tag{6.7}$$

The derivatives are replaced by the equivalent finite-difference formulae, giving a system of linear equations in the samples y_{ij} of $y(\mathbf{r})$. These equations can be written, to first order in the finite-difference formulae, as

$$\sum_{p=i-1}^{i+1} \sum_{q=j-1}^{j+1} a_{pq} y_{pq} = f_{ij}. \tag{6.8}$$

In principle this system can be inverted directly, although iterative methods are often preferable.

The finite-difference method was applied to the solution of equation (6.1), where the electron concentration was given by equations (6.4) and (6.5). Matters are complicated a little by the lack of a radial boundary to the distribution of metastables. This difficulty was overcome by dividing the model discharge volume into two regions, separated by a boundary at $r = R_E$. Equation (6.1) was applied in the inner region. The following differential equation in the inverted coordinate $\varphi = 1/r$ can be derived from equation (6.1):

$$\frac{\partial^2 n_m}{\partial z^2} + \varphi^4 \frac{\partial^2 n_m}{\partial \varphi^2} + \varphi^3 \frac{\partial n_m}{\partial \varphi} - \mu^2 n_m = 0. \tag{6.9}$$

This equation was used in the outer region. Note that the source term in equation (6.1) is equal to zero outside the electrode radius.

The resulting finite-difference equations were solved iteratively. The best rate of convergence was obtained using an extrapolated Liebmann method (Smith 1974). Even with the use of this technique, the convergence was very slow unless the acceleration parameter was chosen with care. As mentioned in the introduction to the present chapter, an example of the approximation obtained for $\phi_m(r,0)$ by using this method is displayed in figure 6.2.

Of the three solution methods considered in the present chapter, the finite-difference method required the least amount of analysis but the most computation.

6.3. Conversion of the metastable-particle diffusion equation to integral form.

Numerical approximation of a differential equation may often be performed more satisfactorily if the equation is first converted to an integral form. This is achieved in the present section by use of a Green's function method.

An equation in the Green's function g can be obtained by replacing the source function of equation (6.1) by a unit impulse source at $r = r'$ and $z = z'$, giving

$$\left(\nabla^2 - \mu^2\right) g = -\delta(r - r')\delta(z - z')/2\pi r', \tag{6.10}$$

where $\mu = G_m/D_m$ as before. The solution to this equation is

$$g(r,r',z,z') = \begin{cases} \dfrac{1}{\pi d}\displaystyle\sum_{n=1}^{\infty} I_0(\rho_n r)\sin(\omega_n z)K_0(\rho_n r')\sin(\omega_n z'), & r < r' \\ \dfrac{1}{\pi d}\displaystyle\sum_{n=1}^{\infty} I_0(\rho_n r')\sin(\omega_n z')K_0(\rho_n r)\sin(\omega_n z), & r > r' \end{cases} \tag{6.11}$$

where $\omega_n = n\pi/d$ and $\rho_n^2 = \mu^2 + \omega_n^2$. The functions $I_j(x)$ and $K_j(x)$ are the modified Bessel functions of order j. By Green's theorem, the solution of equation (6.1) is given by

$$n_m(r,z) = \frac{2\pi w}{D_m}\int_0^{\infty} dr'\, r' \int_0^{d} dz'\, g(r,r',z,z')\,\alpha_m(z')n_e(\mathbf{r}'). \tag{6.12}$$

However, the metastable and electron concentrations are still coupled. In order to progress further, one of these quantities must be eliminated, leaving a single equation in the other. This can be achieved by obtaining from equation (6.12) an equation in $\phi_m(r,0)$, the flux density of metastable atoms into the cathode. Recall that $\phi_m(r,z)$ is given by

$$\phi_m(r,z) = D_m \frac{\partial n_m}{\partial z}; \tag{6.13}$$

from equation (6.12), the cathodic flux density of metastable atoms is therefore

$$\phi_m(r,0)=\frac{2w}{d}\sum_{n=1}^{\infty}\omega_n\int_0^d dz'\sin(\omega_n z')\alpha_m(z')$$
$$\times\left[K_0(\rho_n r)\int_0^r dr'\,r'I_0(\rho_n r')\,n_e(\mathbf{r}')+I_0(\rho_n r)\int_r^{\infty} dr'\,r'K_0(\rho_n r')\,n_e(\mathbf{r}')\right]. \qquad (6.14)$$

This can now be combined with equation (6.3) to give a single equation in either $\phi_m(r,0)$ or $n_e(\mathbf{r})$. These two approaches are explored in the following two sections.

6.4. An integral equation in the metastable-particle flux density $\phi_m(r,0)$.

Substitution of equation (6.4) into equation (6.14) results in the following integral equation in $\phi_m(r,0)$:

$$\phi_m(r,0)=q(r)+L\int_0^{R_E} dr'\,\kappa(r,r')\,\phi_m(r',0) \qquad (6.15)$$

where

$$q(r)=\frac{2\alpha_m S(d) j_{prim}}{d}\times\begin{cases}\displaystyle\sum_{n=1}^{\infty}\omega_n H_n\left[1-\rho_n R_w I_0(\rho_n r)K_1(\rho_n R_w)\right], & r<R_w\\ \displaystyle\sum_{n=1}^{\infty}\omega_n H_n\rho_n R_w I_1(\rho_n R_w)K_0(\rho_n r), & r>R_w,\end{cases} \qquad (6.16)$$

$$L=\frac{2\alpha_m S(d)\gamma_m}{d}, \qquad (6.17)$$

$$\kappa(r,r')=\begin{cases}\displaystyle\sum_{n=1}^{\infty}\omega_n H_n r' I_0(\rho_n r)K_0(\rho_n r'), & r<r'\\ \displaystyle\sum_{n=1}^{\infty}\omega_n H_n r' I_0(\rho_n r')K_0(\rho_n r), & r'<r\end{cases} \qquad (6.18)$$

and

$$H_n = \frac{\omega_n \cos(\omega_n d_0) - \alpha_i \sin(\omega_n d_0) - \omega_n (-1)^n \exp(\alpha_i \Delta d)}{\omega_n^2 + \alpha_i^2}. \tag{6.19}$$

A discrete approximation to equation (6.15) can be found by replacing the integral by an appropriate quadrature formula. The inversion of the resulting equations is, however, most efficiently accomplished if the kernel $\kappa(r,r')$ is first symmetrised. This can be done by replacing ϕ_m in equation (6.15) by the modified flux function

$$f(r) = \phi_m(r,0)\sqrt{r}. \tag{6.20}$$

Equation (6.15) then becomes

$$f(r) = q(r)\sqrt{r} + L\int_0^{R_E} dr' f(r')\kappa'(r,r'). \tag{6.21}$$

The new, symmetric kernel κ' is related to the old kernel κ by

$$\kappa'(r,r') = \kappa(r,r')\sqrt{r/r'}. \tag{6.22}$$

The quantities q and κ' in equation (6.21) can be estimated by truncating the sums in equations (6.16) and (6.18) at convenient points. The sums in equation (6.18) converge for all $r \neq r'$, but the convergence becomes slower as $r \to r'$ and fails completely at $r = r'$. Similar considerations apply to the sums in the definition of $q(r)$ as $r \to R_w$. (The convergence occurs because the product $I_i(x)K_j(y) \sim \exp(x-y)$ at large values of the arguments x and y (Abramowitz and Stegun 1972).) However, the kernel κ is smoothly varying; $\kappa(r,r)$ can therefore be approximated by the average

$$\overline{\kappa}(r,r) = \frac{\kappa(r-\delta,r-\delta,) + \kappa(r+\delta,r+\delta,)}{2} \tag{6.23}$$

for some small δ. Note also that $\kappa(0,0) = \kappa(r,0) = 0$.

A Simpson's rule quadrature of the integral in equation (6.21) gives the following linear system:

$$f_k = q_k\sqrt{kh} + hL\sum_{j=0}^{N} w_j \kappa'_{kj} f_j \qquad (6.24)$$

where f_k is the kth sample of $f(r)$, $h = R_E/3N$ and the w_j are the Simpson's rule weights $\{1,4,2,4,\ldots\}$. (As described above, the diagonal entries κ'_{jj} were approximated by the average of $\kappa'_{j-1,j-1}$ and $\kappa'_{j+1,j+1}$.) The symmetry in the matrix of coefficients can be retained by defining a new quantity $m_k = f_k\sqrt{w_k}$; in these terms, equations (6.24) become

$$m_k = q_k\sqrt{khw_k} + hL\sum_{j=0}^{N} \kappa'_{kj} m_j \sqrt{w_j w_k}\,. \qquad (6.25)$$

Equations (6.25) were inverted using a Cholesky factorisation (Fisher 1988). A typical solution to equation (6.25), using the same discharge parameters as before, is displayed in figure 6.2.

The metastable density distribution throughout the discharge volume can now be found by substituting equations (6.4) and (6.5) into equation (6.12) and numerically integrating.

6.5. Fourier-Bessel expansion.

At large values of the reduced electric field E/N (N here is the gas number density), the distribution of the electrons in a pre-breakdown discharge begins to be affected by their own diffusive motion. This is particularly true of argon: low-energy electrons diffuse very readily in this gas, due to the Ramsauer minimum in its elastic collision cross-section (Ramsauer and Kolluth 1929). The effects of electron diffusion have not been considered in the analyses of sections 6.2 and 6.4 of the present chapter; inclusion of these effects would double the complexity of the former analysis and is impossible in the latter. It would be useful, therefore, to develop a more general analysis in which the diffusive motion of the electrons is included from the start. The Fourier-Bessel method described in the present section answers this requirement. This approach also avoids the both the convergence problems of the integral equation treatment and the discretisation errors implicit in both this and the finite-difference method.

Lucas (1965) solved the electron diffusion equation using a similar series expansion, except that Lucas was modelling an electrode arrangement that included guard rings at $r = R_E$, which are not a feature of the present author's apparatus. Whereas guard rings may be expected to absorb any electrons that come into contact with them, the open boundary of the present experimental arrangement has, instead, the effect of reflecting or confining the electrons, because of the curvature of the electric fields at the edges of the electrodes. Thus, while Lucas used the condition $n_e(R_E) = 0$, the condition

$$\left.\frac{\partial n_e}{\partial r}\right|_{r=R_E} = 0 \tag{6.26}$$

is more appropriate for the present analysis. The anode and cathode boundary conditions which were used in chapter 5 are retained.

Equation (6.2) can be solved by the technique of separation of variables, giving the general solution

$$n_e(r,z) = \sum_{k=0}^{\infty} A_k R_k(r) Z_k(z) \tag{6.27}$$

where R_k and Z_k are, respectively, the solutions of the equations

$$rR_k'' + R_k' + \theta_k^2 r R_k = 0 \tag{6.28}$$

and

$$Z_k'' - 2\lambda Z_k' + \left[\lambda^2 - u_k^2(z)\right] Z_k = 0. \tag{6.29}$$

Here θ_k is the variable of separation,

$$\lambda = w/2D_L \tag{6.30}$$

and

$$u_k^2(z) = \lambda^2 - 2\lambda\alpha_i(z) + \theta_k^2 D/D_L. \tag{6.31}$$

If the step function approximation given in equation (2.4) is used for $\alpha_i(z)$, the solution to equation (6.29) is

$$Z_k(z) = \begin{cases} B_k \exp[(\lambda - u_k')z] + C_k \exp[(\lambda + u_k')z], & z < d_0 \\ \exp[(\lambda - u_k)z] + F_k \exp[(\lambda + u_k)z], & z > d_0 \end{cases} \tag{6.32}$$

where

$$u_k^2 = \lambda^2 - 2\lambda\alpha_i + \theta_k^2 D/D_L \tag{6.33}$$

and

$$u_k'^2 = \lambda^2 + \theta_k^2 D/D_L . \tag{6.34}$$

The anode boundary condition

$$n_e(r,d) = 0 \tag{6.35}$$

requires that F_k = -exp(-2$u_k d$); the other two constants can be evaluated by comparing Z_k and Z_k' at d_0, giving

$$2u_k' B_k = (u_k' + u_k)\exp\left[(u_k' - u_k)d_0\right] - (u_k' - u_k)\exp\left[(u_k' + u_k)d_0 - 2u_k d\right] \tag{6.36}$$

and

$$2u_k' C_k = (u_k' - u_k)\exp\left[-(u_k' + u_k)d_0\right] - (u_k' + u_k)\exp\left[(u_k - u_k')d_0 - 2u_k d\right]. \tag{6.37}$$

The radial equation (6.28) has the solution

$$R_k(r) = J_0(\theta_k r) \tag{6.38}$$

where $J_n(x)$ is the Bessel function of order n. The quantity $\theta_k R_E$ is the kth zero of $J_1(x)$, with $\theta_0 = 0$, in order to satisfy the radial boundary condition (equation (6.26)). Use of equation (5.14) shows that the z-component of the current density $j_z(r,z)$ is given by

$$j_z(r,z) = eD_L \sum_{k=0}^{\infty} A_k J_0(\theta_k r) L_k(z) \tag{6.39}$$

where

$$L_k(z) = \begin{cases} B_k(\lambda + u_k')\exp\left[(\lambda - u_k')z\right] + C_k(\lambda - u_k')\exp\left[(\lambda + u_k')z\right], & z < d_0 \\ (\lambda + u_k)\exp\left[(\lambda - u_k)z\right] - (\lambda - u_k)\exp\left[(\lambda + u_k)z - 2u_k d\right], & z > d_0. \end{cases} \tag{6.40}$$

The expansion of $j_z(r,0)$ in Bessel functions allows the coefficients A_k to be evaluated:

$$A_k = \frac{2}{R_E^2 J_0^2(\theta_k R_E) e D_L \beta_k} \int_0^{R_E} dr\, r\, j_z(r,0) J_0(\theta_k r) \tag{6.41}$$

where

$$\beta_k = B_k(\lambda + u_k') + C_k(\lambda - u_k'). \tag{6.42}$$

An example of the anode current distribution derived from this solution is compared in figure 6.1 with a similar solution calculated using the Hankel transform method of chapter 5. (The transport parameters and the form of cathode current used were the same.) The Fourier-Bessel series was truncated at $k = 4$ as the higher terms were found to make a negligible contribution.

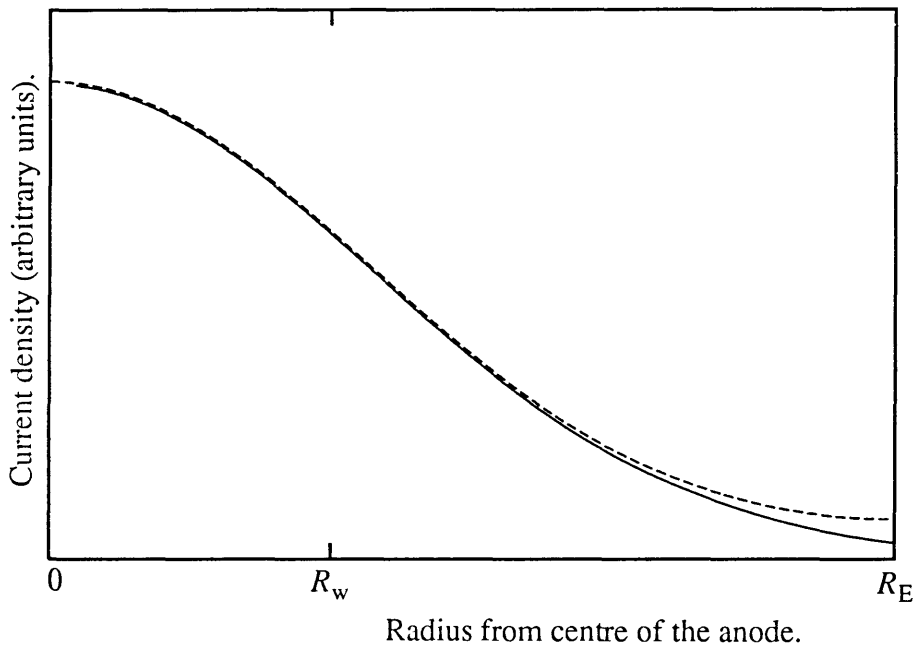

Figure 6.1. Calculated current densities at the anode in argon at 20 Td and 5 Torr, in the absence of secondary processes, with the same electrode dimensions and separation as in figure 5.1. The solid curve is the solution calculated using the Hankel transform method of chapter 5, whereas the dashed line represents the Fourier-Bessel solution of the present chapter. Eight terms of the Fourier-Bessel series were kept. The discharge conditions were chosen so that the two treatments are just distinguishable.

Before attempting to include the contribution of metastables, the ion flux should be considered. This is easy to deduce, since the diffusive motion of the ions can be neglected (see chapter 5). In this approximation the ionic flux becomes

$$\phi_i(r,0) = w\int_0^d dz\, n_e(r,z)\alpha_i(z)$$

$$= w\alpha_i \sum_{j=0}^{\infty} \frac{A_j I_j}{\lambda^2 - u_j^2} J_0(\theta_j r) \tag{6.43}$$

where

$$I_j = 2u_j \exp\left[(\lambda - u_j)d\right] - (\lambda + u_j)\exp\left[(\lambda - u_j)d_0\right] + (\lambda - u_j)\frac{\exp\left[(\lambda + u_j)d_0\right]}{\exp(2u_j d)}. \tag{6.44}$$

The concentration of metastable atoms can be eliminated in the following way. Equation (6.27) is first substituted into equation (6.14). If the functional form of the excitation coefficient $\alpha_m(z)$ is approximated by equation (2.20), the following expression is obtained for the cathodic flux density of the metastables:

$$\phi_m(r,0) = \frac{2}{d}\sum_{n=1}^{\infty}\omega_n \sum_{j=0}^{\infty} A_j H_{nj} M_{nj}(r) \tag{6.45}$$

where

$$H_{nj} = w\alpha_m \int_{d_0}^{d} dz \sin(\omega_n z) Z_j(z), \tag{6.46}$$

the function $Z_j(z)$ being defined in equation (6.12). The integral in equation (6.46) is not difficult to evaluate, but the result is a little complicated:

$$\frac{H_{nj}}{w\alpha_m} = \frac{\left[\omega_n \cos(\omega_n d_0) - (\lambda - u_j)\sin(\omega_n d_0)\right]\exp\left[(\lambda - u_j)d_0\right] - (-1)^n \omega_n \exp\left[(\lambda - u_j)d\right]}{(\lambda - u_j)^2 + \omega_n^2} -$$

$$- \frac{\left[\omega_n \cos(\omega_n d_0) - (\lambda + u_j)\sin(\omega_n d_0)\right]\exp\left[(\lambda + u_j)d_0\right] - (-1)^n \omega_n \exp\left[(\lambda + u_j)d\right]}{\left[(\lambda + u_j)^2 + \omega_n^2\right]\exp(2u_j d)} \tag{6.47}$$

The quantity $M_{nj}(r)$ evaluates to be

$$M_{nj}(r)=\frac{1}{\left(\rho_n^2+\theta_j^2\right)}\begin{cases}\left[J_0\left(\theta_j r\right)-\rho_n R_{\mathrm{E}} I_0(\rho_n r) J_0\left(\theta_j R_{\mathrm{E}}\right) K_1(\rho_n R_{\mathrm{E}})\right], & r<R_{\mathrm{E}}\\ \rho_n R_{\mathrm{E}} K_0(\rho_n r) J_0\left(\theta_j R_{\mathrm{E}}\right) I_1(\rho_n R_{\mathrm{E}}), & r>R_{\mathrm{E}}.\end{cases} \tag{6.48}$$

The final expression is obtained by first substituting equations (6.43) and (6.45) into equation (6.3); the result is then substituted into equation (6.41), giving

$$A_k=\Gamma_k+\Lambda\sum_{j=0}^{\infty}A_j\kappa_{jk} \tag{6.49}$$

where

$$\Gamma_k=\frac{qS_kR_{\mathrm{w}}j_{\mathrm{prim}}}{R_{\mathrm{E}}^2J_0^2(\theta_kR_{\mathrm{E}})eD_{\mathrm{L}}\beta_k}\times\begin{cases}2J_1(\theta_kR_{\mathrm{w}})/\theta_k, & k>0\\ R_{\mathrm{w}}, & k=0,\end{cases} \tag{6.50}$$

$$\Lambda=\frac{4qS_k}{dR_{\mathrm{E}}^2J_0^2(\theta_kR_{\mathrm{E}})D_{\mathrm{L}}\beta_k}, \tag{6.51}$$

$$\kappa_{jk}=\sum_{n=1}^{\infty}\omega_n H_{nj}U_{njk}, \tag{6.52}$$

and

$$U_{njk}=\int_0^{R_{\mathrm{E}}}dr\,r\gamma_{\mathrm{m}}(r)J_0(\theta_kr)M_{nj}(r). \tag{6.53}$$

The factor S_k is given by

$$S_k=\left\{1-\frac{2\lambda q\alpha_{\mathrm{i}}\gamma_{\mathrm{i}}}{\beta_k w\left(2\lambda\alpha_{\mathrm{i}}-\theta_k^2D/D_{\mathrm{L}}\right)}I_k\right\}^{-1}. \tag{6.54}$$

If $\gamma_{\mathrm{m}}(r)=\gamma_{\mathrm{m}}$, U_{njk} evaluates to be

$$U_{njk}=\frac{\gamma_{\mathrm{m}}\rho_n^2R_{\mathrm{E}}^2J_0\left(\theta_jR_{\mathrm{E}}\right)J_0(\theta_kR_{\mathrm{E}})}{\left(\rho_n^2+\theta_k^2\right)\left(\rho_n^2+\theta_j^2\right)}\times\begin{cases}-K_1(\rho_nR_{\mathrm{E}})I_1(\rho_nR_{\mathrm{E}}), & j\neq k\\ \frac{1}{2}\left(1+\theta_k^2/\rho_n^2\right)-K_1(\rho_nR_{\mathrm{E}})I_1(\rho_nR_{\mathrm{E}}), & j=k.\end{cases} \tag{6.55}$$

The A_k spectrum can be approximated by truncating the sums in equations (6.49) and (6.52) at suitable values and inverting the resulting linear system. Both sums are unilaterally convergent. From the definitions of ω_n and ρ_n after equation (6.11) it can be seen that the decrease in H_{nj} is proportional to $1/n$ for large n; since U_{njk} decreases roughly as $1/n^2$, the entire summand in equation (6.52) decreases as $1/n^2$ as well. The convergence of the sum over j is even more rapid, since H_{nj} decreases with j as $\exp\left(-j\pi d_0\sqrt{D/D_L}\right)$ for large j.*

An example solution (calculated in the limit of negligible electron diffusion) is compared in figure 6.2 with the cathodic flux densities of metastables obtained by the previous two methods.

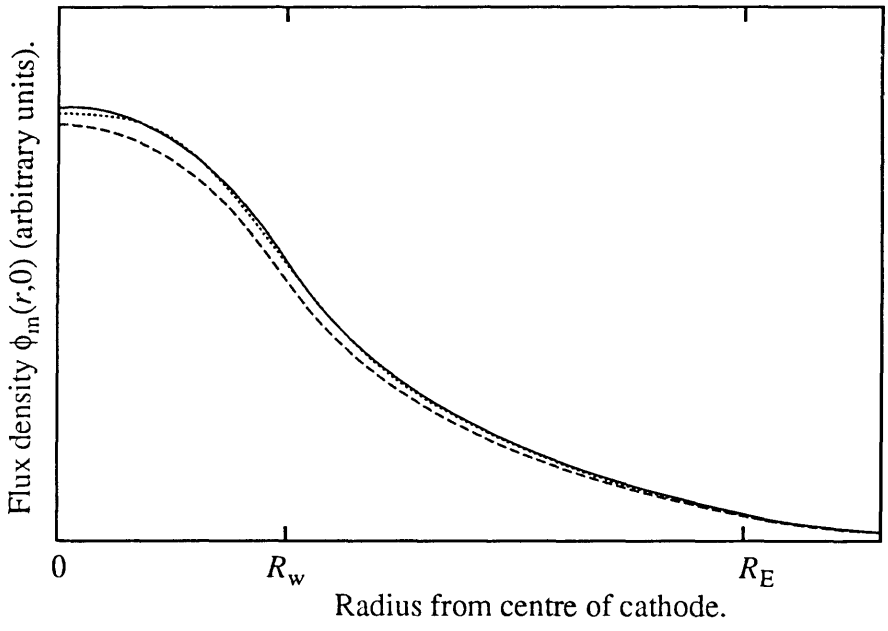

Figure 6.2. A comparison between three different methods of solving the metastable-particle diffusion equation. The solid line is the Fourier-Bessel solution (section 6.4), the dotted line is the integral equation solution (section 6.3) and the dashed line is the finite-difference solution (section 6.1). The Fourier-Bessel series was truncated after the sixth term to allow the Fourier-Bessel and integral equation solutions to be distinguished: at higher orders of truncation, the two curves are almost exactly superimposed. Little attempt was made to choose realistic discharge parameters because of the need to ignore the diffusion of electrons. The parameter μR_E was chosen to be equal to 3 and the breakdown parameter $\chi = I_{slow}/I_{total}$ set at 0.8.

* Note that H_{nj} does not vary with j in the limit of negligible electron diffusion. The series still converges in this case, but at a rate proportional to $j^{-5/2}$. This expression is derived from the asymptotic form of the Bessel functions and zeroes in equation (6.55).

As mentioned in the introduction, the primary purpose of this analysis is to facilitate the determination of the total flux of metastable atoms into the cathode. $\Phi_m(z)$, the total flux of metastables through a plane at z, is related to the flux density $\phi_m(r,z)$ by

$$\Phi_m(z) = 2\pi \int_0^\infty dr\, r\phi_m(r,z). \tag{6.56}$$

However, the amount ΔI of light absorbed by a thin beam directed through the axis of symmetry of the discharge, parallel to the x-y plane, is related to the concentration $n_m(r,z)$ by (see ch. 5, section 5.5)

$$\Delta I(z) = 2I\sigma \int_0^\infty dr\, n_m(r,z). \tag{6.57}$$

If z is small compared with the electrode separation d, the flux density may be approximated by $\phi_m(r,0) \sim n_m(r,z)/z$. (This assumes that $n_m(r,0) \sim 0$. See section 3.1.1 for a discussion of the validity of this assumption.) Hence the net flux into the cathode is given by

$$\Phi_m(0) \sim \frac{\pi \Delta I(z)}{I\sigma z \int_0^\infty dr\, \phi_m(r,0)}. \tag{6.58}$$

From the form of $M_{nj}(r)$ as defined in equation (6.48), it can be seen that the denominator of equation (6.58) involves integrals of the form

$$\int_0^a dr\, J_0(r), \int_0^a dr\, I_0(r) \text{ and } \int_a^\infty dr\, K_0(r).$$

The solutions to these integrals can be written as closed expressions involving Struve functions (Abramowitz and Stegun 1972), but for computational purposes polynomial approximations offer more than adequate precision. The integral of J_0 was approximated by a polynomial given by Hitchcock (1957); polynomial approximations to the other two integrals were found in Luke (1962a). In no case is the error incurred by the use of these formulae greater than 2.4×10^{-6}.

Finally, it is of interest to calculate the limiting shape of the flux density $\phi_m(r,0)$ in the approach to electrical breakdown. Because $\phi_m(r,0)$ is unbounded for all r in the breakdown limit, it is necessary to normalise the function by dividing by some quantity proportional to the total flux $\Phi_m(0)$. From equations (6.56) and (6.45), the total flux is given by

$$\Phi_m(0) = 2\pi \int_0^\infty dr\, r\, \phi_m(r,0)$$

$$= \frac{2\pi R_E^2 A_0}{d} \sum_{n=1}^{\infty} \frac{\omega_n}{\rho_n^2} H_{n0}. \tag{6.59}$$

Because this quantity also grows without bound towards breakdown, it follows from the definitions of the various terms in this equation that A_0 must contain the singularity at this point. Let the normalised coefficients A'_k be defined such that

$$A'_k = A_k / A_0 . \tag{6.60}$$

From equation (6.45), the normalised flux density $\phi_m(r,0)/A_0$ is given by

$$\frac{\phi_m(r,0)}{A_0} = \frac{2}{d} \sum_{n=1}^{\infty} \omega_n \sum_{j=0}^{\infty} A'_j H_{nj} M_{nj}(r), \tag{6.61}$$

whereas the coefficients themselves obey the equation

$$A'_k = \frac{\Gamma_k}{A_0} + \Lambda \sum_{j=0}^{\infty} A'_j \kappa_{jk} . \tag{6.62}$$

As breakdown is approached, the first term becomes negligible. The resulting eigenvalue equation can be solved by inverse iteration (Peters and Wilkinson 1979). The breakdown limit curve in figure 6.3 on the following page was calculated in this way. (The other curves on this figure are explained in section 6.6.)

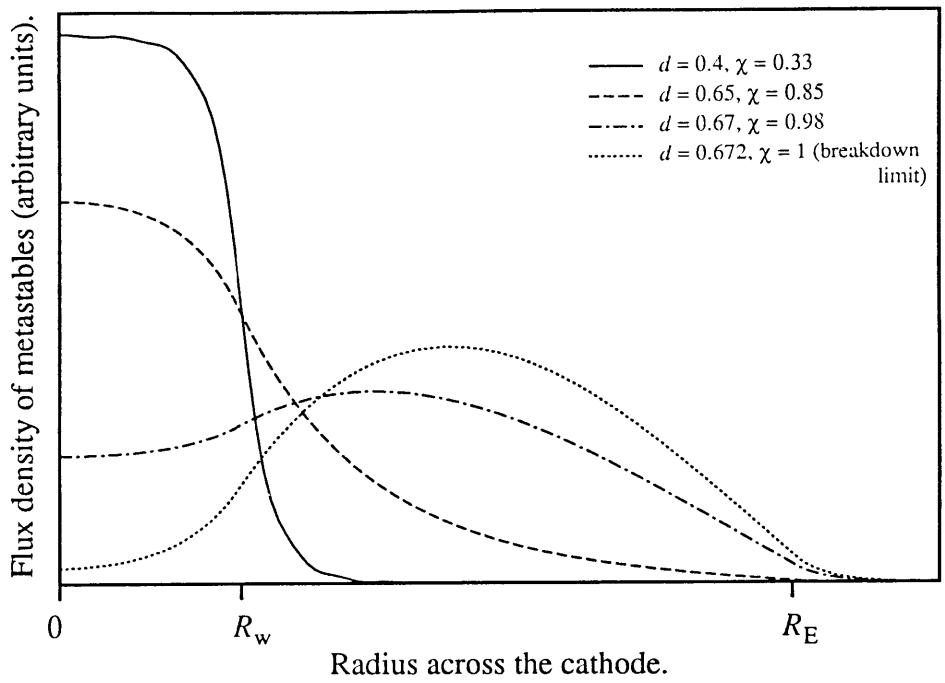

Figure 6.3. The flux density of metastable nitrogen molecules as a function of radius across the cathode, plotted for four different electrode spacings. The discharge parameters are the same as those used to calculate the theoretical curve of figure 6.4. Thirty terms of the Fourier-Bessel series were retained. The heights of the curves have been arbitrarily adjusted so that the total area under each curve is the same. In the legend, d is the electrode separation in cm and χ is the ratio between the 'slow' and total currents.

6.6. The effect on the growth of current of a radial variation in γ_m.

As mentioned in chapter 2, the generalised secondary coefficient $\overline{\omega}(d)$ provides a useful measure of the contribution of metastable-produced secondary electrons to the pre-breakdown discharge. An expression for this quantity within a three-dimensional discharge may be derived from the one-dimensional formula in equation (2.31a):

$$\overline{\omega}(d)=\frac{\alpha_i}{\Phi_i(0)}\left[\int_0^{R_E} dr\, r\gamma_i(r)\phi_i(r,0)+\int_0^{R_E} dr\, r\gamma_m(r)\phi_m(r,0)\right]. \qquad (6.63)$$

This may be expressed as

$$\overline{\omega}(d)=\frac{\alpha_i}{\Phi_i(0)}\left[\gamma_i'\Phi_i(0)+\gamma_m'\Phi_m(0)\right], \qquad (6.64)$$

which has the same form as equation (2.31a), except the secondary coefficients γ are replaced by effective values γ'. The effective secondary coefficients are given by

$$\gamma' = \frac{1}{\Phi(0)} \int_0^{R_E} dr\, r\gamma(r)\phi(r,0). \tag{6.65}$$

Graphs of $\overline{\omega}(d)$ against d at a constant E/N typically exhibit a shape that rises to a maximum before falling away monotonically. (Examples can be found in Haydon and Williams 1973b, 1976). This behaviour can be understood qualitatively as follows. At very small separations, the non-equilibrium region occupies most or all of the discharge volume. Under these conditions, metastable particles are created at a significant rate only near the anode, if at all. The probability is therefore high that these particles will diffuse to the anode, where they are destroyed; only a small proportion diffuse to the cathode. As the electrode separation is increased, the 'centre of gravity' of metastable production moves toward the mid-electrode region, increasing the proportion of metastables that reach the cathode; hence the initial rise in the $\overline{\omega}(d)$ curve. However, as the electrodes are further separated, two mechanisms that oppose this rise begin to have an effect. Firstly, there is the shape of the 'generating' function $\varepsilon(\mathbf{r}) = \alpha_m w n_e(\mathbf{r})$. Because the concentration of electrons increases exponentially across the discharge gap, the shape of ε becomes more and more 'peaked' near the anode as d is increased, which shifts the centre of production of metastables back toward the anode. Secondly, there is the volume quenching coefficient. Metastable particles are destroyed not just at the electrodes, but also within the discharge volume. At larger electrode separations, this process also reduces the ratio between the fluxes of metastables and ions. Both these processes cause the $\overline{\omega}(d)$ curve to fall away from its maximum value at higher values of d.

The above notwithstanding, Ernest (1991) found that, under some conditions, curves of $\overline{\omega}(d)$ in nitrogen discharges reversed direction again as the breakdown separation was approached. This author used the same electrode materials as those in the present study: one of bulk oxygen-free high-conductivity (OFHC) copper; the other constructed of the same material, but including an inlaid disk of fused silica in the centre, the whole electrode being coated with a film of gold of about 10 to 20 nm thickness (see chapter 7 for a more detailed description). Ernest found that the reversal in $\overline{\omega}(d)$ was particularly pronounced when the gold electrode was used as the cathode. (Some evidence of similar upcurving can also be seen in the $\overline{\omega}(d)$ figures of Haydon and Williams (1976).

These authors, although working in the same laboratory as Ernest, used a different discharge chamber.)

Ernest conjectured that the upcurving in $\overline{\omega}(d)$ was due to an increase in the value of the secondary efficiency γ_m toward the edge of the cathode. Some support for a variation in γ_m across the gold-film electrode at least can be found by considering the details of the construction of this electrode. The layer of gold might be expected to produce a uniform value of γ_m across the entire electrode; however, it is normal practice to 'bake' the entire vacuum chamber at several hundred degrees Celsius before performing experiments. This heat treatment may allow the gold which covers the metal part of the electrode to diffuse into the bulk metal. (Gold and copper are miscible in all proportions (Dugdale 1977).) The evidence for this is only qualitative, being based on observation of the colour of the gold-film electrode before and after heat treatment. If this process does in fact take place, it could easily lead to differences in the secondary efficiencies between the pure gold surface over the silica disk and the copper/gold alloy of the remainder of the electrode.

Poor adhesion between the gold film and the silica window is another possible source of a spatial variation in γ_m. It has been known for a long time that the noble metals adhere poorly to oxides and halides (eg Benjamin and Weaver 1961). Problems of this nature have, in fact, been encountered by the present author (see chapter 7). Electrode assemblies used at the University of New England are chemically cleaned prior to coating, but many authors (eg Moore and Thornton 1959, Mattox 1966 and Martin *et al* 1985) have reported that bombardment of the substrate by oxygen ions is the only way to ensure a good bond between silica and gold. Poor adhesion might lead to a partial loss of the gold film over the silica window, which would thereby decrease the effective value of γ_m over this area.

It can therefore be seen that there is some support for the notion that γ_m may not be uniform over the surface of the gold-film electrode. How might this explain the upcurving of the $\overline{\omega}(d)$ curves observed by Ernest? Before embarking on a quantitative description, it is perhaps as well to outline the process in qualitative terms. The fundamental reason for the second increase in $\overline{\omega}(d)$ is the spread of current across the cathode in the approach to electrical breakdown. A series of curves representing the flux density across the cathode at a constant *E/N* but at various values of d is shown in figure 6.3 on page 131. (These curves were calculated using the assumption that γ_m outside the window radius R_w was 1.7 times larger than the value of γ_m within this radius.) As can be seen from this figure, at

small values of the electrode separation d, the discharge is confined mainly to the region of the window, that is, to within $r < R_w$. Because of the proximity of the electrode planes, the majority of metastable particles are quenched by these planes before they can diffuse far in the radial direction. As d is increased, the metastables diffuse out further. There is a feed-back effect operating also, since the proportion of secondary to primary current rises as d is increased toward the breakdown separation. This means that metastables are not just passively diffusing to greater radii at larger d, but are being produced at greater radii as well. From equation (6.65), one can see that, if γ_m increases in value with radius across the cathode, the effective γ_m in equation (6.64) will become larger at greater d. If this increase is large enough to counteract the fall-off in the flux of metastables, then the graph of $\overline{\omega}(d)$ will curve upward again. Similar considerations apply to the ion contribution.

The Fourier-Bessel model described in section 6.5 was modified to accommodate a variation in γ_m of the form

$$\gamma_m(r) = \begin{cases} \gamma_{m,i}, & r < R_w \\ \gamma_{m,o}, & r > R_w. \end{cases} \tag{6.66}$$

(Variations in the ionic secondary coefficient were neglected, because ions produce a negligible proportion of the secondary current under the discharge conditions of present interest.) The expression for U_{njk} in equation (6.52) becomes

$$U_{njk} = T_0 + T_1, \tag{6.67}$$

where

$$T_0 = \frac{\rho_n^2 R_E^2 J_0(\theta_k R_E) K_1(\rho_n R_E)}{(\rho_n^2 + \theta_j^2)(\rho_n^2 + \theta_k^2)}$$

$$\times \left\{ \frac{\gamma_{m,o} - \gamma_{m,i}}{\rho_n R_E} R_w \left[\rho_n I_1(\rho_n R_w) J_0(\theta_j R_w) + \theta_j I_0(\rho_n R_w) J_1(\theta_j R_w) \right] - \gamma_{m,o} I_1(\rho_n R_E) J_0(\theta_j R_E) \right\} \tag{6.68}$$

and

$$T_1 = \frac{0.5}{\rho_n^2 + \theta_k^2} \times \begin{cases} (\gamma_{m,i} - \gamma_{m,o}) R_w^2 \left[J_0^2(\theta_j R_w) + J_1^2(\theta_j R_w) \right] + \gamma_{m,o} R_E^2 J_0^2(\theta_j R_E), & j = k \\ \dfrac{(\gamma_{m,i} - \gamma_{m,o}) R_w \left[\theta_k J_0(\theta_j R_w) J_1(\theta_k R_w) - \theta_j J_0(\theta_k R_w) J_1(\theta_j R_w) \right]}{0.5(\theta_k^2 - \theta_j^2)}, & j \neq k. \end{cases} \tag{6.69}$$

Theoretical $\overline{\omega}(d)$ curves produced using this model are compared in figures 6.4 and 6.5 with experimental measurements made by Ernest (1991) at two different regimes of *E/N*.

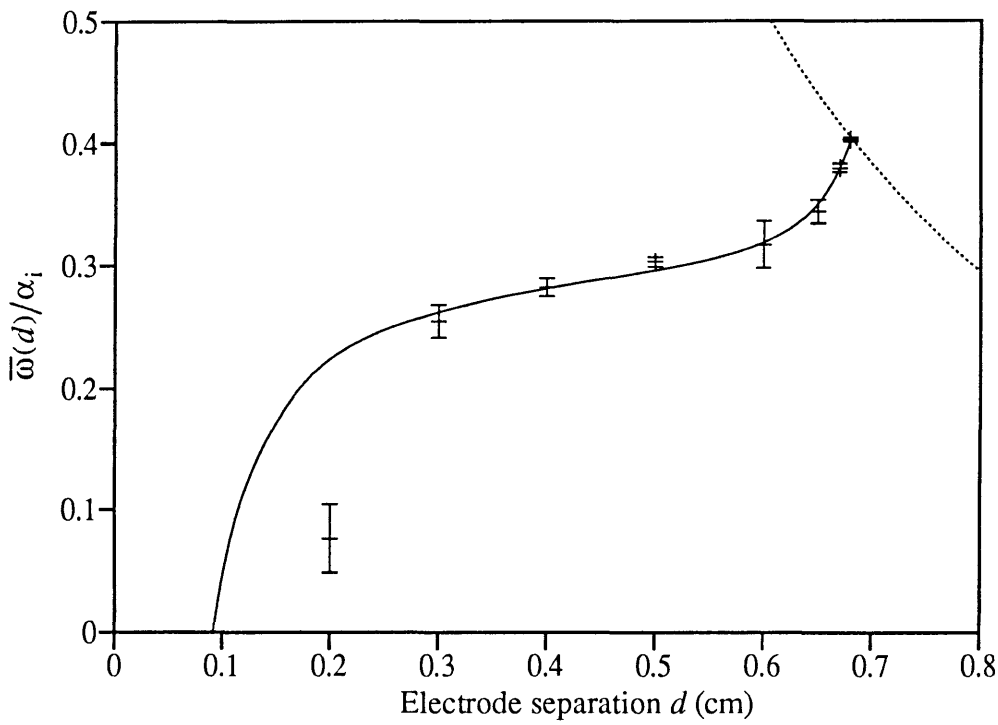

Figure 6.4. The variation of the generalised secondary ionisation coefficient $\overline{\omega}(d)/\alpha_i$ with electrode separation d. The gas is nitrogen at 5 Torr and a reduced electric field *E/N* of 283 Td (~ 100 V cm^{-1} Torr^{-1} at 0 °C). The experimental values of $\overline{\omega}(d)/\alpha_i$ were calculated from measurements of the total discharge current provided by Ernest (1995b). (These data can also be found, in graphical form, in Ernest (1991).) The uncertainties were calculated by assuming an uncertainty of 0.1% in the full scale reading of the ammeter. The dotted line represents the breakdown limit. The theoretical curve (solid line) was calculated using the Fourier-Bessel analysis of section 6.4 with the functional variation of $\gamma_m(r)$ given by equation (6.66). Twenty terms of the series were kept.

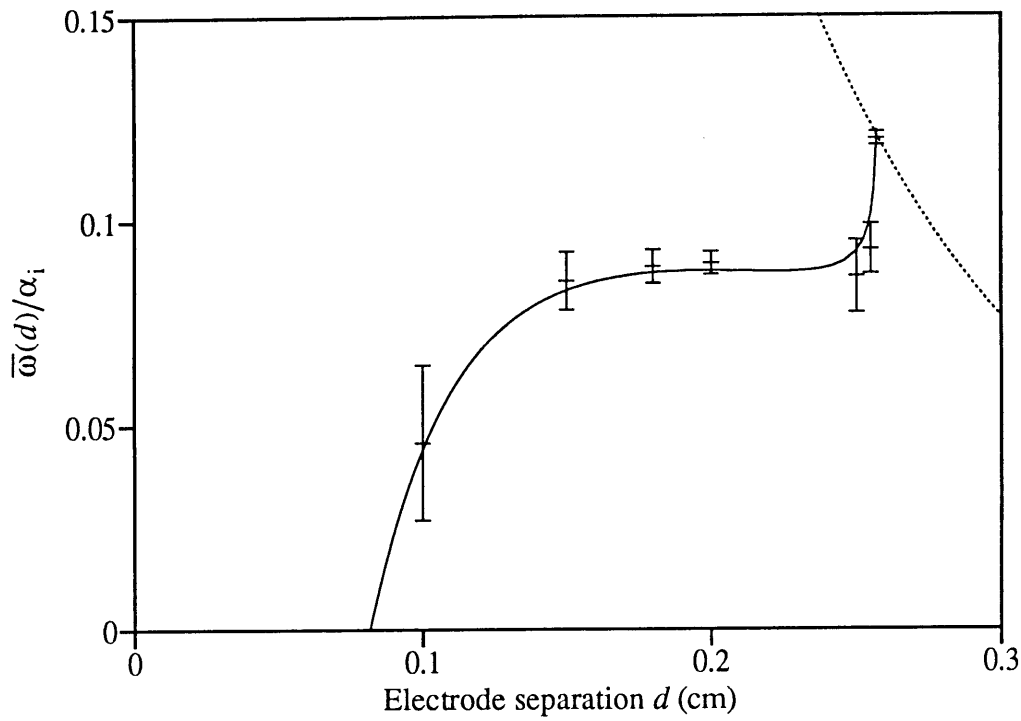

Figure 6.5. Same as figure 6.4, except the reduced electric field E/N in this case is 566 Td (200 V cm^{-1} Torr^{-1} at 0 °C).

Values of the total discharge current were converted to $\overline{\omega}(d)$ values by use of the approximation

$$\frac{\overline{\omega}(d)}{\alpha_i} \sim \left[1-\frac{I_{\text{prim}}\exp(\alpha_i d-\alpha_i d_0)}{I(d)}\right]\times\left[\exp(\alpha_i d-\alpha_i d_0)-1\right]^{-1}, \qquad (6.70)$$

in which electron diffusion is neglected. (The use of equation (6.70) is felt to be justified because the relationship between the experimental and theoretical data is more important for present purposes than the accuracy or otherwise of the $\overline{\omega}(d)$ values.) The discharge parameters which were used in the model were taken from Haydon and Williams (1976), Saelee *et al* (1977), Wedding *et al* (1985) and Nakamura (1987). A value of 5.5×10^{18} cm^{-1} s^{-1} was used for the diffusion coefficient D_m; this is consistent with the recent measurement of Haydon *et al* (1995). The curve was fitted by hand by varying the ratio $\gamma_{m,o}/\gamma_{m,i}$; the values of the primary current I_{prim} and the volume quenching rate G_m were also slightly adjusted. The final values of these parameters were as follows:

	Figure 6.4:	Figure 6.5:
$\gamma_{m,o}/\gamma_{m,i}$:	1.68	1.75
I_{prim} (pA):	13.95	7.0
G_m (s^{-1}):	5.0	5.0

The fit was sensitive to the first two of these parameters, but not to the third. A value of 2 s^{-1} for G_m gave results which were almost as good. A G_m value of several decays per second is consistent with published data for nitrogen at 5 Torr, provided that the gas contains about 10 - 20 parts per million of the common impurities H_2, CO or O_2 (Levron and Phelps 1978, Thomas *et al* 1987).

It can be seen that the agreement between the theoretical and experimental data points is good near to breakdown. However, in figure 6.4 at least, the curves diverge at small values of d. This is not apparent in the figure, because some of the experimental measurements of current generate $\overline{\omega}(d)$ values that fall below the x axis. These have not been plotted. This divergence at small d is probably due to the retention in the model of the step-function approximations for $\alpha_i(z)$ and $\alpha_m(z)$ (equations (2.4) and (2.20)).

It is intended to apply this theory to other sets of data in Ernest (1991), but at the time of writing nothing further has been attempted.

6.7. Conclusion.

Chapters 5 and 6 contrast two methods of solving the electron diffusion equation. The Hankel transform approach of chapter 5 is suited to experiments in which only the concentration of electrons is measured. It was shown in this chapter that the Hankel formulation allows one to deduce electron transport properties from such measurements, without needing any prior knowledge of the form of the cathode current. However, as discussed in the introduction to the present chapter, it is not possible to use the Hankel transform to solve the coupled electron and metastable-particle diffusion equations when the secondary coefficients vary with radius across the cathode plane. An alternative method was therefore developed, which involves the expansion of the electron concentration $n_e(\mathbf{r})$ in a Fourier-Bessel series. It has been shown in the present chapter

that this approach allows a good approximation to be made to the ratio between the total flux of metastables into the cathode and the integrated optical absorption along a single path through the discharge. This, in turn, allows the (effective) value of the secondary coefficient γ_m to be determined without the necessity for a large number of measurements of the metastable-particle concentration.

In the penultimate section, it was shown that the Fourier-Bessel analysis can be used to model a pre-breakdown discharge in which the values of the secondary coefficients exhibit a step-function increase with radius across the cathode. This model generates values of the generalised secondary coefficient $\overline{\omega}(d)$ that become larger in the approach to electrical breakdown, in contrast to the behaviour that is usually observed. The model predictions were compared to $\overline{\omega}(d)$ values derived from experimental data of Ernest (1991) for two separate values of the reduced electric field *E/N*. The agreement between experiment and theory provides support for Ernest's explanation of the upcurving which can be observed in his $\overline{\omega}(d)$ data.

7. Experimental apparatus.

7.1. The vacuum system and ionisation chamber.

Some of this equipment had been used previously by Brunker and is described in his thesis (Brunker 1984).

The vacuum system is sketched in figure 7.1.

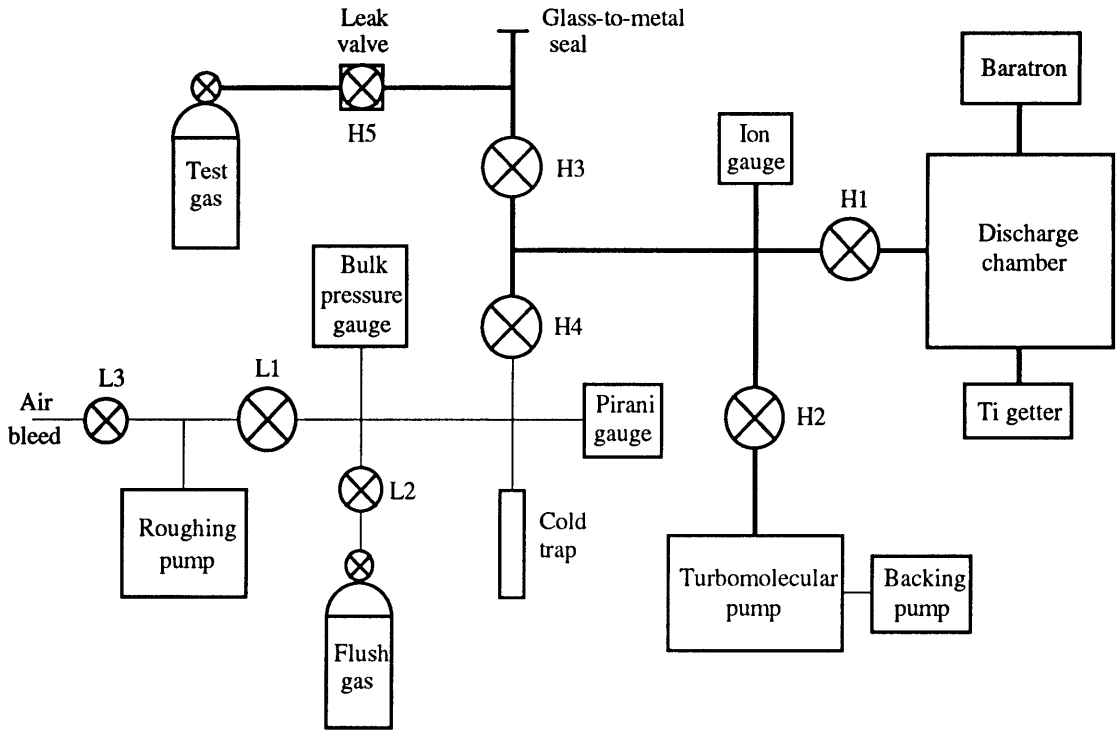

Figure 7.1. The vacuum system. Bold lines represent the sections under high vacuum.

Broadly speaking, the vacuum system can be divided into high- and low-vacuum sections. The high-vacuum section was assembled from prefabricated Varian Conflat sections, sealed with copper gaskets; the low-vacuum side was constructed of copper tubing with O-ring seals. The high-vacuum valves were all of Granville-Phillips type 204, except for one Vacuum Generators valve (tap H3 in figure 7.1). At the start of the study period the high vacuum was maintained by two Varian VacIon pumps, as described by Brunker (1984); later these were replaced by a Pfeiffer turbomolecular pump. The low vacuum section was pumped by a Rud Browne Dynavac rotary pump. The pressure could be measured by one of four gauges, which were: (i) a bulk pressure gauge, useful mainly during the 'cracking' procedure described below; (ii) a pirani

gauge (not used much as it was not considered to be reliable); (iii) a Varian ion gauge, which was used to measure high vacuum; (iv) An MKS Baratron capacitance manometer, which was used to measure the pressure of the argon gas fill. The ion gauge could read (in theory) 10^{-10} Torr FSD, whereas the lowest range of the Baratron gauge was 1 Torr FSD.

Impurities such as CO, N_2, O_2 and H_2O are readily dissolved in or adsorbed onto metals and are therefore frequent trace contaminants in high vacuum apparatus. These species are also efficient quenchers of metastable activity in many gases (see, eg, Velazco *et al* 1978, Thomas *et al* 1987, Tao *et al* 1987 and Golde *et al* 1989). It is therefore desirable to purge the vacuum system of these gases by heat treatment, if a significant concentration of metastable states is to be retained. The efficiency of this 'baking' process increases with temperature in an approximately exponential fashion. Unfortunately, the presence of Torr-Seal vacuum cement in parts of the discharge chamber prevents the chamber itself from being baked at temperatures higher than about 150 °C, although no such restriction applies to the remainder of the system. This restraint prevents the achievement of the desired reduction in the rate of outgassing of impurities from the chamber walls.

Since baking of the chamber was a lengthy and inefficient procedure (due to the temperature restrictions described above), every effort was made to prevent significant quantities of air from entering the vacuum system on the occasions when it became necessary to open the system. The procedure was as follows: Firstly, a bottle of welding-grade argon (99.99% pure) was connected to the low-vacuum section at tap L2. This tap was opened and the low-vacuum section was evacuated by using the roughing pump. The low-vacuum section was then flushed with argon once or twice. After this the cold trap was submerged in liquid air, the roughing pump was isolated by closing tap L1 and then tap H4, which links the low- and high-vacuum sections, was opened. Argon was admitted until the bulk gauge showed a fill pressure of slightly more than the prevailing atmospheric pressure. Tap H4 was then closed. Sections of the high-vacuum system could then be opened without the admission of a significant amount of air to the chamber. It was seldom necessary to re-bake the chamber after using this procedure.

Two gettering arrangements were tried during the study period. The first getter consisted of a single titanium wire about 12 cm long running down the centre of a Pyrex tube, which was mounted directly on the chamber. This wire could be heated by the passage of an electric current. If this was done while the system was evacuated, a

reactive layer of titanium was deposited onto the inner wall of the tube. This getter was found to be effective for a short time: after the first deposit of titanium the base pressure of the system decreased (briefly) from 3×10^{-8} to 4×10^{-9} Torr, while dG_m/dt, the rate of increase in the quenching parameter G_m, decreased from 8.4 to 0.3 sec^{-1} min^{-1}. However, running this getter was a nerve-racking procedure, because the thermal expansion of the heated wire caused it to bow towards the glass envelope. If the white-hot wire had been permitted to contact the envelope, it is likely that the glass would have fractured, resulting in an explosive aspiration of powdered glass into all parts of the vacuum system, including the turbomolecular pump. This risk provided a strong incentive to investigate alternative gettering arrangements. The design which was eventually chosen used twists of titanium wire contained in a fused-silica tube which was heated by an external oven, the operating temperature being about 800 °C. This getter could be used whilst hot, but the electric oven was found to introduce a substantial amount of $n\times50$ Hz noise onto the output of the Tektronix differential amplifier (see figures 7.3 and 7.4), so it was generally heated for only a few hours at a time. The gettering effect persisted for some hours after each heating cycle, the minimum value of dG_m/dt observed with the cold getter being 0.32 sec^{-1} min^{-1}, a similar value to that seen with the old getter. However, a similar reduction in the base pressure was not observed.

Contrary to initial expectations, heating or cooling of the getter was observed to have little effect upon the fill pressure of argon, the change during a complete cycle being about 3%. This is probably because the volume of gas in close contact with the heated surfaces of the getter is much smaller than that which is in contact with the room-temperature surfaces in the discharge chamber.

A typical rate of outgassing, determined by measuring the rise in pressure while the chamber was isolated for several weeks, was about 6.5×10^{-9} Torr litre s^{-1}, or about 2.1×10^{11} molecules per second. This is equivalent to an increase in the concentration of impurities in 1 Torr of argon of about 40 ppm per hour. The second getter appeared to be able to compensate entirely for this rate of outgassing, over a day or two at least. No change in the breakdown voltage could be detected in a sample left overnight with the getter running, whereas a change of 6 to 8 volts was not unusual over a similar time span when the getter was not used.

The argon gas used for experiments was of 99.9995% purity and was supplied by Matheson. It was admitted to the chamber through a Granville-Phillips leak valve (tap H5 on figure 7.1).

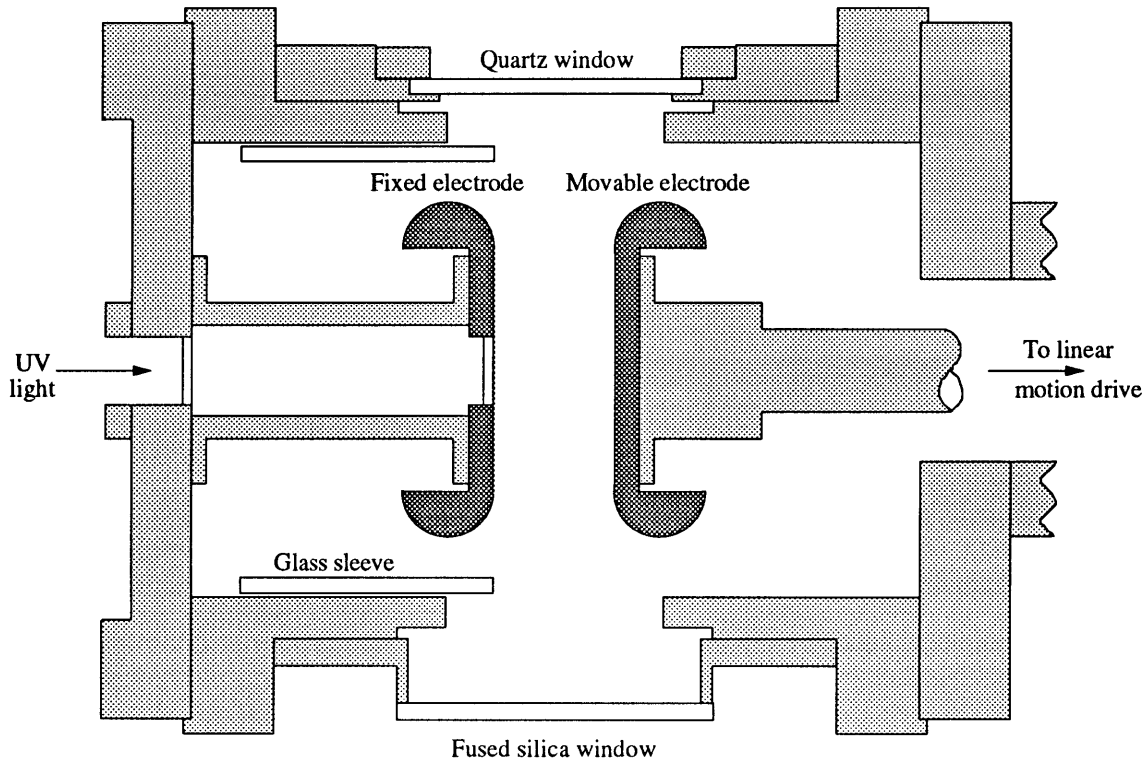

Figure 7.2. A section through the ionisation chamber, seen from above.

The ionisation chamber, described by Brunker (1984) and sketched in plan in figure 7.2, was machined from a nearly cubical parallelepiped of stainless steel of external dimensions 6" by 6" by 6 3/16". A horizontal cylindrical channel of diameter about 7.5 cm diameter was bored between the two square faces. Recessed 5¾" Varian Conflat flanges were machined into these faces. The opposing electrode assemblies are mounted on two 5¾" Conflat plates that seal the ends of this channel. The faces of the two remaining vertical sides have 4½" Conflat flanges machined into them; these faces are connected by a cylindrical bore which is perpendicular to the main channel. The ends of this cross-bore are sealed by windows of approximately 4 cm clear vision diameter. These windows permit a beam of probe radiation to cross the space between the electrodes. The top face of the chamber is machined with a 2¾" Conflat flange to which the Baratron is mounted; the bottom face is occupied by two 2¾" Conflat flanges which allow connection to the titanium getter and to the remainder of the vacuum system. The volume of the chamber was estimated to be about 0.6 lr.

The electrodes, which were cut from OFHC copper, are of cylindrical symmetry. The working faces, which are flat and normal to the symmetry axis, are of 1" radius. The metal at greater radii is machined to a circular profile. This profile is tangent to the plane of the working face and is ⅜" in radius. Both the flat and profiled surfaces are

highly polished. One electrode is fixed in position; the other is moveable through about 2.5 cm by use of a Varian linear motion drive. Ultraviolet light entering the vacuum chamber from behind the fixed electrode is introduced into the discharge volume through a circular hole of radius $\frac{5}{16}$" in the centre of this electrode. This hole is covered by a fused silica disk which is flush with the remainder of this electrode. A similar window seals a hole in the centre of the $5\frac{3}{4}$" Conflat plate to which the fixed electrode is mounted. Uniformity of the electric field within the discharge volume is preserved by coating the whole fixed electrode, including the silica disk, with a conductive, semi-transparent layer of gold, some tens of nanometres in thickness. The transparency of this membrane permits either electrode to be used as a photocathode, depending on the potential difference applied between the electrodes. The optimum degree of transparency is clearly about 50%. Some difficulties were experienced in obtaining the optimum thickness, however. The original film was excellent; unfortunately this was accidentally removed in the course of conversion of the vacuum pumping arrangements. Subsequent attempts to recoat the fixed electrode by vacuum vapour deposition were only partly successful; it was suspected that the gold was not adhering to the quartz substrate and was falling off after re-assembly, possibly during the baking procedure that was invariably necessary after exposing the chamber to the atmosphere. Various authors have commented on the poor adhesion between gold and silica (Moore and Thornton 1959, Benjamin and Weaver 1961, Mattox 1966). The best way to ensure a good bond is reported to be bombardment by oxygen ions (Martin *et al* 1985), but this was not practical with the available equipment. The film used in the present study is probably still a little on the thin side. It allows plenty of ultraviolet light through to the copper electrode, but produces a very small photoelectric current itself. Nevertheless, the film exhibits no evidence of the sort of charging behaviour that might be expected if the conductive layer was discontinuous.

The gap between the electrode edges and the inner diameter of the chamber is only about 1 to 1.5 cm, which can lead to problems with unwanted electrical breakdowns between the fixed electrode and the chamber wall. In an attempt to prevent this from happening, a cylindrical glass sheath or sleeve of outer diameter slightly less than the inner diameter of the chamber bore was placed around the fixed electrode. This sleeve has to be positioned rather carefully. On the one hand, the sleeve should not prevent access by the laser probe to any part of the inter-electrode space; on the other hand, discharges to the chamber wall occur with annoying frequency if the edge of the glass is more than 3 or 4 mm behind the plane of the fixed electrode. The edge of the sleeve was therefore aligned as accurately as possible with the planar surface of the fixed electrode.

Brunker (1984) reported that the side windows were made of quartz, but because the (much cheaper) fused or amorphous silica is often loosely referred to by this name, the present author assumed initially that the windows were in fact made of the latter material. However, tests of the polarisation of the probe laser indicated that at least one of these windows was made of a birefringent (and therefore crystalline) substance. Measurements of the Brewster's angle of each window returned identical values of 1.48 for their refractive indices (a typical value for silica); from back-reflections of polarised light it was further surmised that the window on the side where the diode laser assembly was mounted was of fused silica, whereas the other was crystalline quartz.

7.2. Signal production and acquisition.

7.2.1. Measurements of discharge current.

The signal production and acquisition apparatus is sketched in figure 7.3.

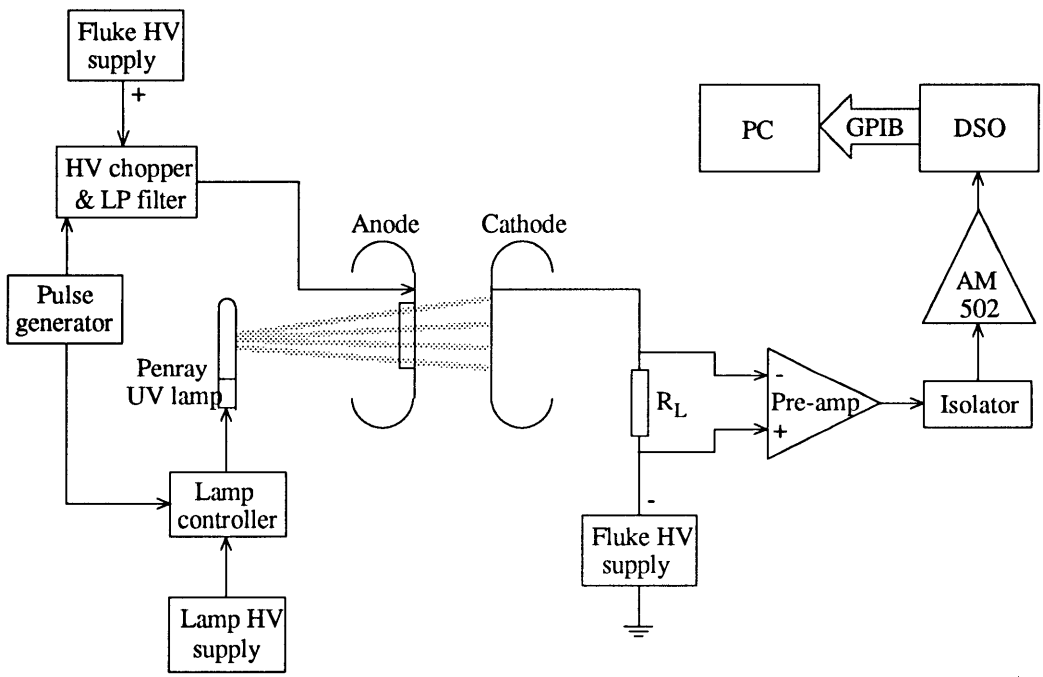

Figure 7.3. Arrangements for generating and detecting the discharge current. The component labelled 'R_L' is a load resistance, 'AM 502' is a differential amplifier, 'DSO' a LeCroy digital storage oscilloscope and 'PC' an IBM type computer.

The primary current is produced by irradiating the electrodes with a Penray mercury ultraviolet lamp producing about 0.06 watts of (mostly 254 nm wavelength) light. This illumination generates about 150 pA of current from the copper electrode. The lamp is powered by a Hipotronics high voltage supply; this is connected to the lamp via a control circuit which permits the lamp to be turned off and on by the application of a small controlling voltage. In practice the lamp was either left on all the time or repetitively 'chopped' by connecting the output of a square-wave generator to the control input. After the lamp is switched on, its output exhibits an initial fast rise in the space of a few μs. The output intensity then remains at a constant level for about 50 ms, after which it begins to decrease, eventually reaching about 2/3 to 1/2 of its initial plateau value. Although there is considerable time jitter in the response of the lamp to the turn-on signal, the turn-off is regular to within 5 μs. It is therefore normal practice to accumulate and average signals from the decay of current after lamp switch-off, at least when the lamp voltage is used to trigger the digital storage oscilloscope.

Because the light source is long and thin in shape, it might be expected to produce a non-circular area of primary current, particularly on the copper (ie, moveable) electrode. If this were true it would lead to difficulties in the analysis of optical absorption data, because the techniques developed in chapters 5 and 6 are only valid if the primary current source has cylindrical symmetry. However, it is thought that the masking effect of the hole in the fixed electrode (this being 10 cm from the lamp, but less than 2.6 cm from the opposite electrode) permits the effect to be neglected.

A Fluke 408B high-voltage supply is used to apply a positive potential to the fixed electrode. High tension can be applied to this electrode either directly, through a 'chopping' circuit or through a low-pass noise filter. Provided that the applied voltage is not such as to cause a spontaneous breakdown of the gas, current passes from cathode to anode only when lamp radiation and a significant potential difference are both present. Any current transmitted by the discharge is detected as a voltage across the load resistor R_L in figure 7.3. This signal is amplified by a Tektronix AM 502 amplifier; the output from this is directed to a LeCroy 9400 digital storage oscilloscope ('DSO' in figure 7.3). The digitised and stored signal can be transferred to an IBM-type microcomputer ('PC') via a General Purpose Interface Bus ('GPIB'). A pre-amplifier stage is occasionally inserted between R_L and the AM 502 amplifier. Measurements of the steady-state discharge current can be made by replacing the resistor, amplifier, DSO and computer by a Vibron 62A electrometer.

The moveable electrode is usually connected directly to ground through R_L, but it can be negatively biased if necessary, as shown in figure 7.3. This was found on occasion to be desirable so as to reduce the risk of electrical breakdown from the fixed electrode to the chamber walls. If both electrodes are biased with opposite polarity, the maximum difference in potential between either electrode and the chamber wall can be reduced to about one half of the PD between the electrodes. By using this arrangement, the chance of a breakdown to the chamber walls is much reduced. Note that, if the second HV supply is connected, some additional circuitry is necessary to isolate both the power supply and output signal of the pre-amplifier.

The geometry of the electrodes leads one to expect that the total amount of light, and therefore the amplitude of the primary current, should be independent of the electrode separation *d*. However, when the primary current from the copper electrode was measured at various values of *d*, there was found to be a slight variation (see figure 7.4). This is perhaps due to differences in the photoelectric efficiency of various regions of the copper surface, and/or to variations in the amount of light reflected from the anode.

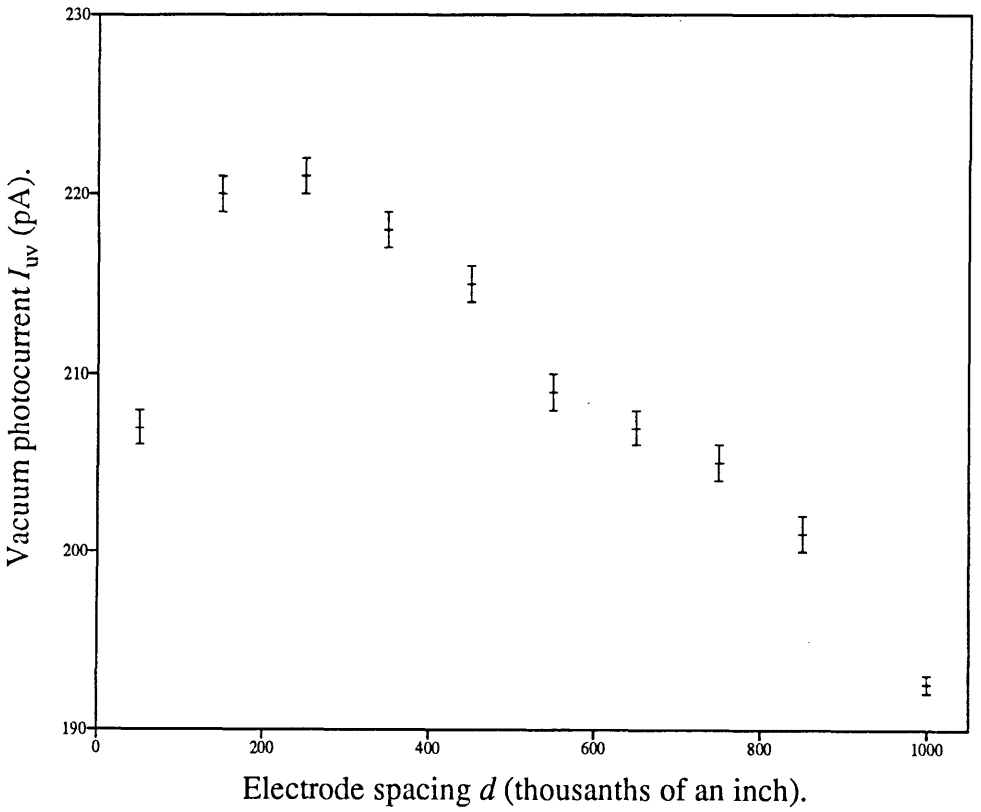

Figure 7.4. The variation in the vacuum photocurrent with electrode spacing. The copper (moveable) electrode is the cathode. Note the suppressed zero on the *y* axis.

7.2.2. The laser absorption experiment.

7.2.2.1. Laser control circuitry.

The lasers used in the experiment were all GaAlAs double heterojunction, gain-guided semiconductor lasers, known colloquially as 'laser diodes'. Among the types used were the Sharp LT010, the Toshiba TOLD9200 and the Hitachi HL7801E. The output spectra of these lasers are characterised by an array of longitudinal modes separated by about 0.3 nm. The individual modes are reported to be about 10-20 MHz wide (eg Takakura *et al* 1980), which is much smaller than the (Doppler-broadened) widths of the argon 1s-2p transitions at room temperature. This was confirmed by scanning the centre wavelength of the laser through an argon 1s-2p transition. The shape of the resulting absorption feature is a convolution of the transition shape and the laser emission profile. If the laser width is significant compared to the transition width, the width of the absorption feature as a whole should be larger than that of the transition. However, this was not found to be the case, within the limits of measurement.

Although all the lasers listed above are technically multi-mode, one of the laser modes is usually dominant, particularly at high output power. However, in no case was the power in the dominant mode observed to exceed 50% of the total output power.

Two further features of diode lasers that make these lasers useful for spectroscopic applications are: (i) the output of the diode laser is linearly polarised; (ii) the mode wavelengths are tuneable by varying either the current or temperature of the diode.

The laser package is mounted on a carriage which is sketched in figure 7.5. This carriage can be moved across the face of the chamber on vertical and horizontal rails. All parts of the discharge are thus accessible to the beam. The laser package is mounted on a thin aluminium plate; behind this on either side of the laser are placed Melcor thermo-electric or Peltier effect coolers. These are sandwiched between the laser mounting plate and a water-cooled plate of approximately the same dimensions. The temperature of the laser can be varied from about 0 to 50 °C by varying the polarity and amplitude of the current through these Peltier devices. Lower temperatures would require either pre-cooling of the water supply or the addition of another layer of thermo-electric coolers.

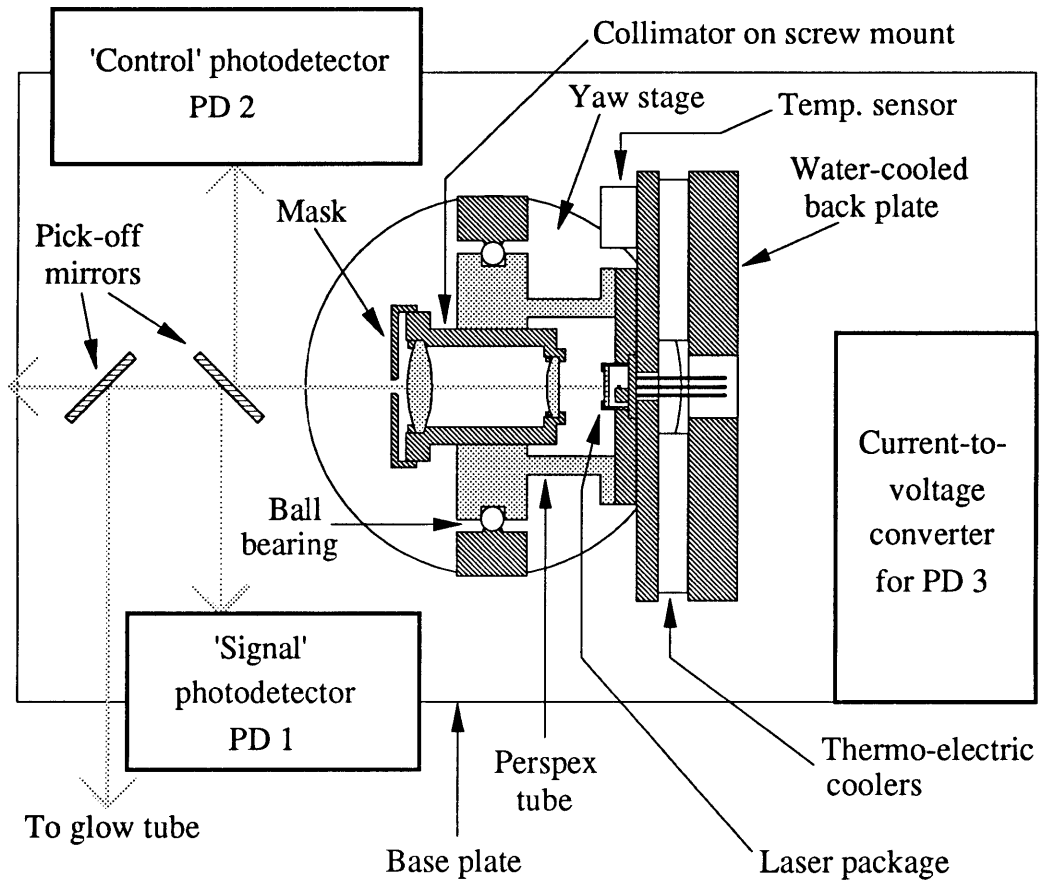

Figure 7.5. The laser carriage, in section, seen from above.

The output of a laser diode is very divergent, spreading at an angle of about 30° to 40° in the plane normal to the plane of orientation of the semiconductor layers. It is therefore necessary to collimate the beam with a short-focal-length lens of considerable numerical aperture if a narrow beam of adequate intensity is to be achieved. A microscope objective of 6 mm focal length and a numerical aperture of about 0.4 is used in the present experiment. The output from this lens is a beam of elongated profile, roughly 10 mm high by 4 mm wide. The laser package is positioned so that the long axis of this profile lies in the vertical direction. A mask with three vertically aligned holes is mounted on the front of the collimator so as to divide the output into three beams of circular cross-section and approximately the same intensity. Although diode lasers are reputed to show a degree of astigmatism (Wieman and Hollberg 1991), this was not found to be a severe limitation in the present experiment: it was found that the beams emerging from the mask maintained an acceptable profile for at least a metre or so. The profile of the middle beam is displayed in figure 7.6.

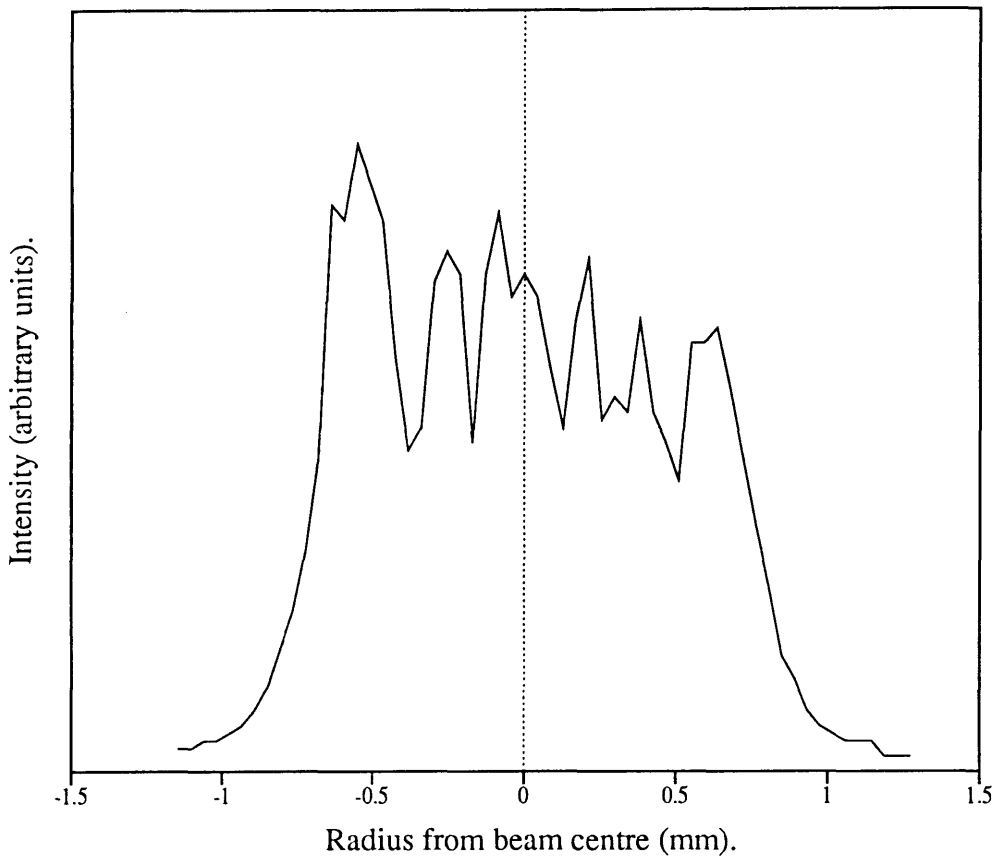

Figure 7.6. The intensity profile of the midmost of the three beams which emerge from the laser mask. The beam was examined at a distance of 12 cm from from the mask, which is approximately the position occupied by the discharge. The profile was measured by translating across the beam diameter a detector comprising a 100 μm pinhole fixed to the front of a photodiode. The superimposed fringes probably arise somewhere in the collimation optics.

The laser package/Peltier assembly is connected to the collimating lens by a Perspex tube. The tube is made of Perspex for the purpose of both thermal and electrical insulation. Space was left in this tube for the introduction of a little silica gel to prevent frosting of the laser package window if it was found to be necessary to cool the laser to sub-zero temperatures. The collimator screws into a threaded hole in the Perspex tube, thereby allowing the focus of the output beam to be adjusted. The whole laser/collimator/cooler assembly is mounted upon a stage which is rotatable about two orthogonal axes which are both perpendicular to the beam direction. These movements are controlled by separate micrometer screws.

Two pick-off mirrors are positioned on the laser carriage to intercept the upper and lower of the three beams. The upper beam is directed through a glow-discharge absorption cell (described below) and thence to a Jarrell Ash 0.5 m spectrometer; the lower beam is directed to a 'control' photodetector which is mounted on the laser

carriage. The middle beam is allowed to continue into the chamber where it crosses the inter-electrode space in a direction parallel to the electrode faces. A mirror could be attached to the far side of the chamber in order to reflect this beam back through the discharge volume. The angle of this mirror is adjusted so that the reflected beam impinges on the back face of the lower pick-off mirror, from which it is reflected into a second, 'signal' photodetector which is also mounted on the laser carriage. This two-pass arrangement doubles the amplitude of the absorption signal and allows the 'signal' photodetector to move with the laser carriage. However, optical feedback from the mirror was found to cause instabilities in the amount of power in the main laser mode. Signals were therefore usually obtained without using the mirror. On these occasions, the 'signal' photodetector was positioned on the far side of the discharge chamber and moved in concert with the laser carriage.

Optical feedback is a factor that requires a lot of attention when using laser diodes. If even a very small amount of the output radiation is reflected back into the laser, both the frequency and amplitude stability of the laser can be compromised (Wieman and Hollberg 1991, and references within). The optical layout is therefore designed to minimise reflections. For this reason, fully silvered, 45 degree pick-off mirrors are used instead of cubical beamsplitters, the main beam is aligned at a small angle (~ 5 deg) to the discharge-chamber windows, and the end windows of the glow-discharge absorption cell are set at angles of several degrees to the beam direction. It was not possible to do anything about reflections from either the flat entry face of the laser collimation lens or the spectrometer entrance slit. However, the feedback produced by these sources was evidently too small to affect the laser stability.

Figure 7.7 shows the general layout of the laser control circuitry and the arrangements for detecting the absorption signal. There are two feedback systems associated with the laser. The first, and by far the most complicated, is used to keep the wavelength of the main output mode as near as possible to the centre of the atomic transition of interest. To do this, the laser driving current is modulated at about 50 kHz, the modulation amplitude being about 10 μW. This gives rise to both amplitude- and wavelength-modulation of the output. As recorded above, the uppermost of the three beams which emerge from the mask is directed through a glow-discharge absorption cell and into a Jarrell Ash monochromator. (The effect of the glow tube on the total laser power is monitored by detector PD4.) The intensity of the light transmitted by the monochromator is measured by an EMI 9658R photomultiplier ('PMT' in figure 7.7). The glow discharge contains argon and exhibits strong absorption features at all the dipole-allowed 1s-2p transitions. The monochromator,

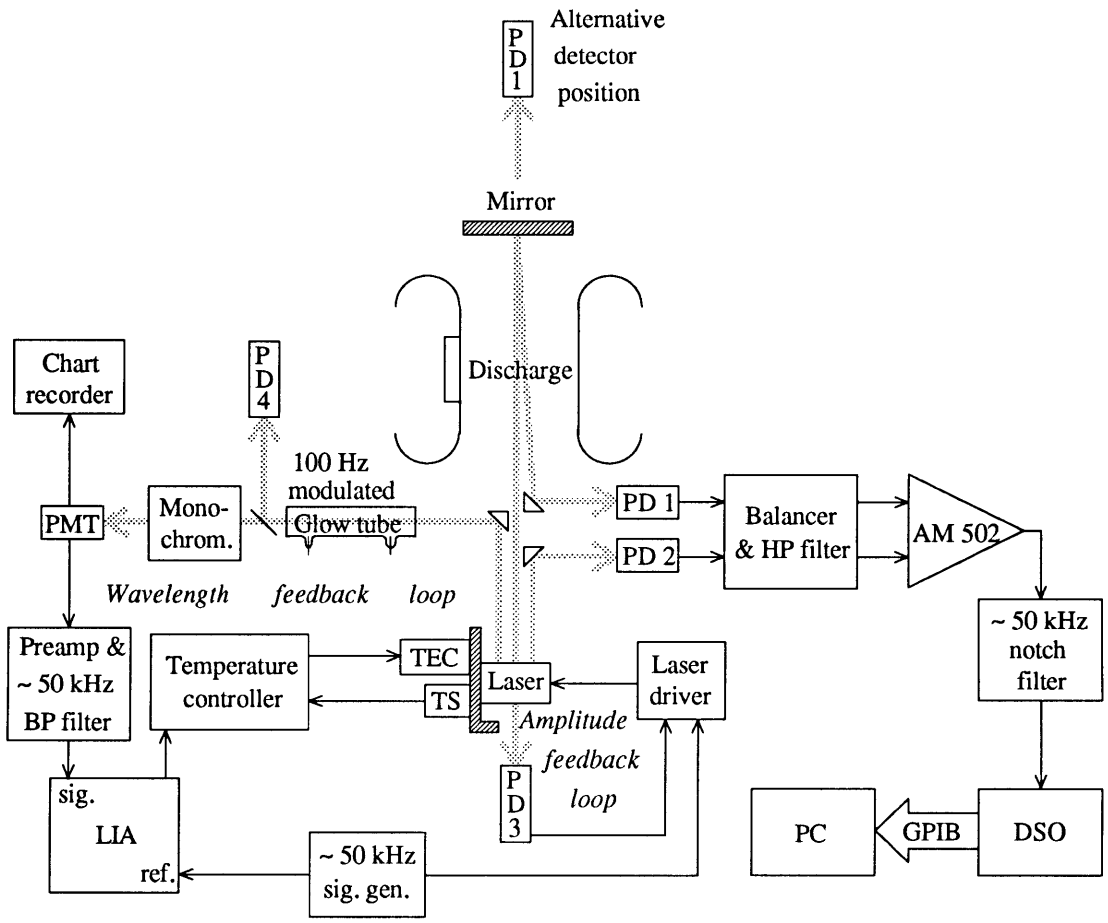

Figure 7.7. Diode laser control and absorption signal detection circuitry. Abbreviations are defined in the text.

which has an instrumental linewidth of about 0.03 nm, serves to select the laser mode which is tuned near to the transition of interest. This mode acquires an additional amplitude modulation during its passage through the absorption cell, with an amplitude which is proportional to the rate of variation of absorption with wavelength; in other words, it is proportional to the first derivative of the lineshape. This modulation is detected by the photodetector PMT; the output signal is passed through a band-pass filter, a pre-amp, a second band-pass filter (EG&G model 189) with a Q value of 50, and finally to an EG&G 128A lock-in amplifier ('LIA' in figure 7.7). The output from the LIA is directed to the temperature controller, where it is added, with an appropriate weighting, to the control voltage, which is set manually. A second input to the controller comes from an LM3911 temperature sensor ('TS'). The difference signal between this input and the combined manual/LIA signal is used to drive the thermo-electric cooler ('TEC' in figure 7.7) and thus control the output wavelength of the laser.

Photodetectors PD 1, PD 2 and PD 4 are all constructed in the same way. A United Detector Technologies 6DI pin photodiode with a collecting area of about 0.2 cm^2 is placed across the terminals of a Zeltex ZA801M1 FET-input op-amp; a feedback resistance of one or two hundred kΩ is used in parallel with an 8.2 pF ceramic capacitor. The Zeltex op-amps were chosen because of their excellent noise properties. For additional noise protection, these op-amps were powered either by batteries or by an isolated DC power supply. The gain of the detectors was found to be independent of the modulation frequency of the input signal up to about 10 kHz. For measurements requiring a higher frequency response, Analogue Devices AD515 op-amps were substituted. The photodetectors as a whole display an excellent linearity of response. This was tested by illuminating the detector with a combination of a variable DC light source and a chopped light source of constant amplitude. The amplitude of the AC part of the signal was examined as the DC illumination was varied. No difference in the AC amplitude could be detected in the range 0 to 10 V DC output. (See also figure 7.8.)

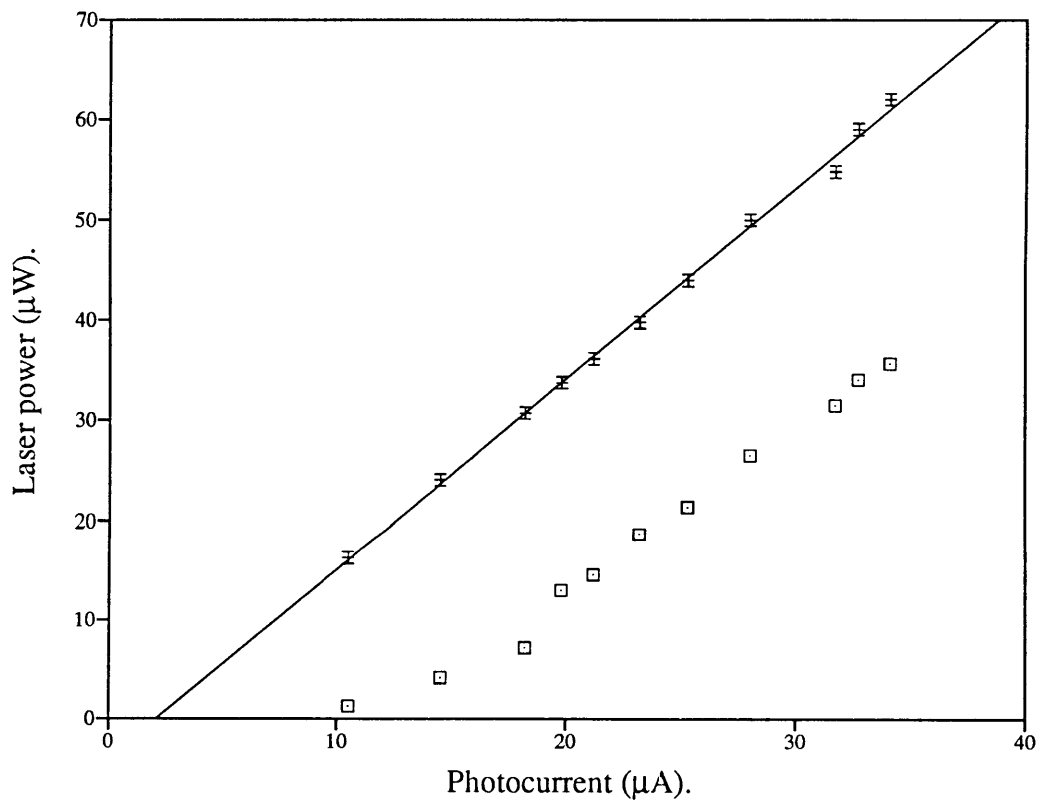

Figure 7.8. The relationship between the photocurrent generated in detector PD1 and the laser power. The crosses with error bars give the total power in the central beam; the solid line was fitted to this data by linear regression. The squares represent the power in the main mode.

In principle the amplitude-modulated signal could be detected by a simple photodetector positioned directly behind the glow tube; however, the large amount of power in the other laser modes makes this impractical. As described above, all the modes leave the laser modulated in both wavelength and amplitude; only the main mode acquires the additional, wavelength-dependent amplitude modulation during its passage through the glow tube. If all the modes were detected indiscriminately the wavelength-dependent part would be swamped. It is therefore necessary to selectively detect the main mode. The best way to do this is by the use of etalons, because a high degree of extinction of the unwanted modes may be obtained without much effect on the selected mode. In the absence of suitable etalons the Jarrell Ash spectrometer had to be used. This extinguishes the unwanted modes completely, but unfortunately it also reduces the power in the selected mode by a large amount. This makes it difficult to ensure an adequate signal-to-noise ratio in the output signal. The maximum photocurrent detected at 50 kHz is only about 100 pA.

Two absorption cells were constructed, both being cold-cathode glow tubes of the Geissler type. For best performance as a wavelength reference, an absorption cell should display lines which are as narrow as possible and neither too shallow nor too deep. In appendix A it is shown that the optimum depth for absorption features is 63%. Unfortunately, neither tube meets these criteria very well. The first, which was filled with about 0.9 Torr of gas, has an optical path length of about 14 cm. This tube can be operated at currents greater than about 0.2 mA and exhibits deep absorption features at all values of current. The depth at line centre of the $1s_5$-$2p_9$ absorption feature is about 90%. The second tube was constructed in an attempt to improve upon the precision available from the first tube. The optical path length of the second tube was restricted to 6 cm and the fill pressure was optimised before the tube was sealed. In addition, provision was made for the tube to be cooled in liquid nitrogen (LN2) or liquid air in order to narrow the absorption lines. This was done by constructing the tube with a 21 cm long U-shaped extension which can be lowered into a Dewar of LN2. Although the condensation temperature of argon is higher than that of either nitrogen or air at atmospheric pressure (87 K as opposed to 80 K for air), its vapour pressure at 80 K is 305 Torr (Clark *et al* 1951), comfortably higher than the fill pressure of 0.85 Torr. Provision was also made to slightly warm the entry and exit windows of this absorption cell to prevent the accumulation of frost.

Despite the precautions which were taken, the absorption lines observed in the second tube are still of greater than optimum depth. Part of the reason for this may lie in the size of the cathode. Tube 1 has a cathode made from an aluminium cylinder of 3

cm radius and 10 cm in length. This provides a large surface area for secondary emission, which allows this tube to be operated at a relatively low current. The second tube, in contrast, has a small cathode of 0.2 cm radius and length 2 cm. The minimum operating current of this tube is about 1 mA. Lower concentrations of metastable particles, and thence shallower absorption lines, may therefore be obtainable by using a tube with a larger cathode (and perhaps also a shorter optical path).

In order to determine the fractional absorption suffered by the beam which traverses the discharge it is necessary to know the DC intensity of the main mode. Unfortunately, the photodetectors are not wavelength-selective; all of the side-modes of the laser are detected as well as the mode which is tuned to the transition. It is therefore necessary to have some means of measuring the fraction of the total power in the main mode at any one time. This is done by applying a rectified but unsmoothed high voltage to the glow tube wavelength reference. The result of this is to produce absorption features in this tube with depths which are modulated at 100 Hz. For example, the depth of the $1s_2$-$2p_9$ absorption feature generally varies from 3% to about 80% during each cycle. A comparison of the 100 Hz modulation on the outputs of PD4 and PMT allows the fraction of power contained in the main mode to be determined. The 100 Hz modulation was found to have no effect on the wavelength stabilization feedback system.

The second major feedback loop in the laser control mechanism acts to reduce amplitude noise which would otherwise be present on the laser output. This noise, which has a pronounced $1/f$ component, probably arises in the transistors within the output stage of the laser driver. The noise is detected by photodetector PD 3, which is a photodiode built into the laser package by the manufacturer; the output from this photodiode is passed through a band-pass filter with cutoffs at 2 kHz and 30 Hz and thence to the laser driver. The amount of feedback can be controlled via a potentiometer on the panel of the driver. The band-pass filter prevents the feedback mechanism from attempting to nullify either the 50 kHz modulation or manual adjustment of the DC driving voltage.

The laser is driven by a current supply of simple design (Tachibana 1991), consisting essentially of a summing amplifier which serves to add the various control voltages; the output of this goes to the control input of an amplifier which provides the driving current necessary to operate the laser diode. This circuit proved to be very efficient, but, as mentioned above, it generated a considerable amount of noise on the laser output. Early attempts to reduce this noise by means of a filter on the output of

the driver were unsuccessful. A simple RC filter is not practical, because the load resistance of a laser diode is only a few ohms. This small value of R requires the use of an enormous capacitance to achieve an adequate band-edge. Inductive filters are more practical in terms of the component values, but were found to pick up a large amount of 50 Hz etc from the mains power supply. However, the feedback loop described above was eventually adopted. This proved very effective in reducing all the major types of noise on the laser output.

An important consideration when operating laser diodes is their extreme sensitivity to transient voltages (Wieman and Hollberg 1991). Hence the driver circuit is shielded from electrostatic noise, and shielded cable is used to connect the driver to the laser carriage. The carriage is electrically isolated from the discharge chamber; its sole connection to ground is via the laser driver. The turn-on procedure is to turn the driver on while the laser is disconnected; the modulation and DC bias are then set to zero, and the connection to the laser established. No laser failures were experienced while these precautions were used.

7.2.2.2. Detection of the absorption signal.

The absorption signal was detected by the 'signal' photodetector ('PD 1' on figure 7.7). The construction of the 'signal' and 'control' detectors has been described above. (UDT 6DI photodiodes were used because of their large light collection area. The large area minimises noise arising from 'walk-off' of the beams due to mechanical vibrations.) The output from these detectors is directed to a balancer/high-pass filter and from there to a Tektronix AM 502 differential amplifier. (The dual-detector arrangement is used to further reduce noise arising from sources common to both detectors, in particular, the 50 kHz modulation of the laser intensity, which was necessarily above the cut-off frequency of the amplitude feedback loop described earlier.) The output from the amplifier is digitised and stored by a LeCroy 9400 digital storage oscilloscope in the same way as the discharge current signal described above.

The photodetectors are shielded and electrically isolated from both the laser carriage and the discharge chamber, being connected to ground only through the AM 502 amplifier.

Some difficulty was encountered in matching the frequency responses of the two photodetectors. It is desirable to do this to assist in the removal of the 50 kHz

modulation which, if left untreated, would have dominated the detected signal. The Zeltex op-amps, as described above, have an upper band edge of about 12 kHz; above this point, however, their individual responses were not found to be uniform. The gain of one fell off at the expected 6dB per octave, but for some unknown reason the gain of the other decreased at about twice this rate. Matching was eventually achieved by a process of trial and error in the selection of filter and feedback resistors.

The 50 kHz modulation was, in fact, a major nuisance. The amplitude of this modulation was more than 40 dB larger than the amplitude of the absorption signal at the output of the photodetectors. Four separate mechanisms were used to reduce this source of noise. Firstly, there is the frequency response of the photodetectors, which at 50 kHz is about 15 dB below their pass-band gain. The second line of defence is the difference scheme described above, which gives an unknown amount of attenuation at 50 kHz. Thirdly, the output from the AM 502 is directed to the DSO through a twin-T type notch filter, which gives a full 40 dB of extinction at 50 kHz. The Q-value of this filter is about 15. This implies that it produces less than 3 dB of unwanted attenuation within the signal acquisition band below 10 kHz. Finally, any remaining 50 kHz noise can usually be removed by digital filtration (see section 7.2.3).

Some form of high-pass filter is necessary between the photodetectors PD 1 and PD 2 and the amplifier: otherwise the DC output of the detectors, which is typically 1 or 2 volts, would overload the amplifier. Although the amplifier possesses an internal high-pass filter, available by selecting 'AC' input coupling, the band edge of this internal filter is about 2 Hz. This was found to be too high for many experimental regimes. The photodetectors are therefore connected to the differential amplifier (with DC coupling selected) through a high-pass filter with a band edge of $\frac{1}{30}$ Hz. Even with the lowered band-edge available by use of this custom-made high-pass filter, a significant amount of 'droop' is noticeable on slower signals. The remaining droop can, however, be almost completely removed by digital means (see chapter 8, section 8.1).

8. Methods of digital signal processing.

Nearly all the experimental results obtained during the study period were recorded as digitised traces of the variation of a signal voltage over time. The ability to transfer these traces to a PC-type computer makes it possible to employ an armory of computational techniques to extract the parameters of interest. These techniques fall into three categories: (i) removal of distortions, (ii) improvement of signal-to-noise ratio and (iii) fitting a specified curve to the data. These categories are discussed individually below. All the techniques described were implemented using computer programs written by the author.

8.1. Removal of 'droop'.

As mentioned above, the laser absorption signal is passed to a differential amplifier through a high-pass filter; it was also mentioned that this filter imposes a noticeable degree of 'droop' on the data. This can be removed by inverse digital filtering, which is described in the present subsection.

Smith and Cohn-Sfetcu (1975) have discussed ways to simulate the effect of several types of filter by digital means. A similar but somewhat simpler technique is described below. The intent of this analysis is to show how the effect of a high-pass RC-type filter may be reversed by a digital manipulation of the data. The first step is to show how the filter itself might be imitated by a digital algorithm; this algorithm is then inverted to arrive at the desired result.

The input and output voltages $V_i(t)$ and $V_o(t)$ of a simple RC-type high-pass filter are related by the differential equation

$$\frac{dV_o}{dt} + \frac{V_o(t)}{RC} = \frac{dV_i}{dt}. \qquad (8.1)$$

Integration of this expression gives

$$V_o(t) = \left[V_o(0) + \int_0^t dt' \frac{dV_i}{dt'} \exp(t'/RC) \right] \exp(-t/RC). \qquad (8.2)$$

A digital simulation of this relationship must necessarily deal with a sampled or discretised input signal. The digital algorithm can therefore be derived from equation (8.2) by modelling the sampled input signal by a step-function defined by

$$V_{\mathrm{i}}(t) = V_{\mathrm{i},n}, \quad n\Delta t < t < (n+1)\Delta t, \tag{8.3}$$

where Δt is the time between successive samples. Although greater attention to detail in modelling the sampling process would result in a more accurate treatment, equation (8.3) is an adequate approximation in the case that $\Delta t << RC$. From equation (8.3), dV_{i}/dt can be shown to be given by

$$\frac{dV_{\mathrm{i}}}{dt} = \sum_{n} \left(V_{\mathrm{i},n} - V_{\mathrm{i},n-1}\right) \delta(t - n\Delta t), \tag{8.4}$$

where $\delta(t)$ is the Dirac impulse function. If this expression is substituted into equation (8.2), and account is taken also of the sampled nature of the output, the kth sample of the output voltage V_{o} is given by

$$V_{\mathrm{o},k} = \left[V_{\mathrm{o},0} + \sum_{n=1}^{k} \left(V_{\mathrm{i},n} - V_{\mathrm{i},n-1}\right) \exp(n\Delta t/RC) \right] \exp(-k\Delta t/RC). \tag{8.5}$$

It is not difficult to show that this expression reduces to

$$V_{\mathrm{o},k} = V_{\mathrm{o},k\text{-}1} \exp(-\Delta t/RC) + V_{\mathrm{i},k} - V_{\mathrm{i},k\text{-}1}. \tag{8.6}$$

The 'droop' imposed by a high-pass RC filter can therefore be reversed by inverting this relationship, the corrected voltages $V_{\mathrm{c},k}$ being given by

$$V_{\mathrm{c},k} = V_{\mathrm{c},k\text{-}1} + V_{\mathrm{o},k} - V_{\mathrm{o},k\text{-}1} \exp(-\Delta t/RC). \tag{8.7}$$

Matters are complicated slightly if there is some DC offset present on the signal at the point where it is digitised. This can occur, for example, if the signal is amplified after passing through the high-pass filter. The offset can be measured by finding the time average of V_{o} (which is assumed to be periodic) over a full period. This value should be subtracted from the data before equation (8.7) is applied.

8.2. Improving the signal-to-noise ratio.

Two principle techniques were used to improve the signal-to-noise ratio, corresponding to the two types of noise encountered, namely periodic and broad-band noise. Periodic noise was reduced using the following Fourier transform technique. First, a preliminary effort was made to fit the desired function to the data. This fitted function was subtracted from the data. This step was necessary to prevent the Fourier components of the signal swamping any periodic noise components that might be present. A discrete Fourier transform was applied to the remainder, using the Cooley-Tukey algorithm (Cooley and Tukey 1965). Any obvious noise components were removed and the remainder then back-transformed. The final result of this process was obtained by adding the fitted curve to the cleaned-up noise trace. A second, more accurate curve fitting could then be performed on the data.

Broad-band noise is often reduced by use of a simple low-pass RC filter. This type of filter is particularly effective when applied to signals that consist of a sum of exponential decays. This is because the time constants (but not the amplitudes) of the decays are unaffected by the filtering action (Smith and Cohn-Sfetcu 1975). There are additional advantages in using a digital simulation of such an RC filter. Firstly, the value of the time constant of a simulated filter can be varied at different parts of the data trace. This can be as simple as turning the filter on after a transient. By this means, the sharp fall in discharge current after the cessation of the primary source may be retained while the subsequent slow decay is smoothed. However, as is shown below, a continuous variation of the time constant may yield further advantages.

The digital filter algorithm can be obtained in a manner similar to that described in section 1. The input and output signals of a low-pass filter are related by the differential equation

$$RC\frac{dV_o}{dt} + V_o(t) = V_i(t), \tag{8.8}$$

where $V_i(t)$ is the sum of a noise term $V_n(t)$ and a signal term $V_s(t)$. Integration and discretisation of this expression gives the required algorithm:

$$V_{o,k} = V_{i,k} + \left(V_{o,k-1} - V_{i,k}\right)\exp(-\Delta t/RC). \tag{8.9}$$

Smith and Cohn-Sfetcu have pointed out that the beginning of the signal may be distorted if the starting value for algorithm is chosen badly. Suppose, for example, the

starting value $V_{o,0}$ is chosen to be equal to $V_{i,0}$. It can be seen from equation (8.9) that there will be little change in V_o for the first few steps. The output signal will thus lag behind the input for a time, distorting the start of the signal. This can be corrected in two ways. The first requires some prior knowledge of the signal $V_s(t)$. This is once again best demonstrated by an example. Let $V_s(t)$ have the form

$$V_s(t) = V_s(0)\exp(-t/\tau). \tag{8.10}$$

In the absence of noise, $V_o(t)$ has the asymptotic form

$$V_o(t) = V_o(0)\exp(-t/\tau); \tag{8.11}$$

from equation (8.8), $V_o(0)$ and $V_s(0)$ are thus related by the equation

$$V_o(0) = \frac{V_s(0)}{1 - RC/\tau}, \tag{8.12}$$

which is approximately equal to $(1 + RC/\tau)V_s(0)$ for $RC << \tau$. Choice of the starting value

$$V_{o,0} = (1 + RC/\tau)V_{i,0} \tag{8.13}$$

ought, therefore, to get rid of most of the distortion.

The second way to correct for this distortion is to apply the RC filter reversed in time (Smith and Cohn-Sfetcu 1975). In this scheme, the output signal is generated by using the algorithm

$$V_{o,k} = V_{i,k} + \left(V_{o,k+1} - V_{i,k}\right)\exp(-\Delta t/RC), \tag{8.14}$$

beginning at $k = N$ rather than $k = 0$. The choice of starting value is now much less critical, because exponentially decaying signals vary at a much slower rate at the end of a time interval than at the beginning. If the decay is allowed to saturate within the acquisition interval, $V_{o,N} = V_{i,N}$ may be selected as the starting value without significant error.

8.2.1. An outline of a new, iterative, nonstationary, distortionless filter.

All of the filtration schemes described so far still create some unavoidable amount of distortion of the signal. For example, as mentioned above, signals composed of multiple exponential decays suffer perturbations in the amplitudes of these decays. An estimate of the form of $V_s(t)$ can, however, permit the filtration of V_s without the introduction of any distortion. This is done by varying the value of the product RC with time so that it remains inversely proportional to $-dV_s/dt$. Substitution of this *ansatz* into equation (8.8) gives

$$-A\frac{V_o'(t)}{V_s'(t)}+V_o(t)=V_s(t)+V_n(t), \tag{8.15}$$

where A is the constant of proportionality. After integration this becomes

$$V_o(t)\exp\left[-\frac{\Delta V_s(t)}{A}\right]=V_o(0)-\frac{1}{A}\int_0^t dx\,[V_s(x)+V_n(x)]V_s'(x)\exp\left[-\frac{\Delta V_s(x)}{A}\right], \tag{8.16}$$

where

$$\Delta V_s(t)=V_s(t)-V_s(0). \tag{8.17}$$

If $V_s(0) + A$ is used as the starting value of V_o, equation (8.16) can be written

$$V_o(t)=V_s(t)+A+V_N(t), \tag{8.18}$$

where the noise term $V_N(t)$ is given by

$$V_N(t)=\int_0^t dx\,V_n(x)\frac{d}{dx}\exp\left[\frac{V_s(t)-V_s(x)}{A}\right]. \tag{8.19}$$

A similar filtering in reversed time gives

$$V_o(t)=V_s(t)-A+V_N(t), \tag{8.20}$$

Adding together the results of equations (8.18) and (8.20) produces an output which is an undistorted version of the original signal plus the noise term V_N.

A filtration scheme is no good unless it reduces the noise amplitude. Although the integral in equation (8.19) can be difficult to evaluate, it is not hard to show that, for certain types of signal V_s, the standard deviation σ_N of the filtered noise voltage V_N is less than σ_n, the standard deviation of the original noise voltage V_n. The first step is to adapt equation (8.19) to the true experimental situation, in which the signals are sampled at intervals Δt. The corresponding equation is

$$V_{N,k} = \Delta t \sum_{j=0}^{k-1} V_{n,j} \frac{d}{dx} \exp\left[\frac{V_s(k\Delta t) - V_s(x)}{A}\right]_{x=j\Delta t}. \qquad (8.21)$$

The variance σ_N^2 in the sampled noise output $V_{N,k}$ is therefore

$$\sigma_N^2 = \sigma_n^2 \Delta t^2 \sum_{j=0}^{k-1} \left\{ \frac{d}{dx} \exp\left[\frac{V_s(k\Delta t) - V_s(x)}{A}\right]_{x=j\Delta t} \right\}^2. \qquad (8.22)$$

For a monotonically decreasing signal V_s, the expression inside the braces is always positive. Therefore the following inequality holds:

$$\sigma_N < \sigma_n \Delta t \sum_{j=0}^{k-1} \frac{d}{dx} \exp\left[\frac{V_s(k\Delta t) - V_s(x)}{A}\right]_{x=j\Delta t}. \qquad (8.23)$$

Approximation of the sum by an integral yields the final expression

$$\frac{\sigma_N}{\sigma_n} < 1 - \exp\left[\frac{V_s(k\Delta t) - V_s(0)}{A}\right] < 1. \qquad (8.24)$$

In practice, the decrease in noise amplitude is usually much better than this inequality indicates. This is because, in the case that V_s decreases exponentially, the expression inside the braces of equation (8.22) is much smaller than 1 for most values of j. Its square is therefore negligible except for values of j very close to k.

Full application of this technique, which does not appear to have been published previously, would involve an iterative procedure in which the filtering operation forms one part and an estimation of dV_s/dt forms the other. This implementation has not yet been attempted.

8.3. Curve fitting techniques.

The reduction of data frequently requires the fitting of theoretical expressions to experimental signals. In the course of the present work it was necessary to fit two sorts of expression, namely, multi-exponential decays and quadratics. A computer program was written to perform this function. (The algorithm is discussed in Appendix B).

Few difficulties were experienced in the fitting of quadratic data; however, multi-exponential decays are much more intractable. The problem of inverting a sum of exponential decays to determine both the time constants and the amplitudes is known to be ill-conditioned (Smith *et al* 1976). Several approaches were pursued in the present study. The easiest situation to deal with occurs when only one exponential is of significant amplitude. (This was frequently found to be the case in decays of both discharge current and optical absorption.) The procedure in this case was to fit the baseline value, then to use a weighted-least-squares analysis to determine the amplitude and time constant. Uncertainties were minimised by careful attention to the size of the time interval within which the signal was acquired and stored. The optimum size of the interval has been discussed by Moore and Yalcin (1973) and Smith and Buckmaster (1975).

Signals that obviously contain more than one exponential were fitted by minimising the value of chi squared between the fitting function and the experimental data (see appendix B). Uncertainties in the parameters of the fitting function (ie, the baseline and the amplitudes and time constants of the exponentials) were calculated by first evaluating the matrix of second derivatives of chi squared with respect to the parameters; this was then inverted to produce the covariance matrix (Bevington 1969).

The biggest difficulty with this fitting approach is that the form of the function must be assumed beforehand. This is a problem if the number of exponentials of significant amplitude is unknown, because it is difficult to verify one's guess about this number from the minimum value of chi square. For example, a signal composed of two components plus noise may often be fitted quite well by a trial function which has three components; furthermore, the amplitudes and time constants of the trial function may not necessarily bear any relation to those of the signal.

An alternative method is being investigated, based on the transformation proposed by Gardner *et al* (1959). This transforms the sum of exponentials into a convolution. In principle, the simplest way to deconvolve this integral is to apply

standard Fourier techniques (eg Provencher 1976, Smith *et al* 1976); however, the presence of noise and the necessity of truncating the data complicate the practice. Techniques of iterative deconvolution (see, for example, Schafer *et al* 1981) may hold some promise. This research is continuing.