

Inelastic Scattering

(with an emphasis on the Bloch wave approach)



Notes to accompany the lectures delivered by L. J. Allen at the Summer School on Electron Microscopy: Fundamental Limits and New Science held at Cornell University, July 13-15, 2006.¹

© L. J. Allen and S. D. Findlay

June 29, 2006

¹Only to be used by students enrolled for the Summer School. The notes may not be transmitted or duplicated in any form.

Contents

1	Inelastic scattering: coupled-channels formulation of scattering theory	3
1.1	The Schrödinger equation	3
1.2	The electron probability current vector and cross section for inelastic scattering	6
1.3	Projected potential and paraxial approximations	7
1.4	Schrödinger equation with local absorptive potential	10
2	Solving the paraxial wave equation with absorption	12
2.1	The Bloch wave method	12
2.2	Boundary conditions	16
2.3	Bloch states and excitation amplitudes	20
2.4	Lattice-resolution contrast in scanning transmission electron microscopy in the Bloch wave model	22
2.5	Multislice solution of the paraxial equation with absorption	22
3	Application: inelastic scattering in scanning transmission electron microscopy (STEM)	25
3.1	Introduction	25
3.2	Inelastic scattering coefficients	26
3.3	Introductory examples of inelastic STEM imaging	29
3.4	Spreading of the probe and quantitative tests of cross-talk	30
3.5	The role of absorption	33
3.6	Nonlocality and “imaging at a distance”	35
3.7	Nonlocality and the importance of detector size	37
3.8	Simulations of STEM images	42
4	Overview	43
5	Appendix - sign conventions	44

Chapter 1

Inelastic scattering: coupled-channels formulation of scattering theory

1.1 The Schrödinger equation

Consider an electron incident on a specimen of condensed matter. The Schrödinger equation is [1–4]

$$\left[-\frac{\hbar^2}{8\pi^2m} \nabla^2 + H(\boldsymbol{\tau}) + H'(\mathbf{r}; \boldsymbol{\tau}) \right] \Psi(\mathbf{r}; \boldsymbol{\tau}) = E\Psi(\mathbf{r}; \boldsymbol{\tau}), \quad (1.1)$$

where \mathbf{r} is the coordinate of the incident electron and $\boldsymbol{\tau} \equiv \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$ are the coordinates of all the particles (nuclei or electrons) in the solid. The term $(-\hbar^2/8\pi^2m)\nabla^2$ is the kinetic energy operator and $H(\boldsymbol{\tau})$ is the Hamiltonian for all the particles in the specimen. The next term $H'(\mathbf{r}; \boldsymbol{\tau})$ describes the interaction of the incident electron with the particles in the specimen. On the right-hand side E is the total energy of the system such that

$$E = E_0 + \varepsilon_0, \quad (1.2)$$

where E_0 is the incident electron energy (in vacuum) and ε_0 is the initial energy of the specimen. We assume the wave function can be expanded in the form

$$\Psi(\mathbf{r}; \boldsymbol{\tau}) = \sum_n \psi_n(\mathbf{r}) a_n(\boldsymbol{\tau}). \quad (1.3)$$

The normalised wave function $a_n(\boldsymbol{\tau})$ represents the n 'th stationary state of the specimen (of energy ε_n) and satisfies the equation

$$H(\boldsymbol{\tau}) a_n(\boldsymbol{\tau}) = \varepsilon_n a_n(\boldsymbol{\tau}). \quad (1.4)$$

One of the states $a_n(\boldsymbol{\tau})$ is regarded as the initial state and denoted $a_0(\boldsymbol{\tau})$. Therefore $\psi_0(\mathbf{r})$ in equation (1.3) describes the elastic scattering. Furthermore $\psi_n(\mathbf{r})$ ($n \neq 0$) describes the inelastic scattering in which the state of the specimen is changed from $a_0(\boldsymbol{\tau})$ to $a_n(\boldsymbol{\tau})$. The energy of the electron in the state $\psi_0(\mathbf{r})$ is given by [cf. equation (1.2)]

$$E_0 = E - \varepsilon_0. \quad (1.5)$$

For the case of inelastic scattering, the energy associated with $\psi_n(\mathbf{r})$, i.e. after the inelastic scattering event, is

$$E_n = E - \varepsilon_n \equiv \frac{\hbar^2}{2m} k_n^2, \quad (1.6)$$

where k_n is the magnitude of the wave vector of the scattered electron. We adopt the convention that $k_n = 1/\lambda_n$, where λ_n is the associated wave length.¹ The energy loss of the incident fast electron after an inelastic scattering event which excites the specimen from the initial to the n 'th excited state is

$$E_{\text{loss}} = \varepsilon_n - \varepsilon_0 . \quad (1.7)$$

Now let us re-arrange equation (1.1) as follows:

$$\nabla^2 \Psi(\mathbf{r}; \boldsymbol{\tau}) + \frac{8\pi^2 m}{h^2} [E - H(\boldsymbol{\tau})] \Psi(\mathbf{r}; \boldsymbol{\tau}) = \frac{8\pi^2 m}{h^2} H'(\mathbf{r}; \boldsymbol{\tau}) \Psi(\mathbf{r}; \boldsymbol{\tau}) . \quad (1.8)$$

Inserting equation (1.3) into equation (1.8) and pre-multiplying by $a_n^*(\boldsymbol{\tau})$ we obtain

$$\begin{aligned} \sum_m \nabla^2 \psi_m(\mathbf{r}) a_n^*(\boldsymbol{\tau}) a_m(\boldsymbol{\tau}) + \frac{8\pi^2 m}{h^2} E \sum_m \psi_m(\mathbf{r}) a_n^*(\boldsymbol{\tau}) a_m(\boldsymbol{\tau}) \\ - \frac{8\pi^2 m}{h^2} \sum_m \psi_m(\mathbf{r}) a_n^*(\boldsymbol{\tau}) H(\boldsymbol{\tau}) a_m(\boldsymbol{\tau}) - \frac{8\pi^2 m}{h^2} \psi_n(\mathbf{r}) a_n^*(\boldsymbol{\tau}) H'(\mathbf{r}; \boldsymbol{\tau}) a_n(\boldsymbol{\tau}) \\ = \frac{8\pi^2 m}{h^2} \sum_{m \neq n} \psi_m(\mathbf{r}) a_n^*(\boldsymbol{\tau}) H'(\mathbf{r}; \boldsymbol{\tau}) a_m(\boldsymbol{\tau}) . \end{aligned} \quad (1.9)$$

Using equation (1.4) and integrating over $d\boldsymbol{\tau} \equiv d\mathbf{r}_1 \dots d\mathbf{r}_N$ we obtain

$$\begin{aligned} \nabla^2 \psi_n(\mathbf{r}) + \frac{8\pi^2 m}{h^2} (E - \varepsilon_n) \psi_n(\mathbf{r}) - \frac{8\pi^2 m}{h^2} \psi_n(\mathbf{r}) \int a_n^*(\boldsymbol{\tau}) H'(\mathbf{r}, \boldsymbol{\tau}) a_n(\boldsymbol{\tau}) d\boldsymbol{\tau} \\ = \frac{8\pi^2 m}{h} \sum_{m \neq n} \left[\int a_n^*(\boldsymbol{\tau}) H'(\mathbf{r}, \boldsymbol{\tau}) a_m(\boldsymbol{\tau}) d\boldsymbol{\tau} \right] \psi_m(\mathbf{r}) . \end{aligned} \quad (1.10)$$

Using equation (1.6) this may be rewritten as

$$\left[\nabla^2 + 4\pi^2 k_n^2 - \frac{8\pi^2 m}{h^2} H_{nn}(\mathbf{r}) \right] \psi_n(\mathbf{r}) = \frac{8\pi^2 m}{h^2} \sum_{m \neq n} H_{nm}(\mathbf{r}) \psi_m(\mathbf{r}) , \quad (1.11)$$

where

$$H_{nm}(\mathbf{r}) = \int a_n^*(\boldsymbol{\tau}) H'(\mathbf{r}; \boldsymbol{\tau}) a_m(\boldsymbol{\tau}) d\boldsymbol{\tau} . \quad (1.12)$$

Note that $H_{nm}(\mathbf{r}) = H_{mn}^*(\mathbf{r})$. In the special case where $n = 0$ we have

$$\left[\nabla^2 + 4\pi^2 k_0^2 - \frac{8\pi^2 m}{h^2} H_{00}(\mathbf{r}) \right] \psi_0(\mathbf{r}) = \frac{8\pi^2 m}{h^2} \sum_{m \neq 0} H_{0m}(\mathbf{r}) \psi_m(\mathbf{r}) . \quad (1.13)$$

If inelastic scattering is neglected this becomes

$$\left[\nabla^2 + 4\pi^2 k_0^2 - \frac{8\pi^2 m}{h^2} H_{00}(\mathbf{r}) \right] \psi_0(\mathbf{r}) = 0 . \quad (1.14)$$

Thus $H_{00}(\mathbf{r})$ is the potential for the elastic scattering of the incident electron.

¹Relativistic corrections are usually applied to wavelengths and the mass m .

We now make the *assumption* that, in equation (1.11), only terms with $m = 0$ give a significant contribution, i.e. only excitations from the ground state contribute significantly to the scattering (see Wang [5]). This means we assume $H_{n0}(\mathbf{r}) \gg H_{nm}(\mathbf{r})$ ($m \neq 0$ and $m \neq n$). We also *assume* that we can neglect H_{nn} ($\approx H_{00}$) when $n \neq 0$ in equation (1.11), which means that we neglect dynamical diffraction (channelling) effects for the inelastically scattered electrons.² Equation (1.11) then reduces to

$$(\nabla^2 + 4\pi^2 k_n^2) \psi_n(\mathbf{r}) = \frac{8\pi^2 m}{h^2} H_{n0}(\mathbf{r}) \psi_0(\mathbf{r}) \quad (n \neq 0). \quad (1.15)$$

The solution of equation (1.15) can be written in the standard form (outgoing wave solution):

$$\psi_n(\mathbf{r}) = -\frac{2\pi m}{h^2} \int \frac{\exp(2\pi i k_n |\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|} H_{n0}(\mathbf{r}') \psi_0(\mathbf{r}') d\mathbf{r}'. \quad (1.16)$$

Substituting this into equation (1.13) we obtain

$$\begin{aligned} & \left[\nabla^2 + 4\pi^2 k_0^2 - \frac{8\pi^2 m}{h^2} H_{00}(\mathbf{r}) \right] \psi_0(\mathbf{r}) \\ & + \frac{8\pi^2 m}{h^2} \sum_{m \neq 0} H_{0m}(\mathbf{r}) \frac{2\pi m}{h^2} \int \frac{\exp(2\pi i k_m |\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|} H_{m0}(\mathbf{r}') \psi_0(\mathbf{r}') d\mathbf{r}' = 0 \end{aligned} \quad (1.17)$$

or, equivalently,

$$\left[\nabla^2 + 4\pi^2 k_0^2 - \frac{8\pi^2 m}{h^2} H_{00}(\mathbf{r}) \right] \psi_0(\mathbf{r}) - \frac{8\pi^2 m}{h^2} \int A(\mathbf{r}, \mathbf{r}') \psi_0(\mathbf{r}') d\mathbf{r}' = 0, \quad (1.18)$$

where

$$A(\mathbf{r}, \mathbf{r}') = -\frac{2\pi m}{h^2} \sum_{m \neq 0} H_{0m}(\mathbf{r}) H_{m0}(\mathbf{r}') \frac{\exp(2\pi i k_m |\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|}. \quad (1.19)$$

Equation (1.18) was derived by considering the effect of inelastic scattering on the elastic scattering wave function. The elastic wave function obtained as a solution to equation (1.18) is modified by the presence of the inelastic potential. A treatment of inelastic scattering which does not neglect the terms $H_{nm}(\mathbf{r})$ ($m \neq 0$ and $m \neq n$) [*cf.* assumptions above equation (1.15)] is discussed in Ref. [6]. More exact but more complicated forms for $A(\mathbf{r}, \mathbf{r}')$ can be obtained [3,6–8]. Equation (1.18) is just a Schrödinger equation containing a nonlocal potential term.

Now $A(\mathbf{r}, \mathbf{r}')$ is not Hermitian but can be written as the sum of two Hermitian terms as follows:

$$A(\mathbf{r}, \mathbf{r}') = \frac{1}{2} \{ A(\mathbf{r}, \mathbf{r}') + A^*(\mathbf{r}', \mathbf{r}) \} + \frac{1}{2i} \{ i [A(\mathbf{r}, \mathbf{r}') - A^*(\mathbf{r}', \mathbf{r})] \}. \quad (1.20)$$

The first Hermitian term corresponds to virtual inelastic scattering and is known to be small [9,10]. Therefore we neglect this term, obtaining

$$A(\mathbf{r}, \mathbf{r}') \approx -\frac{i}{2} \{ i [A(\mathbf{r}, \mathbf{r}') - A^*(\mathbf{r}', \mathbf{r})] \} \equiv -\frac{i}{2} W(\mathbf{r}, \mathbf{r}'). \quad (1.21)$$

²This is never a good approximation in detail, but is greatly simplifying and proves adequate to describe the total absorption from the elastic channel and to estimate the inelastic scattering into a detector with a moderately large acceptance angle.

With this approximation we can rewrite equation (1.18) in the form

$$\left[\nabla^2 + 4\pi^2 k_0^2 - \frac{8\pi^2 m}{h^2} H_{00}(\mathbf{r}) \right] \psi_0(\mathbf{r}) + \frac{4i\pi^2 m}{h^2} \int W(\mathbf{r}, \mathbf{r}') \psi_0(\mathbf{r}') d\mathbf{r}' = 0. \quad (1.22)$$

The set of coupled second order differential equations (1.11) has been reduced to a single integro-differential equation, where the term containing the nonlocal absorptive potential $W(\mathbf{r}, \mathbf{r}')$ gives the effect of inelastic scattering on the elastic scattering channel.

Using equation (1.19) and the fact that

$$\frac{\exp(2\pi i k_m r) - \exp(-2\pi i k_m r)}{r} = \frac{i}{k_m} \int \delta(k_m - K') \exp(2\pi i \mathbf{K}' \cdot \mathbf{r}) d\mathbf{K}', \quad (1.23)$$

we may write $W(\mathbf{r}, \mathbf{r}')$ as

$$W(\mathbf{r}, \mathbf{r}') = \frac{2\pi m}{h^2} \sum_{m \neq 0} H_{0m}(\mathbf{r}) H_{m0}(\mathbf{r}') k_m \int e^{2\pi i \mathbf{K}' \cdot (\mathbf{r} - \mathbf{r}')} \delta(k_m - K') d\Omega_{\mathbf{K}'} dK', \quad (1.24)$$

where we have used $d\mathbf{K}' = K'^2 d\Omega_{\mathbf{K}'} dK'$.

1.2 The electron probability current vector and cross section for inelastic scattering

The electron current density vector $\mathbf{j}_0(\mathbf{r})$ is defined as [11]

$$\mathbf{j}_0(\mathbf{r}) = \frac{h}{4\pi i m} [\psi_0^*(\mathbf{r}) \nabla \psi_0(\mathbf{r}) - \psi_0(\mathbf{r}) \nabla \psi_0^*(\mathbf{r})]. \quad (1.25)$$

In the theory of scattering in electron microscopy it is usual to normalize the (dimensionless) wave function over the volume V of the specimen as $\frac{1}{V} \int_V |\psi_0(\mathbf{r})|^2 d\mathbf{r}$. Then $\mathbf{j}_0(\mathbf{r})$ has units ms^{-1} . Noting that

$$\psi_0^*(\mathbf{r}) \nabla^2 \psi_0(\mathbf{r}) - \psi_0(\mathbf{r}) \nabla^2 \psi_0^*(\mathbf{r}) = \nabla \cdot [\psi_0^*(\mathbf{r}) \nabla \psi_0(\mathbf{r}) - \psi_0(\mathbf{r}) \nabla \psi_0^*(\mathbf{r})], \quad (1.26)$$

it follows from equation (1.22) that

$$\int_S \mathbf{j}_0(\mathbf{r}) \cdot d\mathbf{s} = \int_V \nabla \cdot \mathbf{j}_0(\mathbf{r}) d\mathbf{r} = -\frac{2\pi}{h} \int_V \int_V \psi_0^*(\mathbf{r}) W(\mathbf{r}, \mathbf{r}') \psi_0(\mathbf{r}') d\mathbf{r} d\mathbf{r}', \quad (1.27)$$

where S denotes the surface of the specimen. The quantity $(2\pi/h)W(\mathbf{r}, \mathbf{r}')$ expresses the spatial distribution of the nonlocal ‘‘absorption power’’, with the amount of absorption depending on the value of $\psi_0(\mathbf{r})$ at the points \mathbf{r} and \mathbf{r}' .

Conservation of electrons can be stated in the form

$$\sum_{n \neq 0} \int_S \mathbf{j}_n(\mathbf{r}) \cdot d\mathbf{s} = - \int_S \mathbf{j}_0(\mathbf{r}) \cdot d\mathbf{s}, \quad (1.28)$$

where $\mathbf{j}_n(\mathbf{r})$ is the electron probability current density vector corresponding to an event that leaves the specimen in the state n . In other words, the sum of the electron currents flowing out from the

specimen by inelastic scattering is equal to the *net* flux of electrons from the elastic scattered part $\psi_0(\mathbf{r})$ *into* (minus sign) the specimen. The electrons that contribute to the inelastic scattering are effectively absorbed in the specimen as far as the elastic scattering part $\psi_0(\mathbf{r})$ is concerned.

The cross section for absorptive scattering, or “absorption” from the elastic scattered electron flux, is given by the number of electrons per unit volume multiplied by the probability that an inelastically scattered electron will cross the surface of the specimen, per incident flux per unit area, i.e.

$$\sigma = \left[\rho \sum_{n \neq 0} \int_S \mathbf{j}_n(\mathbf{r}) \cdot d\mathbf{s} \right] / (\rho v), \quad (1.29)$$

where ρ is the number of electrons per unit volume and $v = \hbar k_0 / m$ is the incident electron velocity.

The integral over the specimen surface S in equation (1.29) summed over all $n \neq 0$ is the probability that an inelastically scattered electron will exit the specimen surface. Using equation (1.28), we write the inelastic cross section in terms of the elastic electron probability current density vector as

$$\sigma = -\frac{1}{v} \int_S \mathbf{j}_0(\mathbf{r}) \cdot d\mathbf{s}. \quad (1.30)$$

Note that $\int \mathbf{j}_0(\mathbf{r}) \cdot d\mathbf{s}$ has units of $\text{m}^3 \text{s}^{-1}$ and the cross section therefore has units of m^2 . Consequently, using equation (1.27), the general expression for the cross section for inelastic scattering is

$$\sigma = \frac{2\pi}{\hbar v} \int_V \int_V \psi_0^*(\mathbf{r}) W(\mathbf{r}, \mathbf{r}') \psi_0(\mathbf{r}') d\mathbf{r} d\mathbf{r}'. \quad (1.31)$$

1.3 Projected potential and paraxial approximations

Let us *assume* that we have a beam of electrons incident on the specimen where the wave vector \mathbf{k}_0 is along the z -axis. The entrance surface of the slab-like specimen coincides with the xy -plane. This geometry may readily be generalized [12, 13]. Assume that, on interacting with the specimen, the electrons are not greatly deviated from the z -axis. It is then convenient to rewrite the wave function as a modulated plane wave

$$\psi_0(\mathbf{r}) \equiv \psi_0(\mathbf{r}_\perp, z) = e^{2\pi i K z} \phi_0(\mathbf{r}_\perp, z), \quad (1.32)$$

where \mathbf{r}_\perp denotes a vector parallel to the xy -plane and

$$K^2 = k_0^2 + \frac{2m}{\hbar^2} V_0, \quad \text{with} \quad V_0 = -\frac{1}{V} \int_V H_{00}(\mathbf{r}) d\mathbf{r}. \quad (1.33)$$

Here V_0 is (minus) the mean value of the elastic scattering potential $H_{00}(\mathbf{r})$ in the crystal. The value of K so defined describes the magnitude of the wave vector as corrected for refraction in the specimen. The minus sign is consistent with introducing the potential

$$V(\mathbf{r}) \equiv -H_{00}(\mathbf{r}) - V_0 \quad (1.34)$$

to agree with a convention in the theory of scattering in electron microscopy that an attractive potential is positive. The mean value of $V(\mathbf{r})$ so introduced is zero, with what would otherwise be the

mean value (V_0) rather being associated with a modification of the incident wave number, as defined in equation (1.33).

Using equations (1.32) to (1.34) in the Schrödinger equation (1.22) we can show that the modulating function $\phi_0(\mathbf{r})$ satisfies

$$\frac{\partial^2 \phi_0}{\partial z^2} + 4\pi i K \frac{\partial \phi_0}{\partial z} + \nabla_{\perp}^2 \phi_0 + \frac{8\pi^2 m}{h^2} V(\mathbf{r}) \phi_0 + \frac{4\pi^2 i m}{h^2} \int \int e^{-2\pi i K z} W(\mathbf{r}, \mathbf{r}') e^{2\pi i K z'} \phi_0(\mathbf{r}'_{\perp}, z') d\mathbf{r}'_{\perp} dz' = 0. \quad (1.35)$$

We now *assume* that the second derivative with respect to z in the above equation is very small in comparison to the other terms, i.e. the modulating function ϕ_0 is only slowly varying in the z -direction [which would not have been true of ψ_0 in equation (1.32)]. We then have

$$4\pi i K \frac{\partial \phi_0}{\partial z} + \nabla_{\perp}^2 \phi_0 + \frac{8\pi^2 m}{h^2} V(\mathbf{r}) \phi_0 + \frac{4\pi^2 i m}{h^2} \int \int e^{-2\pi i K z} W(\mathbf{r}, \mathbf{r}') e^{2\pi i K z'} \phi_0(\mathbf{r}'_{\perp}, z') d\mathbf{r}'_{\perp} dz' = 0. \quad (1.36)$$

Discarding the second derivative in z is known as the paraxial approximation (in particular back scattering is no longer included). In the absence of potential terms, the above equation becomes the governing equation for paraxial (i.e. small angle) propagation in free space. Reducing a second order partial differential equation in z to a first order partial differential equation in z means only one boundary condition is necessary. The wave function should be continuous but $\partial\phi_0/\partial z$ need not be continuous.

The potentials $V(\mathbf{r})$ and $W(\mathbf{r}, \mathbf{r}')$ in equation (1.36) vary in three dimensions. However, given that $\phi_0(\mathbf{r}_{\perp}, z)$ varies slowly with z , the electron is insensitive to the variation of the potential along the z -direction. Therefore, as shown in figure 1.1, we will divide the slab-like sample into slab-like slices such that a projected potential approximation may be separately applied within each slice, i.e. the potential in each slice will be replaced by its average along the z -direction. Generally such slices must be chosen such that multiple scattering occurs on a length scale larger than the z -dimension of the slice. In a crystalline sample the repeat distance is usually a natural and convenient distance over which to project.

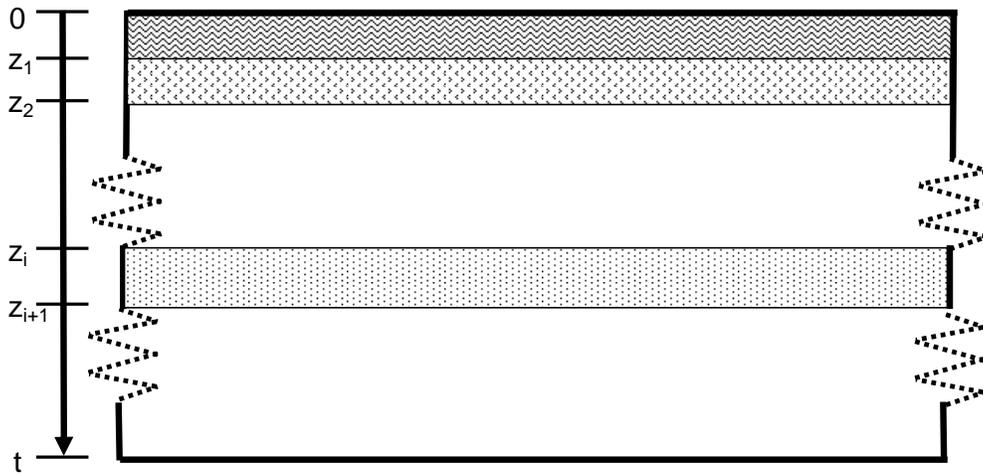


Figure 1.1: Schematic of the conceptual division of the slab-like sample into slab-like slices such that the projected potential approximation may be separately applied within each slice.

So our approach is to solve equation (1.36) sequentially in each slice where we make a projected potential approximation, paying attention to the boundary conditions at the interfaces between slices. (Having made the paraxial approximation and discarded back-scattering through the discarding of the second derivative, we are fully justified in considering a single slice at a time.) Let us consider one such slice and *assume* that inelastic scattering can be associated with individual atoms in that slice and that the contributions from each atom contribute incoherently with respect to each other. Labelling the atoms in the i 'th slice by α_i we can write equation (1.36) as

$$4\pi i K \frac{\partial \phi_0}{\partial z} + \nabla_{\perp}^2 \phi_0 + \frac{8\pi^2 m}{h^2} V(\mathbf{r}) \phi_0 + \frac{4\pi^2 i m}{h^2} \int \int e^{-2\pi i K z} \sum_{\alpha_i} W_{\alpha_i}(\mathbf{r}, \mathbf{r}') e^{2\pi i K z'} \phi_0(\mathbf{r}'_{\perp}, z') d\mathbf{r}'_{\perp} dz' = 0. \quad (1.37)$$

Now $W_{\alpha_i}(\mathbf{r}, \mathbf{r}')$ will only have non-zero contributions when both z and z' are in the vicinity of atom α_i . The extent of the "vicinity" will depend on the range of the interaction, but we shall assume that it is of a distance smaller than that on which $\phi_0(\mathbf{r}'_{\perp}, z')$ varies with z' . So to a good approximation $W_{\alpha_i}(\mathbf{r}, \mathbf{r}') \phi_0(\mathbf{r}'_{\perp}, z') \approx W_{\alpha_i}(\mathbf{r}, \mathbf{r}') \phi_0(\mathbf{r}'_{\perp}, z)$, and we may rewrite equation (1.37) as

$$4\pi i K \frac{\partial \phi_0}{\partial z} + \nabla_{\perp}^2 \phi_0 + \frac{8\pi^2 m}{h^2} V(\mathbf{r}) \phi_0 + \frac{4\pi^2 i m}{h^2} \int \phi_0(\mathbf{r}'_{\perp}, z) \int e^{-2\pi i K z} \sum_{\alpha_i} W_{\alpha_i}(\mathbf{r}, \mathbf{r}') e^{2\pi i K z'} dz' d\mathbf{r}'_{\perp} = 0. \quad (1.38)$$

Let us now make the projected potential approximation within the i 'th slice. Within this slice we replace the elastic potential $V(\mathbf{r})$ in equation (1.38) with

$$V_i(\mathbf{r}_{\perp}) = \frac{1}{d_i} \int_{z_i}^{z_{i+1}} V(\mathbf{r}_{\perp}, z) dz, \quad (1.39)$$

where $d_i = z_{i+1} - z_i$. Similarly, for the absorptive term we define

$$W_i(\mathbf{r}_{\perp}, \mathbf{r}'_{\perp}) = \frac{1}{d_i} \int_{z_i}^{z_{i+1}} \int_{z_i}^{z_{i+1}} e^{-2\pi i K z} \sum_{\alpha_i} W_{\alpha_i}(\mathbf{r}, \mathbf{r}') e^{2\pi i K z'} dz' dz. \quad (1.40)$$

Hence within the i 'th slice the paraxial Schrödinger equation is

$$4\pi i K \frac{\partial \phi_0}{\partial z} + \nabla_{\perp}^2 \phi_0 + \frac{8\pi^2 m}{h^2} V_i(\mathbf{r}_{\perp}) \phi_0 + \frac{4\pi^2 i m}{h^2} \int W_i(\mathbf{r}_{\perp}, \mathbf{r}'_{\perp}) \phi_0(\mathbf{r}'_{\perp}, z) d\mathbf{r}'_{\perp} = 0. \quad (1.41)$$

Let us now consider the use of the modulated plane wave ansatz of equation (1.32) in the cross section expression for inelastic scattering in equation (1.31). It is clear that applying these to the i 'th slice we may write

$$\sigma_i = \frac{2\pi}{h\nu} \int_A \int_{z_i}^{z_{i+1}} \int_A \int_{z_i}^{z_{i+1}} \phi_0^*(\mathbf{r}_{\perp}, z) e^{-2\pi i K z} \sum_{\alpha_i} W_{\alpha_i}(\mathbf{r}, \mathbf{r}') e^{2\pi i K z'} \phi_0(\mathbf{r}'_{\perp}, z') d\mathbf{r}_{\perp} dz d\mathbf{r}'_{\perp} dz'. \quad (1.42)$$

Assuming that ϕ_0 varies slowly with z over the distance d_i between z_i and z_{i+1} , we may make the approximation

$$\begin{aligned} \sigma_i &\approx \frac{2\pi}{h\nu} d_i \int_A \int_A \phi_0^*(\mathbf{r}_{\perp}, z_i) \phi_0(\mathbf{r}'_{\perp}, z_i) \left[\frac{1}{d_i} \int_{z_i}^{z_{i+1}} \int_{z_i}^{z_{i+1}} e^{-2\pi i K z} \sum_{\alpha_i} W_{\alpha_i}(\mathbf{r}, \mathbf{r}') e^{2\pi i K z'} dz dz' \right] d\mathbf{r}_{\perp} d\mathbf{r}'_{\perp} \\ &= \frac{2\pi}{h\nu} d_i \int_A \int_A \phi_0^*(\mathbf{r}_{\perp}, z_i) W_i(\mathbf{r}_{\perp}, \mathbf{r}'_{\perp}) \phi_0(\mathbf{r}'_{\perp}, z_i) d\mathbf{r}_{\perp} d\mathbf{r}'_{\perp}. \end{aligned} \quad (1.43)$$

The total cross section for the specimen is given by the sum over slices as

$$\sigma = \sum_i \sigma_i = \sum_i d_i \frac{2\pi}{h\nu} \int_A \int_A \phi_0^*(\mathbf{r}_\perp, z_i) W_i(\mathbf{r}_\perp, \mathbf{r}'_\perp) \phi_0(\mathbf{r}'_\perp, z_i) d\mathbf{r}_\perp d\mathbf{r}'_\perp . \quad (1.44)$$

If the sample is crystalline and d_i is taken to be the repeat distance d , then $W_i(\mathbf{r}_\perp, \mathbf{r}'_\perp)$ is independent of i and we may write

$$\sigma = \frac{2\pi}{h\nu} \int_0^t \int_A \int_A \phi_0^*(\mathbf{r}_\perp, z) W(\mathbf{r}_\perp, \mathbf{r}'_\perp) \phi_0(\mathbf{r}'_\perp, z) d\mathbf{r}_\perp d\mathbf{r}'_\perp dz . \quad (1.45)$$

where

$$W(\mathbf{r}_\perp, \mathbf{r}'_\perp) = \frac{1}{t} \int_0^t \int_0^t e^{-2\pi i K z} \sum_\alpha W_\alpha(\mathbf{r}, \mathbf{r}') e^{2\pi i K z'} dz' dz . \quad (1.46)$$

and the sum over α now runs over all the atoms in the crystal.

1.4 Schrödinger equation with local absorptive potential

Now using equation (1.24) for each atom in the crystal in equation (1.46) we can write the projected nonlocal potential $W(\mathbf{r}_\perp, \mathbf{r}'_\perp)$ in the form

$$\begin{aligned} W(\mathbf{r}_\perp, \mathbf{r}'_\perp) &= \frac{2\pi m}{h^2 t} \sum_\alpha \int_0^t \int_0^t e^{-2\pi i K z} e^{2\pi i K z'} \sum_{n \neq 0} H_{\alpha, 0n}(\mathbf{r}) H_{\alpha, n0}(\mathbf{r}') k_n \\ &\quad \times \int e^{2\pi i \mathbf{K}' \cdot (\mathbf{r} - \mathbf{r}')} \delta(k_n - K') d\Omega_{K'} dK' dz' dz \\ &= \frac{2\pi m}{h^2 t} \sum_\alpha \sum_{n \neq 0} k_n \left[\int_0^t H_{\alpha, 0n}(\mathbf{r}) e^{-2\pi i (K - K') z} dz \right] \left[\int_0^t H_{\alpha, n0}(\mathbf{r}') e^{2\pi i (K - K') z'} dz' \right] \\ &\quad \times \int e^{2\pi i \mathbf{K}'_\perp \cdot (\mathbf{r}_\perp - \mathbf{r}'_\perp)} \delta(k_n - K') d\Omega_{K'} dK' \\ &\equiv \frac{2\pi m}{h^2 t} \sum_\alpha \sum_{n \neq 0} k_n H_{\alpha, n0}^*(\mathbf{r}_\perp) H_{\alpha, n0}(\mathbf{r}'_\perp) \int e^{2\pi i \mathbf{K}'_\perp \cdot (\mathbf{r}_\perp - \mathbf{r}'_\perp)} \delta(k_n - K') d\Omega_{K'} dK' , \quad (1.47) \end{aligned}$$

where we have defined the projected transition matrix element via

$$H_{\alpha, n0}(\mathbf{r}_\perp) \equiv \int_0^t H_{\alpha, n0}(\mathbf{r}) e^{2\pi i (K - K') z} dz . \quad (1.48)$$

The projected transition matrix element has been obtained in the third line of equation (1.47) using the approximation that $K'_z \approx K'$ (a good approximation for scattering which is predominantly in the forward direction). Making the flat Ewald sphere approximation this can be written as

$$W(\mathbf{r}_\perp, \mathbf{r}'_\perp) \approx \frac{2\pi m}{h^2 t} \sum_{n \neq 0} \frac{1}{k_n} \sum_\alpha H_{\alpha, n0}^*(\mathbf{r}_\perp) H_{\alpha, n0}(\mathbf{r}'_\perp) \int e^{2\pi i \mathbf{K}'_\perp \cdot (\mathbf{r}_\perp - \mathbf{r}'_\perp)} d^2 \mathbf{K}'_\perp . \quad (1.49)$$

If the range of integration in the plane is large enough (large aperture) then the integral reduces to $\delta(\mathbf{r}_\perp - \mathbf{r}'_\perp)$ (we will return to this point in section 3.7) and we may make the approximation

$$W(\mathbf{r}_\perp, \mathbf{r}'_\perp) \approx 2V'(\mathbf{r}_\perp) \delta(\mathbf{r}_\perp - \mathbf{r}'_\perp) \equiv \frac{2\pi m}{h^2 t} \sum_{n \neq 0} \frac{1}{k_n} \sum_\alpha H_{\alpha, n0}^*(\mathbf{r}_\perp) H_{\alpha, n0}(\mathbf{r}_\perp) \delta(\mathbf{r}_\perp - \mathbf{r}'_\perp) . \quad (1.50)$$

The inelastic term in the paraxial Schrödinger equation must account for all inelastic scattering. Thus the effective aperture size is a maximum and we may take the approximation in equation (1.50) as given. The paraxial equation (1.41) then reduces to

$$4\pi iK \frac{\partial \phi_0}{\partial z} + \nabla_{\perp}^2 \phi_0 + \frac{8\pi^2 m}{h^2} V(\mathbf{r}_{\perp}) \phi_0 + \frac{8i\pi^2 m}{h^2} V'(\mathbf{r}_{\perp}) \phi_0 = 0. \quad (1.51)$$

Letting

$$U(\mathbf{r}_{\perp}) = \frac{2m}{h^2} V(\mathbf{r}_{\perp}) \quad \text{and} \quad U'(\mathbf{r}_{\perp}) = \frac{2m}{h^2} V'(\mathbf{r}_{\perp}), \quad (1.52)$$

we can rewrite Eq. (1.51) in the form

$$4\pi iK \frac{\partial \phi_0}{\partial z} + \nabla_{\perp}^2 \phi_0 + 4\pi^2 [U(\mathbf{r}_{\perp}) + iU'(\mathbf{r}_{\perp})] \phi_0 = 0. \quad (1.53)$$

Using equation (1.50), equation (1.45) for the cross section in a crystal reduces to

$$\sigma = \frac{4\pi}{h\nu} \int_0^t \int_A |\phi_0(\mathbf{r}_{\perp}, z)|^2 V'(\mathbf{r}_{\perp}) d\mathbf{r}_{\perp} dz. \quad (1.54)$$

Chapter 2

Solving the paraxial wave equation with absorption

2.1 The Bloch wave method

We will now discuss a Bloch wave approach to the solution of the paraxial wave equation (1.53), which we restate here for convenience:

$$4\pi iK \frac{\partial \phi_0}{\partial z} + \nabla_{\perp}^2 \phi_0 + 4\pi^2 [U(\mathbf{r}_{\perp}) + iU'(\mathbf{r}_{\perp})] \phi_0 = 0. \quad (2.1)$$

The Bloch wave method, which proceeds in reciprocal space, is suited to samples possessing a periodicity condition of the form

$$U(\mathbf{r}_{\perp}) + iU'(\mathbf{r}_{\perp}) = U(\mathbf{r}_{\perp} + \mathbf{R}) + iU'(\mathbf{r}_{\perp} + \mathbf{R}), \quad (2.2)$$

where \mathbf{R} is any lattice translation vector, i.e. to crystalline specimens. This periodicity permits the Fourier series expansions

$$U(\mathbf{r}_{\perp}) + iU'(\mathbf{r}_{\perp}) = \sum_{\mathbf{g} \neq 0} U_{\mathbf{g}} e^{2\pi i \mathbf{g} \cdot \mathbf{r}_{\perp}} + i \sum_{\mathbf{g}} U'_{\mathbf{g}} e^{2\pi i \mathbf{g} \cdot \mathbf{r}_{\perp}} \equiv \sum_{\mathbf{g} \neq 0} W_{\mathbf{g}} e^{2\pi i \mathbf{g} \cdot \mathbf{r}_{\perp}} + iU'_0, \quad (2.3)$$

where we have defined $W_{\mathbf{g}} = U_{\mathbf{g}} + iU'_{\mathbf{g}}$. Consistent with this, the eigenstates of the paraxial wave equation, equation (2.1), are assumed to be Bloch states of the form¹

$$\phi_0^i(\mathbf{r}) = e^{2\pi i \lambda^i z} \sum_{\mathbf{g}} C_{\mathbf{g}}^i e^{2\pi i \mathbf{g} \cdot \mathbf{r}_{\perp}}. \quad (2.4)$$

Each Bloch state $\phi_0^i(\mathbf{r})$ is characterized by the complex eigenvalue $\lambda^i = \gamma^i + i\eta^i$ [12, 14]. The γ^i are the *anpassung* and the η^i the absorption coefficients. Since thermal scattering provides the dominant contribution to overall absorption, η^i can to a good approximation be calculated for TDS only, using an Einstein model [12, 15].

¹The form presented assumes the net current to be perpendicular to the surface. An incident plane wave with component \mathbf{K}_{\perp} in the plane of the crystal surface can readily be handled: a phase factor of $\exp(2\pi i \mathbf{K}_{\perp} \cdot \mathbf{r}_{\perp})$ is introduced, and the λ^i and $C_{\mathbf{g}}^i$ become functions of \mathbf{K}_{\perp} . We opt for the simpler notation for the present purpose. An example of the generalization may be found in Ref. [12].

Substituting equations (2.3) and (2.4) into equation (2.1) gives

$$\begin{aligned}
-8\pi^2 K\lambda^i e^{2\pi i\lambda^i z} \sum_{\mathbf{g}} C_{\mathbf{g}}^i e^{2\pi i\mathbf{g}\cdot\mathbf{r}_{\perp}} + e^{2\pi i\lambda^i z} \sum_{\mathbf{g}} C_{\mathbf{g}}^i \left[-4\pi^2(\mathbf{g}^2 - iU'_0) \right] e^{2\pi i\mathbf{g}\cdot\mathbf{r}_{\perp}} \\
+ 4\pi^2 \sum_{\mathbf{h}\neq\mathbf{0}} W_{\mathbf{h}} e^{2\pi i\mathbf{h}\cdot\mathbf{r}_{\perp}} e^{2\pi i\lambda^i z} \sum_{\mathbf{g}} C_{\mathbf{g}}^i e^{2\pi i\mathbf{g}\cdot\mathbf{r}_{\perp}} = 0. \quad (2.5)
\end{aligned}$$

Removing the common factor of $e^{2\pi i\lambda^i z}$, we may multiply by $e^{-2\pi i\mathbf{f}\cdot\mathbf{r}_{\perp}}$ and integrate over \mathbf{r}_{\perp} . This results in

$$\begin{aligned}
-8\pi^2 K\lambda^i \sum_{\mathbf{g}} C_{\mathbf{g}}^i \delta_{\mathbf{g}-\mathbf{f}} + \sum_{\mathbf{g}} C_{\mathbf{g}}^i \left[-4\pi^2(\mathbf{g}^2 - iU'_0) \right] \delta_{\mathbf{g}-\mathbf{f}} + 4\pi^2 \sum_{\mathbf{h}\neq\mathbf{0},\mathbf{g}} W_{\mathbf{h}} C_{\mathbf{g}}^i \delta_{\mathbf{h}+\mathbf{g}-\mathbf{f}} = 0, \\
\left[-4\pi^2(\mathbf{f}^2 - iU'_0) - 8\pi^2 K\lambda^i \right] C_{\mathbf{f}}^i + 4\pi^2 \sum_{\mathbf{h}\neq\mathbf{0}} W_{\mathbf{h}} C_{\mathbf{f}-\mathbf{h}}^i = 0, \\
\left[-4\pi^2(\mathbf{f}^2 - iU'_0) - 8\pi^2 K\lambda^i \right] C_{\mathbf{f}}^i + 4\pi^2 \sum_{\mathbf{h}\neq\mathbf{f}} W_{\mathbf{f}-\mathbf{h}} C_{\mathbf{h}}^i = 0, \\
-(\mathbf{g}^2 - iU'_0) C_{\mathbf{g}}^i + \sum_{\mathbf{h}\neq\mathbf{g}} W_{\mathbf{g}-\mathbf{h}} C_{\mathbf{h}}^i = 2K\lambda^i C_{\mathbf{g}}^i. \quad (2.6)
\end{aligned}$$

The final result constitutes a set of simultaneous equations for the λ^i and $C_{\mathbf{g}}^i$ values. They are typically referred to as the dynamical equations, and may be recast into matrix form [14]

$$\mathcal{A}\mathcal{C} = \mathcal{C} \left[2K\lambda^i \right]_D. \quad (2.7)$$

Writing simultaneous equations in matrix form allows much freedom as to the ordering of rows and columns in the matrix. Allen *et al.* [16–18] write the structure (Bethe) matrix \mathcal{A} as

$$\mathcal{A} = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & -\mathbf{h}^2 + iU'_0 & W_{\mathbf{h}-\mathbf{g}} & W_{\mathbf{h}} & W_{\mathbf{h}+\mathbf{g}} & W_{2\mathbf{h}} & \dots \\ \dots & W_{\mathbf{g}-\mathbf{h}} & -\mathbf{g}^2 + iU'_0 & W_{\mathbf{g}} & W_{2\mathbf{g}} & W_{\mathbf{g}+\mathbf{h}} & \dots \\ \dots & W_{-\mathbf{h}} & W_{-\mathbf{g}} & iU'_0 & W_{\mathbf{g}} & W_{\mathbf{h}} & \dots \\ \dots & W_{-\mathbf{g}-\mathbf{h}} & W_{-2\mathbf{g}} & W_{-\mathbf{g}} & -\mathbf{g}^2 + iU'_0 & W_{-\mathbf{g}+\mathbf{h}} & \dots \\ \dots & W_{-2\mathbf{h}} & W_{-\mathbf{h}-\mathbf{g}} & W_{-\mathbf{h}} & W_{-\mathbf{h}+\mathbf{g}} & -\mathbf{h}^2 + iU'_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \quad (2.8)$$

where \mathbf{g} and \mathbf{h} are reciprocal lattice vectors, and we make the assumption that for every reciprocal lattice vector \mathbf{g} , the vector $-\mathbf{g}$ is included in the representation of \mathcal{A} .

The matrix \mathcal{C} has as columns the eigenvectors of \mathcal{A} and can be explicitly written as

$$\mathcal{C} = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{\mathbf{h}}^1 & C_{\mathbf{h}}^2 & \dots & C_{\mathbf{h}}^i & \dots \\ C_{\mathbf{g}}^1 & C_{\mathbf{g}}^2 & \dots & C_{\mathbf{g}}^i & \dots \\ C_{\mathbf{0}}^1 & C_{\mathbf{0}}^2 & \dots & C_{\mathbf{0}}^i & \dots \\ C_{-\mathbf{g}}^1 & C_{-\mathbf{g}}^2 & \dots & C_{-\mathbf{g}}^i & \dots \\ C_{-\mathbf{h}}^1 & C_{-\mathbf{h}}^2 & \dots & C_{-\mathbf{h}}^i & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \quad (2.9)$$

Lastly, in equation (2.7), $[2K\lambda^i]_D$ is a diagonal matrix containing the (complex) eigenvalues of \mathcal{A} .

Since the potentials $U(\mathbf{r})$ and $U'(\mathbf{r})$ given in equation (2.3) are real, $U_{\mathbf{g}} = U_{-\mathbf{g}}^*$ and $U'_{\mathbf{g}} = U'_{-\mathbf{g}}^*$. However, in the presence of absorption, $(U_{\mathbf{g}} + iU'_{\mathbf{g}}) \neq (U_{-\mathbf{g}} + iU'_{-\mathbf{g}})^*$ and \mathcal{A} is not Hermitian. Thus the matrix of eigenvectors given by equation (2.9) is not unitary.

Let us assume an N -beam approximation (\mathcal{A} becomes an $N \times N$ matrix). Furthermore let us relabel the elements of the corresponding $N \times N$ eigenvector matrix \mathcal{C} as follows:

$$\mathcal{C} = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1i} & \dots & C_{1N} \\ C_{21} & C_{22} & \dots & C_{2i} & \dots & C_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{N1} & C_{N2} & \dots & C_{Ni} & \dots & C_{NN} \end{pmatrix}. \quad (2.10)$$

This relabelling facilitates writing down the representation for the elements of \mathcal{A} , which follows from equation (2.7). Let the reciprocal lattice vector \mathbf{g}_n occur in the n 'th row of equation (2.8) and similarly for \mathbf{g}_m . Then

$$\mathcal{A}_{\mathbf{g}_n, \mathbf{g}_m} = 2K \sum_i C_{ni} \lambda^i [C^{-1}]_{im}, \quad (2.11)$$

where the sum extends over N terms. $[C^{-1}]_{im}$ is the element in the i 'th row and m 'th column of the inverse matrix of \mathcal{C} .

Thus solution of the eigenvalue/eigenvector problem in equation (2.7), giving the elements of the matrix \mathcal{C} and the eigenvalues $2K\lambda^i$, allows us to construct the Bloch states in equation (2.4). The Bloch states form a complete set in the plane and as such the wave function of the fast electron in the crystal can be written as a sum of Bloch states with excitation amplitudes α^i :

$$\phi_0(\mathbf{r}) = \sum_i \alpha^i \phi_0^i(\mathbf{r}) = \sum_i \alpha^i e^{2\pi i \lambda^i z} \sum_{\mathbf{g}} C_{\mathbf{g}}^i e^{2\pi i \mathbf{g} \cdot \mathbf{r}_{\perp}}. \quad (2.12)$$

The α^i are obtained from the boundary conditions: the continuity of the wave function at the entrance surface of the crystal. For a plane wave this requires that the amplitude of the directly transmitted beam is 1 and the amplitudes of the diffracted beams are 0. It is found that $\alpha^i = [C^{-1}]_{i, (N+1)/2}$, i.e. it is the i 'th element in the central column of C^{-1} . It can be shown that, in the absence of absorption, $\alpha^i = C_{\mathbf{0}}^{i*}$. We will delay providing a proof (many exist in the literature [14]) until section 2.2 where it arises as a limiting case of the result for a focused probe.

At the exit surface of the crystal the Bloch waves decouple into plane waves. At this transition the tangential components remain unchanged and therefore the amplitude of the beam \mathbf{g} for a crystal of thickness t is obtained from equation (2.12) as

$$v_{\mathbf{g}}(t) = \sum_i \alpha^i C_{\mathbf{g}}^i \exp(2\pi i \lambda^i t). \quad (2.13)$$

Introducing the vector $\mathbf{v} = (v_{\mathbf{g}})$ this equation is contained in the more general result [14]

$$\mathbf{v} = \mathcal{S} \mathbf{u}, \quad (2.14)$$

where \mathcal{S} is the scattering matrix and the vector \mathbf{u} contains the Fourier components of the incident wave [$\mathbf{u} = (\delta_{\mathbf{g}\mathbf{0}})$ for a plane wave] and it can be shown that

$$\mathcal{S} = \exp\left(\frac{i\pi t}{K} \mathcal{A}\right) = \mathcal{C} [\exp(2\pi i \lambda^i t)]_D \mathcal{C}^{-1}, \quad (2.15)$$

where once again $[]_D$ denotes a diagonal matrix. The scattering matrix relates the incident electron wave at the entrance surface of the crystal to the diffracted wave at the exit surface of the crystal of thickness t .² Since \mathcal{A} is not Hermitian when absorption is included, \mathcal{S} is not unitary (i.e. in the presence of absorption the number of electrons in the elastic channels decreases with increased propagation distance into the crystal). Schematically we can represent \mathcal{S} as

$$\mathcal{S} = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & S_{\mathbf{h},\mathbf{h}} & S_{\mathbf{h},\mathbf{g}} & S_{\mathbf{h},\mathbf{0}} & S_{\mathbf{h},-\mathbf{g}} & S_{\mathbf{h},-\mathbf{h}} & \dots \\ \dots & S_{\mathbf{g},\mathbf{h}} & S_{\mathbf{g},\mathbf{g}} & S_{\mathbf{g},\mathbf{0}} & S_{\mathbf{g},-\mathbf{g}} & S_{\mathbf{g},-\mathbf{h}} & \dots \\ \dots & S_{\mathbf{0},\mathbf{h}} & S_{\mathbf{0},\mathbf{g}} & S_{\mathbf{0},\mathbf{0}} & S_{\mathbf{0},-\mathbf{g}} & S_{\mathbf{0},-\mathbf{h}} & \dots \\ \dots & S_{-\mathbf{g},\mathbf{h}} & S_{-\mathbf{g},\mathbf{g}} & S_{-\mathbf{g},\mathbf{0}} & S_{-\mathbf{g},-\mathbf{g}} & S_{-\mathbf{g},-\mathbf{h}} & \dots \\ \dots & S_{-\mathbf{h},\mathbf{h}} & S_{-\mathbf{h},\mathbf{g}} & S_{-\mathbf{h},\mathbf{0}} & S_{-\mathbf{h},-\mathbf{g}} & S_{-\mathbf{h},-\mathbf{h}} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \quad (2.16)$$

In the N -beam approximation, we have the representation

$$S_{\mathbf{g}_n,\mathbf{g}_m} = \sum_i C_{ni} \exp(2\pi i \lambda^i t) [C^{-1}]_{im} \quad (2.17)$$

for the \mathcal{S} matrix [cf. equation (2.11)]

Let us now return to the expression for the inelastic scattering cross section given by equation (1.46), which we write down again for convenience:

$$\sigma = \frac{2\pi}{\hbar v} \int_0^t \int_A \int_A \phi_0^*(\mathbf{r}_\perp, z) W(\mathbf{r}_\perp, \mathbf{r}'_\perp) \phi_0(\mathbf{r}'_\perp, z) d\mathbf{r}_\perp d\mathbf{r}'_\perp dz. \quad (2.18)$$

If we insert the Bloch wave ansatz for the wave function given by equation (2.12) into this equation then we obtain

$$\sigma = \frac{2\pi t}{\hbar v} \sum_{i,j} B^{ij}(t) \sum_{\mathbf{h},\mathbf{g}} C_{\mathbf{h}}^{j*} C_{\mathbf{g}}^i \int_A \int_A e^{-2\pi i \mathbf{h} \cdot \mathbf{r}_\perp} W(\mathbf{r}_\perp, \mathbf{r}'_\perp) e^{2\pi i \mathbf{g} \cdot \mathbf{r}'_\perp} d\mathbf{r}_\perp d\mathbf{r}'_\perp. \quad (2.19)$$

where the $B^{ij}(t)$ is given by

$$B^{ij}(t) = \alpha^i \alpha^{j*} \frac{\exp[2\pi i (\lambda^i - \lambda^{j*}) t] - 1}{2\pi i (\lambda^i - \lambda^{j*}) t}. \quad (2.20)$$

Defining

$$\mu_{\mathbf{h},\mathbf{g}} \equiv \frac{2\pi}{\hbar v A} W_{\mathbf{h},\mathbf{g}} = \frac{2\pi}{\hbar v A} \int_A \int_A e^{-2\pi i \mathbf{h} \cdot \mathbf{r}_\perp} W(\mathbf{r}_\perp, \mathbf{r}'_\perp) e^{2\pi i \mathbf{g} \cdot \mathbf{r}'_\perp} d\mathbf{r}_\perp d\mathbf{r}'_\perp, \quad (2.21)$$

allows us to write equation (2.19) in the compact form

$$\sigma = \frac{2\pi t}{\hbar v} \sum_{i,j} B^{ij}(t) \sum_{\mathbf{g},\mathbf{h}} C_{\mathbf{g}}^i C_{\mathbf{h}}^{j*} W_{\mathbf{h},\mathbf{g}} = A t \sum_{i,j} B^{ij}(t) \sum_{\mathbf{g},\mathbf{h}} C_{\mathbf{g}}^i C_{\mathbf{h}}^{j*} \mu_{\mathbf{h},\mathbf{g}}. \quad (2.22)$$

²As we have derived it, only the central column of the scattering matrix defined in equation (2.15) follows directly from equation (2.13). That the remaining columns are fully consistent with equation (2.14) is shown in Ref. [18].

Note that the area factor A introduced in the denominator in the definition of $\mu_{\mathbf{h},\mathbf{g}}$ in equation (2.22) serves to place a factor of $V = At$ in the cross-section expression above. The remainder of the expression may therefore be viewed as a cross section per unit volume. We will discuss the evaluation of the inelastic scattering coefficients $W_{\mathbf{h},\mathbf{g}}$ or $\mu_{\mathbf{h},\mathbf{g}}$ in section 3.2.

For the case of plane wave incidence the contribution to elastic scattering from electrons already absorbed from the elastic channel (mainly by TDS) can be modelled by the additional term [19]

$$\sigma' = \frac{2\pi t}{h\nu} \left[1 - \sum_{i,j} B^{ij}(t) \sum_{\mathbf{g}} C_{\mathbf{g}}^i C_{\mathbf{g}}^{j*} \right] W_{0,0} = At \left[1 - \sum_{i,j} B^{ij}(t) \sum_{\mathbf{g}} C_{\mathbf{g}}^i C_{\mathbf{g}}^{j*} \right] \mu_{0,0}. \quad (2.23)$$

[Note that $\sum_{i,j} B^{ij}(t) \sum_{\mathbf{g}} C_{\mathbf{g}}^i C_{\mathbf{g}}^{j*} = \frac{1}{V} \int_V |\psi_0(\mathbf{r})|^2 d\mathbf{r}$, and so the bracketed term simply expresses the proportion of electrons lost from the elastic wave function due to absorption.]

2.2 Boundary conditions

In section 2.1 we touched upon the boundary condition at the entrance surface of the specimen for an incident plane wave. To model propagation of a general wave across the entrance surface of a crystal we start with the (reciprocal space) complex amplitude $T(\mathbf{p})$. The most relevant example is that of a focused spherical wave in scanning transmission electron microscopy (STEM), the distorted coherent probe, a schematic of which is given in figure 2.1. In this case $T(\mathbf{p})$ is simply the contrast transfer function and may be written as [20]

$$T(\mathbf{p}) = A(\mathbf{p}) \exp \left[-i \frac{2\pi}{\lambda} \chi(\mathbf{p}) \right], \quad (2.24)$$

where the objective aperture pupil function $A(\mathbf{p})$ is equal to unity if $p \leq p_{\max}$ and zero otherwise. The aperture cutoff p_{\max} is related to the aperture semi-angle via $\alpha = \tan^{-1}(p_{\max}/k_0) \approx p_{\max}/k_0$, as seen in figure 2.1. The effect of chromatic aberration has been omitted (but may be introduced at any stage). The phase distortion χ is expressed as $2\pi/\lambda$ times the path difference travelled by off-axis waves due to lens aberration. For a lens with cylindrical symmetry, 3rd order spherical aberration coefficient C_s , 5th order spherical aberration coefficient C_5 , and defocus Δf (overfocus positive)³

$$\chi(p) = \frac{1}{2} \Delta f (\lambda p)^2 + \frac{1}{4} C_s (\lambda p)^4 + \frac{1}{6} C_5 (\lambda p)^6. \quad (2.25)$$

Let us rewrite equation (2.12) for the total elastic wave function ϕ_0 in the crystal in the form

$$\begin{aligned} \phi_0(\mathbf{R}, \mathbf{r}_{\perp}, z) &= \sum_i \alpha^i(\mathbf{R}) \phi_0^i(\mathbf{r}_{\perp}, z) \\ &= \sum_i \alpha^i(\mathbf{R}) \exp(2\pi i \lambda^i z) \sum_{\mathbf{g}} C_{\mathbf{g}}^i \exp(2\pi i \mathbf{g} \cdot \mathbf{r}_{\perp}) \\ &\equiv \sum_i \alpha^i(\mathbf{R}) \exp(2\pi i \lambda^i z) \phi_{xy}^i(\mathbf{r}_{\perp}). \end{aligned} \quad (2.26)$$

³In lenses without a spherical aberration corrector, the C_5 term is generally negligible in effect compared to the remaining terms. Only with the advent of aberration correctors has its inclusion become important.

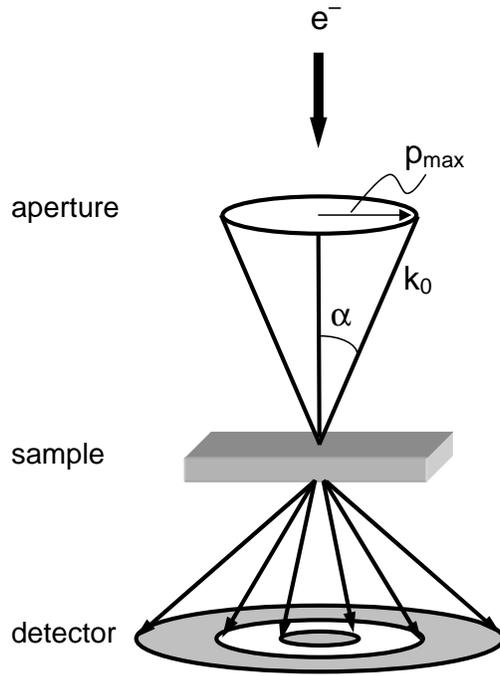


Figure 2.1: Schematic diagram of a scanning transmission electron microscope.

The excitation amplitudes $\alpha^i(\mathbf{R})$ are to be derived from the boundary conditions, and it is anticipated that they will depend on the phase distortion and focus position \mathbf{R} of the probe in the xy -plane.⁴

The real space probe amplitude may be constructed from the Fourier transform

$$T(\mathbf{r}_\perp, \mathbf{R}) = \sum_{\mathbf{p}} T(\mathbf{p}) \exp(-2\pi i \mathbf{p} \cdot \mathbf{R}) \exp(2\pi i \mathbf{p} \cdot \mathbf{r}_\perp), \quad (2.27)$$

where the phase factor $\exp(-2\pi i \mathbf{p} \cdot \mathbf{R})$ accounts for the position of the focused probe.

The use of a discrete Fourier sum in equation (2.27), necessary for numerical work, introduces an effective periodicity. In STEM this is unphysical. We can eliminate the consequences of the spurious repetitions of the probe introduced by working with Fourier frequencies such that the unavoidable periodicity is such that the duplicate probes are sufficiently far apart that they do not significantly interact. The same approach is taken in handling crystal defects in the conventional transmission electron microscope regime [21]. The total repeating unit is referred to as a supercell, and for a crystal sample will comprise several physical unit cells. Reciprocal lattice vectors \mathbf{g} are constructed from a sufficiently large supercell, with linear dimension $1/\Delta\mathbf{p}$, such that a reciprocal space mesh $\Delta\mathbf{p}$ is sampled for coupling the focused probe amplitude $T(\mathbf{p})$ directly into reflections \mathbf{g} . Thus direct excitation of beams, including those with null structure factors, occurs by direct coupling at the top surface. Each beam propagates through the perfect undistorted crystal without coupling to neighbouring beams separated in reciprocal space by $\Delta\mathbf{q}$ (a multiple of $\Delta\mathbf{p}$), unless $\Delta\mathbf{q} = \mathbf{G}$, a reciprocal lattice vector derived from the fundamental unit cell, as discussed further in reference

⁴This is in contrast to the Bloch wave approach in which a spherical wave is constructed from a coherent summation of phase-linked plane waves, and where an excitation amplitude $\alpha^i = C_0^{i*}$ (ignoring absorption) is associated with each phase-linked component [23–26], C_0^i being the zeroth order Fourier coefficient associated with the Bloch state of index i . The equivalence between these two approaches is shown in [22].

[22]. The summation over \mathbf{p} in equation (2.27) is thus taken over the mesh in \mathbf{g} defined by the supercell.

Figure 2.2 shows the intensity, phase, and aberration function for three 100 keV STEM probes. The parameters are given in the figure caption. For the first probe, the value of C_5 is unimportant because, within the aperture used, the aberration function is dominated by the behaviour of the C_s term. For the second probe, the values of both C_s and Δf have been chosen to offset the fixed value for C_5 in order to simulate an aberration-corrected probe. The third is an aberration-free probe, with aperture semi-angle 25 mrad, corresponding to $q_{\max} = 0.676 \text{ \AA}^{-1}$. The intensities have been normalized to a common maximum of unity. The phases, which are only meaningfully defined up to an additive constant, have been pinned to zero at the centre of the probe. The intensity distributions get narrower with increasing probe aperture, and the phase variation of the probes in the region of significant probe intensity is minimal. The dashed line on the plots of the aberration function denotes the position of the aperture cutoff. For both the aberrated probe and the aberration-balanced probe the aberration function varies significantly for large momentum transfers, but the aperture is positioned such that this variation does not significantly distort the resultant probe. For the aberrated probe this positioning is given by the standard Scherzer conditions.

In the absence of absorption the structure matrix \mathcal{A} is Hermitian and therefore the eigenvector matrix \mathcal{C} is unitary. This translates into an orthogonality property of the Bloch states:

$$\int_A \phi_0^{j*}(\mathbf{r}_\perp) \phi_0^i(\mathbf{r}_\perp) d\mathbf{r}_\perp = \delta_{i,j} \quad (2.28)$$

(where the states have been normalized appropriately). Applying continuity of the wave function across the crystal entrance surface in the Bloch wave model for STEM is a question of finding the excitation amplitudes such that

$$T(\mathbf{r}_\perp, \mathbf{R}) = \sum_i \alpha^i(\mathbf{R}) \phi_0^i(\mathbf{r}_\perp). \quad (2.29)$$

From the orthogonality condition in equation (2.28), it follows that the excitation amplitude is given by the overlap integral [27]

$$\alpha^i(\mathbf{R}) = \int \phi_0^{i*}(\mathbf{r}_\perp, z=0) T(\mathbf{r}_\perp, \mathbf{R}) d\mathbf{r}_\perp. \quad (2.30)$$

Taking this one step further, by use of equations (2.26) and (2.27), yields the excitation amplitude

$$\alpha^i(\mathbf{R}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{i*} \sum_{\mathbf{p}} T(\mathbf{p}) \exp(-2\pi i \mathbf{p} \cdot \mathbf{R}) \int \exp[2\pi i (\mathbf{p} - \mathbf{g}) \cdot \mathbf{r}_\perp] d\mathbf{r}_\perp. \quad (2.31)$$

This expression is approximate when absorption is present and a rigorous result replaces the term $C_{\mathbf{g}}^{i*}$ by the appropriate element in the inverse of the matrix of eigenvectors [12]. However, we prefer the clarity of exposition provided by the simpler and more usual notation. A discussion using the more rigorous notation may be found in [28].

The integral in equation (2.31) is proportional to the Kronecker delta $\delta_{\mathbf{p},\mathbf{g}}$. Absorbing the proportionality constant, which can be considered as being determined by the incident electron flux, into the $\alpha^i(\mathbf{R})$ we obtain the expression for the excitation amplitudes derived from an arbitrarily distorted incoming spherical wave

$$\alpha^i(\mathbf{R}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{i*} \exp(-2\pi i \mathbf{g} \cdot \mathbf{R}) T(\mathbf{g}) \quad (\text{general case}). \quad (2.32)$$

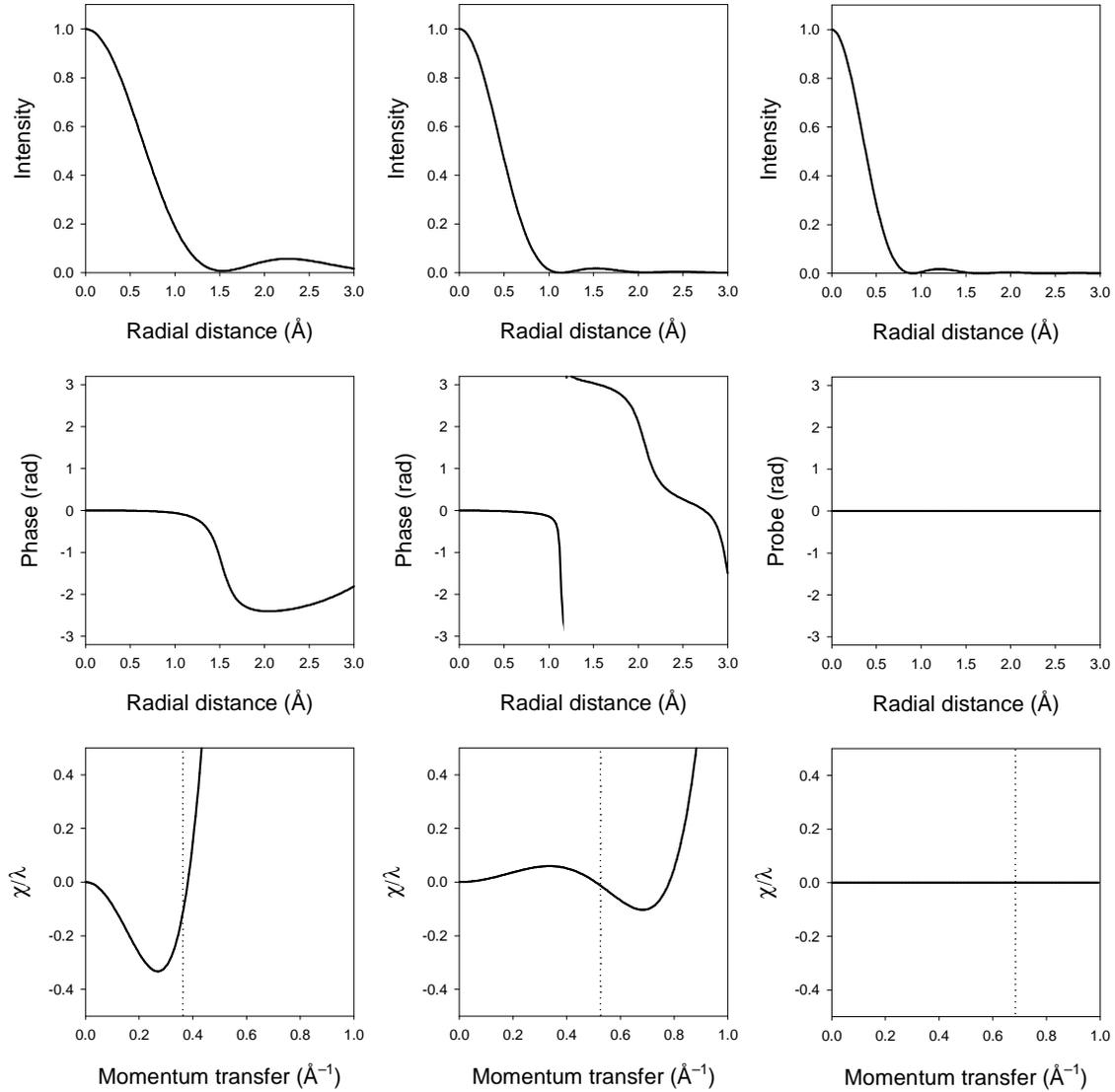


Figure 2.2: Intensity (top row), phase (middle row) and aberration function (bottom row) for: the aberrated probe, with $\Delta f = -496.7 \text{ \AA}$, $C_s = 0.5 \text{ mm}$, and aperture semi-angle of 14 mrad (left column); the aberration-balanced probe, with $\Delta f = 62 \text{ \AA}$, $C_s = -0.05 \text{ mm}$, $C_5 = 63 \text{ mm}$, and aperture semi-angle of 20 mrad (centre column); and the aberration-free probe, with aperture semi-angle of 25 mrad (right column).

This equation, in which each Fourier component $C_{\mathbf{g}}^i$ is coupled via a position-dependent phase factor to the amplitude $T(\mathbf{g})$ of the incident probe, constitutes a most fundamental point in the theory of coherent and incoherent STEM contrast.

Two limiting cases are of special interest. Plane wave incidence is characterized by setting $T(\mathbf{g})$ equal to unity for $\mathbf{g} = \mathbf{0}$ and zero for $\mathbf{g} \neq \mathbf{0}$, leading to the usual plane wave boundary condition

$$\alpha^i = C_0^{i*} \quad (\text{plane wave}). \quad (2.33)$$

A delta function (or a point source emitter of spherical waves) is characterized by equating $T(\mathbf{g})$ to unity for all \mathbf{g} , leading to the expression

$$\alpha^i(\mathbf{R}) = \sum_{\mathbf{g}} C_{\mathbf{g}}^{i*} \exp(-2\pi i \mathbf{g} \cdot \mathbf{R}) \quad (\text{point source}). \quad (2.34)$$

It is instructive to realise that equation (2.34) is merely the complex conjugate of $\phi_{xy}^i(\mathbf{r}_{\perp})$ from equation (2.26). Thus the Bloch wave excitation amplitude invoked by a point probe at a position \mathbf{R} is simply the complex conjugate of the amplitude in the Bloch wave itself at that point. The general expression leads to a good intuitive understanding of the probe wave function as a function of focus, position and depth, as well as lattice-resolution contrast derived from both coherent and incoherent signals.

2.3 Bloch states and excitation amplitudes

Figure 2.3 shows the excitation amplitudes for a few select Bloch states in ZnS, viewed along the [110] zone axis. For simplicity we will neglect absorption and so $\alpha^i = C_0^{i*}$. We use two STEM probes from figure 2.2. Figure 2.3 shows results for the 100 keV aberration-balanced probe (first column) and the 100 keV aberration-free probe (second column), as well as a delta function probe as per equation (2.34) (third column).

In figure 2.3, the number below the state label for each row is the value of $|C_0^{i*}|$, which is the magnitude of the excitation amplitude for plane wave incidence. Thus in conventional transmission electron microscopy with normal incidence, the first four Bloch states in figure 2.3 account for 99% of the electron density. However, depending on the probe position, it is seen that in the STEM case these states may not be significantly excited. Additionally, state 5, which being antisymmetric has zero excitation amplitude for normal plane wave incidence, is significantly excited for certain probe positions. Indeed all available antisymmetric states are excited depending on probe position. Contributions from all states must therefore be taken into account for the correct dispersion of the probe in real space with increasing depth.

Relative magnitudes are shown below the images. State 1 may be regarded as an s-state for the zinc column, following the parlance of Buxton *et al.* [29], and state 2 may be regarded as an s-state for the sulfur column (though in both cases faint contributions should be noted on the adjacent column site for the aberration free case). In these states especially, though also in the others, the variations in the excitation amplitudes with probe position become more pronounced for the finer probes. The excitation amplitudes suggest that the s-states incorporate most of the electron density when the probe is situated upon the column. Much has been made of s-state models in trying to find schemes which balance accuracy with ease of interpretation. However it is clear from figure 2.3 that for other

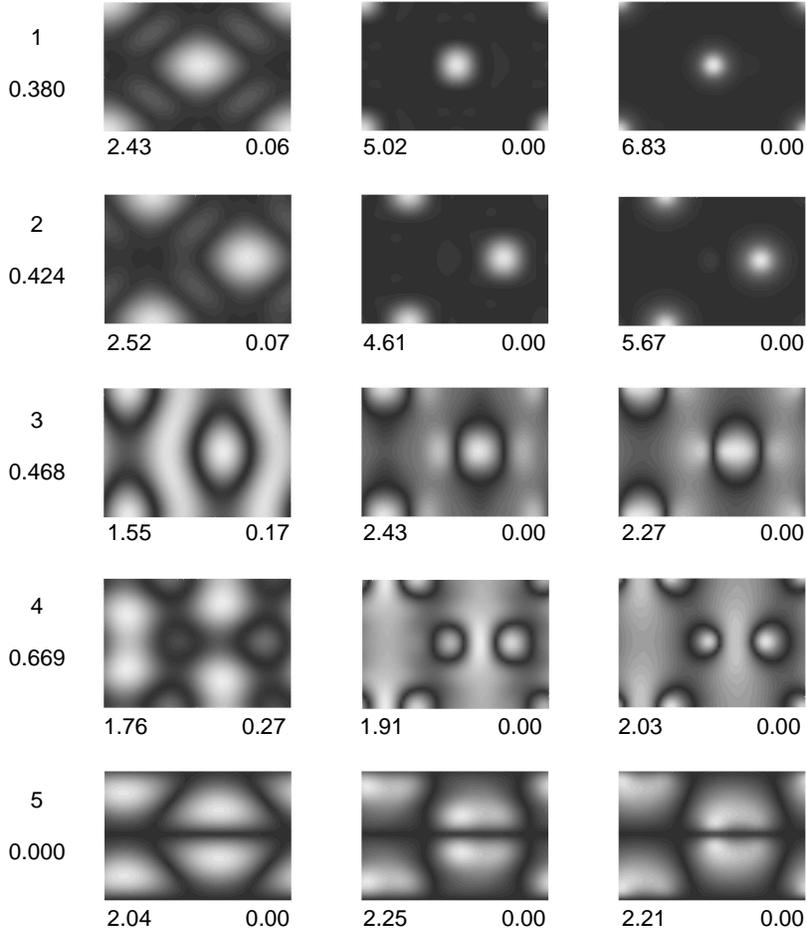


Figure 2.3: Bloch state excitation amplitudes in ZnS, viewed along the [110] zone axis, using 100 keV STEM probes. The rows are labelled with an identifying state number, below which is given the magnitude of $|C_0^{i*}|$ which would be the excitation amplitude in the case of normal plane wave incidence. The first column corresponds to the aberration-balanced probe, the second an aberration-free probe, and the third a delta-function probe. Maximum and minimum values are given below the images to provide a sense of scale.

probe positions it will be necessary to include many more Bloch states for an adequate description of the wave function and the signals arising from it [30,31]. Put another way, only for the case of plane wave incidence (in symmetrical zone axis orientation) is it true that the wave function may be adequately represented by just a few states near the top of the dispersion curve. In particular, the extent of the range of Bloch states excited is critical in assessing the spreading of the wave function about the probe location.

2.4 Lattice-resolution contrast in scanning transmission electron microscopy in the Bloch wave model

Incoherent lattice contrast can be calculated in the Bloch wave model using equation (2.22) provided that the $B^{ij}(t)$ in equation (2.20) depend on the probe position \mathbf{R} via equation (2.32). The diffuse contribution to the cross section given by equation (2.23), which proved important in quantitative analysis in the conventional transmission electron microscopy geometry [32], has been shown to be inadequate for the case of fine probes in STEM [33]. We shall give examples of the evaluation of the inelastic cross section expression for STEM in chapter 3.

Coherent (i.e. elastic scattering) bright field contrast in STEM is derived by writing the amplitude $A_{\mathbf{h}}$ scattered to a point \mathbf{h} as the Fourier transform of the exit surface wave function. Using equation (2.32) in equation (2.12) and extracting the amplitudes of the Fourier components gives

$$A_{\mathbf{h}}(\mathbf{R}) = \sum_i C_{\mathbf{h}}^i \exp(2\pi i \lambda^i t) \sum_{\mathbf{g}} C_{\mathbf{g}}^{i*} \exp(-2\pi i \mathbf{g} \cdot \mathbf{R}) T(\mathbf{g}), \quad (2.35)$$

where the index \mathbf{g} is used for points which lie within the objective aperture. A ronchigram [34] is obtained from the intensity distribution $|A_{\mathbf{h}}(\mathbf{R})|^2$ as a function of \mathbf{h} for fixed \mathbf{R} . Consider a point detector located at the origin ($\mathbf{h} = \mathbf{0}$) with p_{\max} sufficiently large that overlap between neighbouring fundamental reflections \mathbf{G} occurs at the centre of the zeroth order disk [20]. A coherent STEM lattice image $|A_0(\mathbf{R})|^2$ may be calculated as a function of probe position \mathbf{R} :

$$A_0(\mathbf{R}) = \sum_i C_0^i \exp(2\pi i \lambda^i t) \sum_{\mathbf{g}} C_{\mathbf{g}}^{i*} \exp(-2\pi i \mathbf{g} \cdot \mathbf{R}) T(\mathbf{g}). \quad (2.36)$$

2.5 Multislice solution of the paraxial equation with absorption

Let us return to equation (2.1), which for convenience we recast as

$$\frac{\partial \phi_0}{\partial z} = \frac{i}{4\pi K} \left\{ \nabla_{\perp}^2 + 4\pi^2 [U(\mathbf{r}_{\perp}) + iU'(\mathbf{r}_{\perp})] \right\} \phi_0. \quad (2.37)$$

This is an operator equation of the form

$$\frac{d\phi_0}{dz} = Q\phi_0, \quad (2.38)$$

where we have replaced the partial derivative in equation (2.37) with an exact derivative since the operator Q does not depend on the value of z . The formal, operator solution of this equation is

$$\phi_0(\mathbf{r}_{\perp}, z) = \exp(Qz)\phi_0(\mathbf{r}_{\perp}, 0). \quad (2.39)$$

Using the standard property of exponentials, this may also be written as

$$\phi_0(\mathbf{r}_{\perp}, z) = \left[\exp\left(\frac{Qz}{N}\right) \right]^N \phi_0(0) = \prod_{i=1}^N \exp\left(\frac{Qz}{N}\right) \phi_0(\mathbf{r}_{\perp}, 0) \quad (2.40)$$

or, defining $\Delta z = \frac{z}{N}$,

$$\phi_0(\mathbf{r}_{\perp}, z + \Delta z) = \exp(Q\Delta z) \phi_0(\mathbf{r}_{\perp}, z). \quad (2.41)$$

Comparing equation (2.37) and equation (2.38) we identify

$$\mathcal{Q} = \frac{i}{4\pi K} \nabla_{\perp}^2 + \frac{i\pi}{K} [U(\mathbf{r}_{\perp}) + iU'(\mathbf{r}_{\perp})] \equiv \mathcal{T} + \mathcal{U} . \quad (2.42)$$

Consider then the expression

$$\exp(\mathcal{Q}\Delta z) = \exp[(\mathcal{T} + \mathcal{U})\Delta z] . \quad (2.43)$$

It is tempting to factorise the exponential to write

$$\exp(\mathcal{Q}\Delta z) = \exp(\mathcal{T}\Delta z) \exp(\mathcal{U}\Delta z) . \quad (2.44)$$

Strictly speaking this is not justified, since the operators \mathcal{T} and \mathcal{U} do not commute. However for sufficiently small values of Δz , the result is a good approximation. Thus using the identifications of equation (2.42) in the approximation of equation (2.44), equation (2.41) may be written as

$$\phi_0(\mathbf{r}_{\perp}, z + \Delta z) = \exp\left(\frac{i\Delta z}{4\pi K} \nabla_{\perp}^2\right) \exp\left[\frac{i\pi\Delta z}{K} U(\mathbf{r}_{\perp})\right] \phi_0(\mathbf{r}_{\perp}, z) . \quad (2.45)$$

Applying the standard result

$$\nabla_{\perp}^2 f(\mathbf{r}_{\perp}) = -\mathcal{F}^{-1} \left\{ (2\pi q_{\perp})^2 \mathcal{F} [f(\mathbf{r}_{\perp})] \right\} \quad (2.46)$$

to the first exponential in equation (2.45) we obtain

$$\phi_0(\mathbf{r}_{\perp}, z + \Delta z) = \mathcal{F}^{-1} \left[\exp\left(-\frac{i\pi\Delta z q_{\perp}^2}{K}\right) \mathcal{F} \left\{ \exp\left[\frac{i\pi\Delta z}{K} U(\mathbf{r}_{\perp})\right] \phi_0(\mathbf{r}_{\perp}, z) \right\} \right] . \quad (2.47)$$

For convenience we define

$$\begin{aligned} P(\mathbf{q}_{\perp}) &= \exp\left(-\frac{i\pi\Delta z q_{\perp}^2}{K}\right) , \\ q(\mathbf{r}_{\perp}) &= \exp\left\{\frac{i\pi\Delta z}{K} [U(\mathbf{r}_{\perp}) + iU'(\mathbf{r}_{\perp})]\right\} . \end{aligned} \quad (2.48)$$

Using this notation in equation (2.47),

$$\phi_0(\mathbf{r}_{\perp}, z + \Delta z) = \mathcal{F}^{-1} \{ P(\mathbf{q}_{\perp}) \mathcal{F} [q(\mathbf{r}_{\perp}) \phi_0(\mathbf{r}_{\perp}, z)] \} . \quad (2.49)$$

Using the convolution theorem, equation (2.49) may be written as

$$\phi_0(\mathbf{r}_{\perp}, z + \Delta z) = p(\mathbf{r}_{\perp}) \otimes [q(\mathbf{r}_{\perp}) \phi_0(\mathbf{r}_{\perp}, z)] , \quad (2.50)$$

where $p(\mathbf{r}_{\perp})$ is the inverse Fourier transform of $P(\mathbf{q}_{\perp})$ and \otimes denotes a convolution. Equation (2.50) is more standard notation, but it is equation (2.49), in which the Fourier transform may be performed numerically via numerical fast Fourier transform techniques, which is usually encoded and makes the multislice a competitive technique for simulating electron diffraction. Iteration of equation (2.49) allows the wave function to be determined in a series of planes, beginning with the entrance surface and propagating through the crystal to the exit surface. In this fashion the wave function throughout

the crystal (or at least in a series of closely spaced planes) is determined. There are many numerical intricacies in performing a properly converged multislice calculation. We mention here only that a value of $\Delta z \approx 2 \text{ \AA}$ generally suffices to give fairly well converged results.

The inelastic scattering cross section given by equation (1.44) may be written in a form suitable for multislice evaluation. The wave function may be expanded in a Fourier series as

$$\phi_0(\mathbf{r}_\perp, z_i) = \sum_{\mathbf{g}} \Phi_0(\mathbf{g}, z_i) e^{2\pi\mathbf{g}\cdot\mathbf{r}_\perp} . \quad (2.51)$$

Substituting equation (2.51) into equation (1.44) gives

$$\begin{aligned} \sigma &= \sum_i d_i \frac{2\pi}{h\nu} \int_A \int_A \left[\sum_{\mathbf{h}} \Phi_0(\mathbf{h}, z_i) e^{2\pi\mathbf{h}\cdot\mathbf{r}_\perp} \right]^* W_i(\mathbf{r}_\perp, \mathbf{r}'_\perp) \left[\sum_{\mathbf{g}} \Phi_0(\mathbf{g}, z_i) e^{2\pi\mathbf{g}\cdot\mathbf{r}_\perp} \right] d\mathbf{r}_\perp d\mathbf{r}'_\perp \\ &= \sum_i d_i \sum_{\mathbf{g}, \mathbf{h}} \Phi_0^*(\mathbf{h}, z_i) \Phi_0(\mathbf{g}, z_i) \frac{2\pi}{h\nu} \int_A \int_A e^{-2\pi\mathbf{h}\cdot\mathbf{r}_\perp} W_i(\mathbf{r}_\perp, \mathbf{r}'_\perp) e^{2\pi\mathbf{g}\cdot\mathbf{r}_\perp} d\mathbf{r}_\perp d\mathbf{r}'_\perp . \end{aligned} \quad (2.52)$$

Using equation (2.21) this reduces to

$$\sigma = A \sum_i d_i \sum_{\mathbf{g}, \mathbf{h}} \Phi_0^*(\mathbf{h}, z_i) \Phi_0(\mathbf{g}, z_i) \mu_{\mathbf{h}, \mathbf{g}}^i . \quad (2.53)$$

Note that we have retained an i dependence in the inelastic scattering coefficients $\mu_{\mathbf{h}, \mathbf{g}}^i$. In the multislice construction it is straight forward to use different projected potentials in different slices, and therefore $\mu_{\mathbf{h}, \mathbf{g}}^i$ varies accordingly. In the crystalline case we assume that the same projection holds for all slices, and the i dependence may then be dropped.

It is worth emphasizing at this point that the multislice and Bloch wave methods are not in opposition. Having been derived from the same paraxial Schrödinger equation they are simply different ways of calculating the same thing. Selection between the methods should therefore be based on the application of interest. The Bloch wave method is very efficient in the case of a perfect crystal. However the multislice algorithm handles periodic and nonperiodic samples alike and proves the more adaptable to nonperiodic specimens. These issues will be further explored in the next chapter.

Chapter 3

Application: inelastic scattering in scanning transmission electron microscopy (STEM)

3.1 Introduction

The purpose of imaging models in STEM is to relate the measured signal (we usually detect inelastically scattered electrons) to the structure of the specimen.

Let us return to equation (1.46) for the cross section for inelastic scattering in our specimen, which we restate here for convenience:

$$\sigma = \frac{2\pi}{h\nu} \int_0^t \int_A \int_A \phi_0(\mathbf{r}_\perp, z) W(\mathbf{r}_\perp, \mathbf{r}'_\perp) \phi_0(\mathbf{r}'_\perp, z) d\mathbf{r}_\perp d\mathbf{r}'_\perp dz, \quad (3.1)$$

where [a reprise of equation (1.49)]

$$W(\mathbf{r}_\perp, \mathbf{r}'_\perp) \approx \frac{2\pi m}{h^2} \sum_{n \neq 0} \frac{1}{k_n} \sum_{\alpha} H_{\alpha, n0}^*(\mathbf{r}_\perp) H_{\alpha, n0}(\mathbf{r}'_\perp) \int e^{2\pi i \mathbf{K}'_\perp \cdot (\mathbf{r}_\perp - \mathbf{r}'_\perp)} d^2 \mathbf{K}'_\perp. \quad (3.2)$$

We note that the different inelastic processes leaving the crystal in different excited states and from different atomic sites contribute additively to the nonlocal potential $W(\mathbf{r}_\perp, \mathbf{r}'_\perp)$. For inelastic scattering of a given type, e.g. a specific inner-shell ionization from a particular atomic species in the specimen, we obtain the cross section by summing over the appropriate subset of transitions $0 \rightarrow n$ and subset of atomic sites α . Though less clear in equation (3.2) having made the flat Ewald sphere approximation than it was in equation (1.24), we may keep track of only those electrons that scatter into a particular solid angle (or in other words detector aperture).

If this detector aperture is large enough then the integral reduces to $\delta(\mathbf{r}_\perp - \mathbf{r}'_\perp)$ [cf. equation (1.50)] and the expression for the cross section given by equation (3.1) reduces to

$$\sigma = \frac{4\pi}{h\nu} \int_0^t \int_A |\phi_0(\mathbf{r}_\perp, z)|^2 V'(\mathbf{r}_\perp) d\mathbf{r}_\perp dz. \quad (3.3)$$

We will address the issue of how large the aperture needs to be to use equation (3.3) in section 3.4. If the experimental situation is such that we need to use equation (3.1) then the connection between the signal and the structure is subtle and not straight forward. The connection between the probe

intensity and the structure can be indirect. For example strong signals can be obtained from atoms which are effectively not illuminated by the probe [35]. In the case of the last expression there is a clearer relationship between the probe intensity and the inelastic potential (which we assume represents the structure). However the quantitative disentanglement of the probe information from the structure information (the inelastic potential) is not generally possible. It only becomes so if the probe does not channel in the specimen (thin specimen) and we can write

$$\sigma = \frac{4\pi t}{h\nu} \int_A P(\mathbf{R} - \mathbf{r}_\perp) V'(\mathbf{r}_\perp) d\mathbf{r}_\perp = \frac{4\pi t}{h\nu} P(\mathbf{R}) \otimes V'(\mathbf{R}) , \quad (3.4)$$

where $P(\mathbf{R} - \mathbf{r}_\perp)$ is the probe intensity and we can deconvolve to obtain $V'(\mathbf{R})$. This is the so-called object function approximation. Unfortunately the conditions for its validity are not often met in practice. Hence the importance of modelling in interpreting experimental results.

3.2 Inelastic scattering coefficients

Let us make a Fourier expansion for $W(\mathbf{r}_\perp, \mathbf{r}'_\perp)$ in equation (3.1) as follows:

$$W(\mathbf{r}_\perp, \mathbf{r}'_\perp) = \frac{1}{A^2} \sum_{\mathbf{h}, \mathbf{g}} W_{\mathbf{h}, \mathbf{g}} e^{2\pi i \mathbf{h} \cdot \mathbf{r}_\perp} e^{-2\pi i \mathbf{g} \cdot \mathbf{r}'_\perp} . \quad (3.5)$$

The coefficients $W_{\mathbf{h}, \mathbf{g}}$ are the inelastic scattering coefficients and \mathbf{g} and \mathbf{h} are vectors in the plane in Fourier space. Fourier transforming we have

$$W_{\mathbf{h}, \mathbf{g}} = \int_A \int_A e^{-2\pi i \mathbf{h} \cdot \mathbf{r}_\perp} W(\mathbf{r}_\perp, \mathbf{r}'_\perp) e^{2\pi i \mathbf{g} \cdot \mathbf{r}'_\perp} d\mathbf{r}_\perp d\mathbf{r}'_\perp . \quad (3.6)$$

Using the first line of equation (1.47)

$$\begin{aligned} W_{\mathbf{h}, \mathbf{g}} &= \frac{2\pi m}{h^2 t} \int_A \int_A e^{-2\pi i \mathbf{h} \cdot \mathbf{r}_\perp} e^{2\pi i \mathbf{g} \cdot \mathbf{r}'_\perp} \sum_{\alpha} \sum_{n \neq 0} H_{\alpha, n 0}^*(\mathbf{r}_\perp) H_{\alpha, n 0}(\mathbf{r}'_\perp) k_n \\ &\quad \times \int e^{2\pi i \mathbf{K}'_\perp \cdot (\mathbf{r}_\perp - \mathbf{r}'_\perp)} \delta(k_n - K') d\Omega_{K'} dK' d\mathbf{r}_\perp d\mathbf{r}'_\perp \end{aligned} \quad (3.7)$$

(note that we have reverted to integration over the sphere as determined by energy conservation). We rewrite this as follows:

$$\begin{aligned} W_{\mathbf{h}, \mathbf{g}} &= \frac{2\pi m}{h^2 t} \sum_{\alpha} \sum_{n \neq 0} k_n \int \left[\int_A H_{n 0}^*(\mathbf{r}_\perp - \boldsymbol{\tau}_\alpha) e^{-2\pi i (-\mathbf{K}'_\perp + \mathbf{h}) \cdot \mathbf{r}_\perp} d\mathbf{r}_\perp \right] \\ &\quad \times \left[\int_A H_{n 0}(\mathbf{r}'_\perp - \boldsymbol{\tau}_\alpha) e^{2\pi i (-\mathbf{K}'_\perp + \mathbf{g}) \cdot \mathbf{r}'_\perp} d\mathbf{r}'_\perp \right] \delta(k_n - K') d\Omega_{K'} dK' , \end{aligned} \quad (3.8)$$

where $\boldsymbol{\tau}_\alpha$ denotes the position of atom α in the xy -plane. Letting $\mathbf{r}_\perp - \boldsymbol{\tau}_\alpha \rightarrow \mathbf{r}_\perp$ and $\mathbf{r}'_\perp - \boldsymbol{\tau}_\alpha \rightarrow \mathbf{r}'_\perp$ and $\mathbf{Q} = \mathbf{K} - \mathbf{K}'$ this becomes

$$W_{\mathbf{h}, \mathbf{g}} = \frac{2\pi m}{h^2 t} \sum_{\alpha} e^{2\pi i (\mathbf{g} - \mathbf{h}) \cdot \boldsymbol{\tau}_\alpha} \sum_{n \neq 0} k_n \int H_{n 0}^*(\mathbf{Q} + \mathbf{h}) H_{n 0}(\mathbf{Q} + \mathbf{g}) \delta(k_n - K') d\Omega_{K'} dK' , \quad (3.9)$$

where $H_{n0}(\mathbf{Q} + \mathbf{g})$ is defined by

$$\begin{aligned}
H_{n0}(\mathbf{Q} + \mathbf{g}) &= \int H_{n0}(\mathbf{r}_\perp) e^{2\pi i(\mathbf{Q} + \mathbf{g}) \cdot \mathbf{r}_\perp} d\mathbf{r}_\perp \\
&= \int_A \left[\int H_{n0}(\mathbf{r}) e^{2\pi i(K - K')z} dz \right] e^{2\pi i(\mathbf{Q} + \mathbf{g}) \cdot \mathbf{r}_\perp} d\mathbf{r}_\perp \\
&= \int_V H_{n0}(\mathbf{r}) e^{2\pi i(\mathbf{Q} + \mathbf{g}) \cdot \mathbf{r}} d\mathbf{r}, \tag{3.10}
\end{aligned}$$

where in the second step we have made use of equation (1.48) and in the final step have identified $Q_z \approx K - K'$ [consistent with the approximation made in and the discussions following equation (1.48)]. We have also assumed that the incident electron beam has wave vector \mathbf{K} in the z -direction.

Incoherence of the contributions to the cross section from inelastic scattering at different atomic sites in the xy -plane is catered for in the construction of the $W_{\mathbf{h},\mathbf{g}}$. Assuming that the transition $0 \rightarrow n$ is simply an atomic transition which proceeds via the Coulomb interaction we can write

$$H_{n0}(\mathbf{r}) = \frac{e^2}{4\pi\epsilon_0} \int u_n^*(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} u_0(\mathbf{r}') d\mathbf{r}', \tag{3.11}$$

where u_0 and u_n denote the initial and final state wave functions respectively of the atom. Using this and the standard result

$$\int \frac{e^{2\pi i\mathbf{Q} \cdot \mathbf{r}}}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} = \frac{e^{2\pi i\mathbf{Q} \cdot \mathbf{r}'}}{\pi Q^2} \tag{3.12}$$

we can show that

$$H_{n0}(\mathbf{Q} + \mathbf{h}) = \frac{e^2}{4\pi^2\epsilon_0} \frac{F_{n0}(\mathbf{Q} + \mathbf{h})}{|\mathbf{Q} + \mathbf{h}|^2}, \quad \text{where} \quad F_{n0}(\mathbf{Q}) = \int u_n^*(\mathbf{r}) e^{2\pi i\mathbf{Q} \cdot \mathbf{r}} u_0(\mathbf{r}) d\mathbf{r}. \tag{3.13}$$

So we can rewrite equation (3.9) in the form

$$W_{\mathbf{h},\mathbf{g}} = \frac{2\pi m}{h^2 t} \left(\frac{e^2}{4\pi^2\epsilon_0} \right)^2 \sum_\alpha e^{2\pi i(\mathbf{g} - \mathbf{h}) \cdot \boldsymbol{\tau}_\alpha} \sum_{n \neq 0} k_n \int \frac{F_{n0}^*(\mathbf{Q} + \mathbf{h})}{|\mathbf{Q} + \mathbf{h}|^2} \frac{F_{n0}(\mathbf{Q} + \mathbf{g})}{|\mathbf{Q} + \mathbf{g}|^2} \delta(k_n - K') d\Omega_{K'} dK', \tag{3.14}$$

Introducing the Bohr radius $a_0 = \epsilon_0 h^2 / \pi m e^2$, the constants on the right-hand side simplify to

$$W_{\mathbf{h},\mathbf{g}} = \frac{h^2}{8\pi^5 m a_0^2 t} \sum_\alpha e^{2\pi i(\mathbf{g} - \mathbf{h}) \cdot \boldsymbol{\tau}_\alpha} \sum_{n \neq 0} k_n \int \frac{F_{n0}^*(\mathbf{Q} + \mathbf{h})}{|\mathbf{Q} + \mathbf{h}|^2} \frac{F_{n0}(\mathbf{Q} + \mathbf{g})}{|\mathbf{Q} + \mathbf{g}|^2} \delta(k_n - K') d\Omega_{K'} dK', \tag{3.15}$$

In previous work the Fourier coefficients have been redefined to absorb the factor in front of the integrals in equation (3.1) to write the inelastic scattering coefficients as [cf. equations (2.21) and (2.22)]

$$\begin{aligned}
\mu_{\mathbf{h},\mathbf{g}} &= \frac{2\pi}{h\nu A} W_{\mathbf{h},\mathbf{g}} = \frac{2\pi m}{h^2 k_0 A} W_{\mathbf{h},\mathbf{g}} \\
&= \frac{N}{4\pi^4 k_0 a_0^2 A t} \sum_{\substack{\alpha \text{ in} \\ \text{unit cell}}} e^{2\pi i(\mathbf{g} - \mathbf{h}) \cdot \boldsymbol{\tau}_\alpha} \sum_{n \neq 0} k_n \int \frac{F_{n0}^*(\mathbf{Q} + \mathbf{h})}{|\mathbf{Q} + \mathbf{h}|^2} \frac{F_{n0}(\mathbf{Q} + \mathbf{g})}{|\mathbf{Q} + \mathbf{g}|^2} \delta(k_n - K') d\Omega_{K'} dK' \tag{3.16}
\end{aligned}$$

where N is the number of unit cells in the crystal.

The inelastic scattering coefficients $W_{\mathbf{h},\mathbf{g}}$ (or $\mu_{\mathbf{h},\mathbf{g}}$, since, the difference being only a dimensional constant, the term is applied to both) contain the mixed dynamic form factors (MDFFs) which are required for all forms of incoherent scattering. They are defined by

$$S(\mathbf{Q} + \mathbf{h}, \mathbf{Q} + \mathbf{g}, E_{n'}) \equiv \sum_{n' \neq 0} F_{n'0}^*(\mathbf{Q} + \mathbf{h}) F_{n'0}(\mathbf{Q} + \mathbf{g}), \quad (3.17)$$

where the sum over n' is a restriction to the sum over states n for the particular energy loss $E_{n'}$. Discussions about the MDFFs can be found in [36–39].¹

Specific equations for the inelastic scattering coefficients, which form an essential part of the underlying physics for both thermal diffuse scattering (TDS) and ionization will now be given.

For annular dark field (ADF) or back-scattered electron (BSE) contrast, an Einstein model for TDS may be used to obtain inelastic scattering factors [15, 40]. The inelastic scattering factors are given by²

$$\mu_{\mathbf{h},\mathbf{g}} = \frac{1}{V_c} \sum_n \exp[2\pi i(\mathbf{g} - \mathbf{h}) \cdot \boldsymbol{\tau}_n] \int f_n(\mathbf{Q} + \mathbf{g}) f_n^*(\mathbf{Q} + \mathbf{h}) \times \\ \{ \exp[-M_n(\mathbf{g} - \mathbf{h})] - \exp[-M_n(\mathbf{Q} + \mathbf{g}) - M_n(\mathbf{Q} + \mathbf{h})] \} d\Omega_{K'}, \quad (3.18)$$

where $V_c = At/N$ is the volume of a unit cell. The sum over n encompasses all atoms with scattering factors f_n in positions $\boldsymbol{\tau}_n$ in the unit cell of volume V_c . The Debye-Waller factor $M_n(\mathbf{g}) = 2\pi^2 \langle u_n^2 \rangle g^2$, where $\langle u_n^2 \rangle$ is the projected mean square thermal displacement. The range of integration over $d\Omega_{K'}$ is determined by the geometry of the detector.³ The zeroth order component $\mu_{0,0}$ is just the reciprocal of the mean free path for scattering of an (undiffracted) plane wave into the detector.

For contrast by EELS or X-ray emissions in energy dispersive X-ray (EDX) analysis, the inelastic scattering coefficients for ionization of atoms of a particular type are written

$$\mu_{\mathbf{h},\mathbf{g}} = \frac{1}{2\pi K V_c} \sum_n \exp[-M(\mathbf{g} - \mathbf{h})] \exp[2\pi i(\mathbf{g} - \mathbf{h}) \cdot \boldsymbol{\tau}_n] f(\mathbf{h}, \mathbf{g}), \quad (3.19)$$

where the sum over n is now restricted to the pertinent species in the unit cell. The Debye-Waller term accounts for thermal smearing of the ionization potential [41]. The form factor $f(\mathbf{h}, \mathbf{g})$ contains a product of a pair of non-diagonal transition amplitudes, each derived from an isolated transition matrix element for an (e,2e) interaction [19, 42–45]. We may write $f(\mathbf{h}, \mathbf{g})$ as [46, 47]

$$f(\mathbf{h}, \mathbf{g}) = \frac{1}{2\pi^3 a_0^2} \int \int K' \frac{F^*(\mathbf{Q} + \mathbf{h}, \boldsymbol{\kappa}) F(\mathbf{Q} + \mathbf{g}, \boldsymbol{\kappa})}{|\mathbf{Q} + \mathbf{h}|^2 |\mathbf{Q} + \mathbf{g}|^2} d\boldsymbol{\kappa} d\Omega_{K'}, \quad (3.20)$$

where a_0 is the relativistic Bohr radius and the transition matrix element

$$F(\mathbf{Q} + \mathbf{g}, \boldsymbol{\kappa}) = \int u_f(\boldsymbol{\kappa}, \mathbf{r}') \exp[2\pi i(\mathbf{Q} + \mathbf{g}) \cdot \mathbf{r}'] u_i(\mathbf{r}') d\mathbf{r}'. \quad (3.21)$$

¹Note that occasionally the $W_{\mathbf{h},\mathbf{g}}$ or the $\mu_{\mathbf{h},\mathbf{g}}$ are referred to as the MDFFs. This is not strictly correct, but the difference is only the inclusion of the detector geometry and possibly an integration over an energy window.

²It should be noted that equation (3.18) is seldom derived explicitly from equation (3.16), however the similarity in form is notable.

³In calculating the contribution from thermal scattering to the terms $W_{\mathbf{g},\mathbf{h}} = W_{\mathbf{g}-\mathbf{h}}$ in equation (2.8) the integration is over the whole solid angle.

Here $u_i(\mathbf{r}')$ and $u_f(\boldsymbol{\kappa}, \mathbf{r}')$ denote appropriately normalized bound and continuum state wave functions of the target electron \mathbf{r}' , ejected with momentum $\boldsymbol{\kappa}$. These transition matrix elements may be evaluated using realistic wave functions in an angular momentum representation [45, 46, 48]. Relative quantum mechanical phase between the two transition matrix elements $F(\mathbf{Q} + \mathbf{g}, \boldsymbol{\kappa})$ and $F(\mathbf{Q} + \mathbf{h}, \boldsymbol{\kappa})$ depends on both magnitude and direction of $\boldsymbol{\kappa}$ with respect to the non-aligned vectors $\mathbf{Q} + \mathbf{g}$ and $\mathbf{Q} + \mathbf{h}$. In fact, this phasing may cause individual terms in the summation over intensity contributions to be negative, although the summed intensity over all components is inevitably greater than zero. This may play a vital role in the integration over an appropriate range of energy and momentum transfers $\hbar\mathbf{Q}$ as defined by the EELS detection geometry and energy window. For ionization events which are detected by the emission of characteristic X-rays (i.e. the EDX signal), integration over the full energy window (from ionization threshold to the probe energy) and all momentum transfers is necessary. For further discussion of inelastic scattering coefficients and their importance see [36, 38, 39].

3.3 Introductory examples of inelastic STEM imaging

As an example of the sorts of STEM images which may be simulated, let us take ZnS in the $\langle 110 \rangle$ zone axis orientation probed with 100 keV electrons. Lattice images for ADF, EELS and EDX calculated using equation (2.22) assuming STEM probe incidence are shown in figure 3.1. The projected $\langle 110 \rangle$ potential for the zinc-blende structure ZnS at 100 K is shown in figure 3.1(a), where Debye-Waller factors are obtained from Reid [49], with characteristic Zn-S dumb-bells evident. Figure 3.1(b) shows the ADF image assuming a 60-160 mrad detector. The EELS images in figures 3.1(c) and (d), the Zn L -shell and S K -shell ionization respectively, are simulated for an on-axis detector with 20 mrad aperture semi-angle and integrating over an energy window of 40 eV above the ionization threshold.⁴ Figures 3.1(e) and (f) show EDX images. The maximum and minimum values in units of fractional intensity (i.e. detected current as a fraction of the incident current) are provided below each image. Note that those for EDX assume a full 4π steradian detector, and so must be re-scaled for realistic detectors, and neglect the attenuation of signal due to absorption of the X-rays in leaving the crystal.

These simulations were performed using the Bloch wave method. For the case of perfect crystals, a block diagonalization approach to the solution of the eigenvalue/eigenvector problem makes the Bloch wave method very efficient for the calculation of STEM images [22, 30], which, in particular, means calculating a full two-dimensional scan is tractable. For such cases the Bloch wave method is to be preferred. However, consider now the simulation of diffraction through a crystal defect. The standard method of treating such structures is to use supercells (the method of periodic continuation [50, 51]) and so is naturally suited to the requirements of multislice calculations. For the Bloch wave method this is a problem, because determination of the Bloch states must now be carried out on the full supercell. But the multislice calculation of the wave function on the supercell is equally amenable to this case as to the periodic case. For such cases it is therefore the multislice method that is to be preferred. Another consideration is that the multislice calculation time scales linearly with

⁴Integration over a relatively large collection angle averages the dynamical final states of the scattered electron in such a way that an effective plane wave representation [single channelling, cf. the discussion above equation (1.15)] is a good approximation [55], enabling EELS contrast to be calculated in this single channelling formulation. The smaller the collection aperture, however, the poorer this approximation, and it may be necessary to extend the single channelling formulation to a full double channelling description [43, 44, 55].

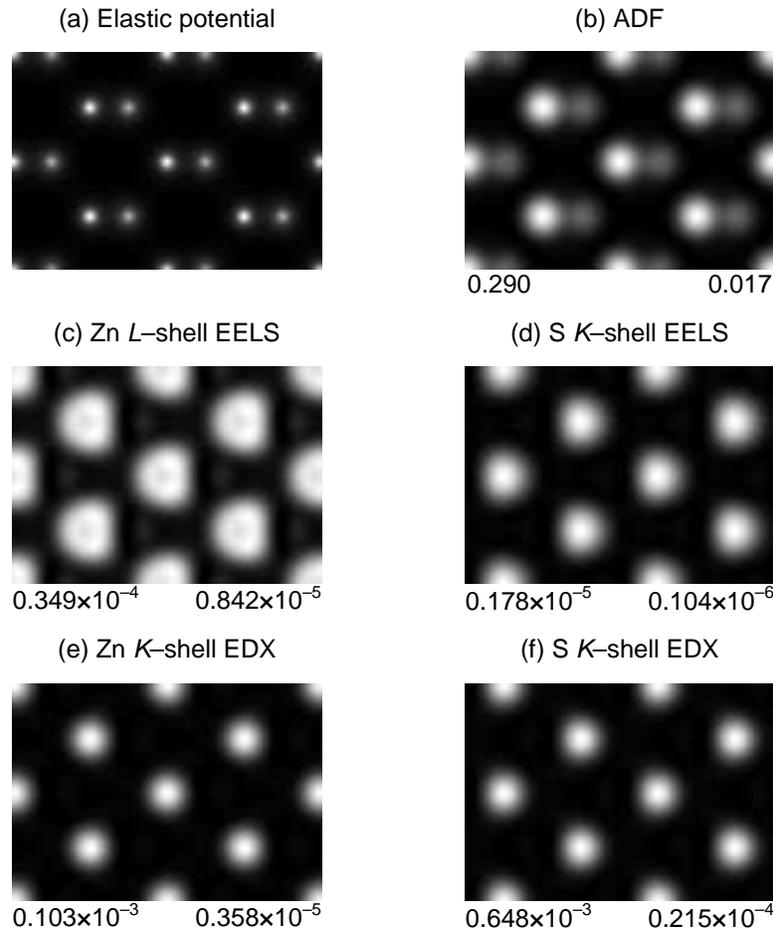


Figure 3.1: STEM images for the case of electrons incident along the $[110]$ zone axis of ZnS, thickness of 122.4 \AA , with incident energy of 100 keV . The probe is aberration-balanced. (a) Elastic potential. (b) ADF image with $60\text{-}160 \text{ mrad}$ detector. (c) Zn L -shell EELS image. (d) S K -shell EELS image. (e) Zn K -shell EDX image. (f) S K -shell EDX image. Maximum and minimum fractional intensities are given below the images.

thickness whilst the Bloch wave calculation time is independent of the value of the thickness. Thus if the periodicity makes the Bloch wave method at all tractable for simulating the wave function then for thick crystals it is to be strongly preferred over the multislice method.

3.4 Spreading of the probe and quantitative tests of cross-talk

Image simulation in STEM has long been concerned with the extent to which ADF images may be visually interpreted. At lower resolutions, false peaks may arise due to probe tails situated over adjacent columns [52]. However the scattering in such cases may conform to the requirements of the object function approximation [equation (3.4)] and this class of artifacts may be removed by probe deconvolution in the incoherent image inversion procedure [53, 54]. Improvements in resolution, most notably through the advent of C_s correctors, have allowed for probes capable of sub-angstrom resolution and as such column-by-column imaging, particularly in the ADF imaging

mode, is routine. Since the interpretation of such images is predominantly visual bright spots being interpreted as atomic columns more recent theoretical investigations have considered the extent to which false spots may arise as the result of probe spreading, the significant dynamical effect which simplified incoherent imaging models such as that given by equation (3.4) perforce neglect. Hillyard *et al.* [56], Plamann and Hýtch [57] and Dwyer and Etheridge [58] have demonstrated that the probe spreads considerably for thicker specimens, with the implication that column-by-column spectroscopy may be questionable in those circumstances. We refer to the possibility that a probe, initially focused onto an atomic column, may interact with atoms in neighbouring columns as cross-talk. Ishizuka [59,60] has explored the quantitative validity of some qualitative features often attributed to STEM images, such as their insensitivity to thickness and defocus, and the approximate Z^2 scaling of the column contrast.

The formulation for incoherent contrast presented in the section 3.2 allows this issue to be addressed in a quantitative manner. This is done by altering the summation over positions τ_n in the inelastic scattering coefficients to include or exclude various projected columns from contributing to the incoherent scattering process, whilst propagation of the wave function through the crystal remains unaltered [28,30]. To demonstrate this we consider multislice simulations for ADF imaging were performed on ZnS in [110] zone axis orientation at room temperature, using the 100 keV, aberration-balanced probe described in section 2.2.

Figure 3.2(a) shows the product of the probe intensity and the ADF potential as a function of thickness along a single, horizontal line intersecting the dumbbells when the probe is situated on the zinc column. The contribution from the uppermost slices is strong but for increasing depths decays with a *pendellösung* effect, until the contribution becomes quite small above $z = 200$ Å. By this depth, using an Einstein model for absorption by thermal scattering, the integrated probe intensity has dropped to about half the input intensity. Figure 3.2(b) shows the product of probe intensity and ADF scattering potential with the probe centred on the sulfur column. This signal is qualitatively similar in shape but smaller in magnitude than that on the zinc column, and persists to greater depths. This is a manifestation of the weakened degree of absorption, in this instance due to the difference in atomic number rather than probe energy. The rate of decay along the sulfur column is slower than for the zinc columns, this being a consequence of smaller thermal attenuation for s-type waves on sulfur compared with similar waves on zinc. The z -dependent *pendellösung* has a longer period on columns of sulfur than on columns of zinc.

Real space dispersion of the focused probe occurs within the dynamically diffracting environment, as has been shown (on a multislice model) by a number of authors [56–59]. When focused on an atomic column, the central probe initially becomes more strongly localized on that atomic column, and then gradually disperses somewhat onto neighbouring columns. Note the slight contributions from the neighbouring zinc column in figure 3.2(b) located around depths in the range 100-300 Å. This is direct evidence of crosstalk, a contribution to the signal deriving from the zinc column when the probe is situated upon the sulfur column. Note however that this cross-talk signal is small. Thus while the dispersion of intensity away from the sulfur column may be significant, only a small portion is sufficiently concentrated upon the adjacent column to give a contribution to the ADF image.

Figure 3.2(c) and (d) demonstrate quantitatively the amount of cross-talk incurred for a probe focused onto either of neighbouring columns of zinc or sulfur. The different plots represent the contributions per slice to the ADF signal for the following four scenarios, in which the ADF potential is: (i) taken into account for the full crystal; (ii) limited to the column alone; (iii) limited to the neigh-

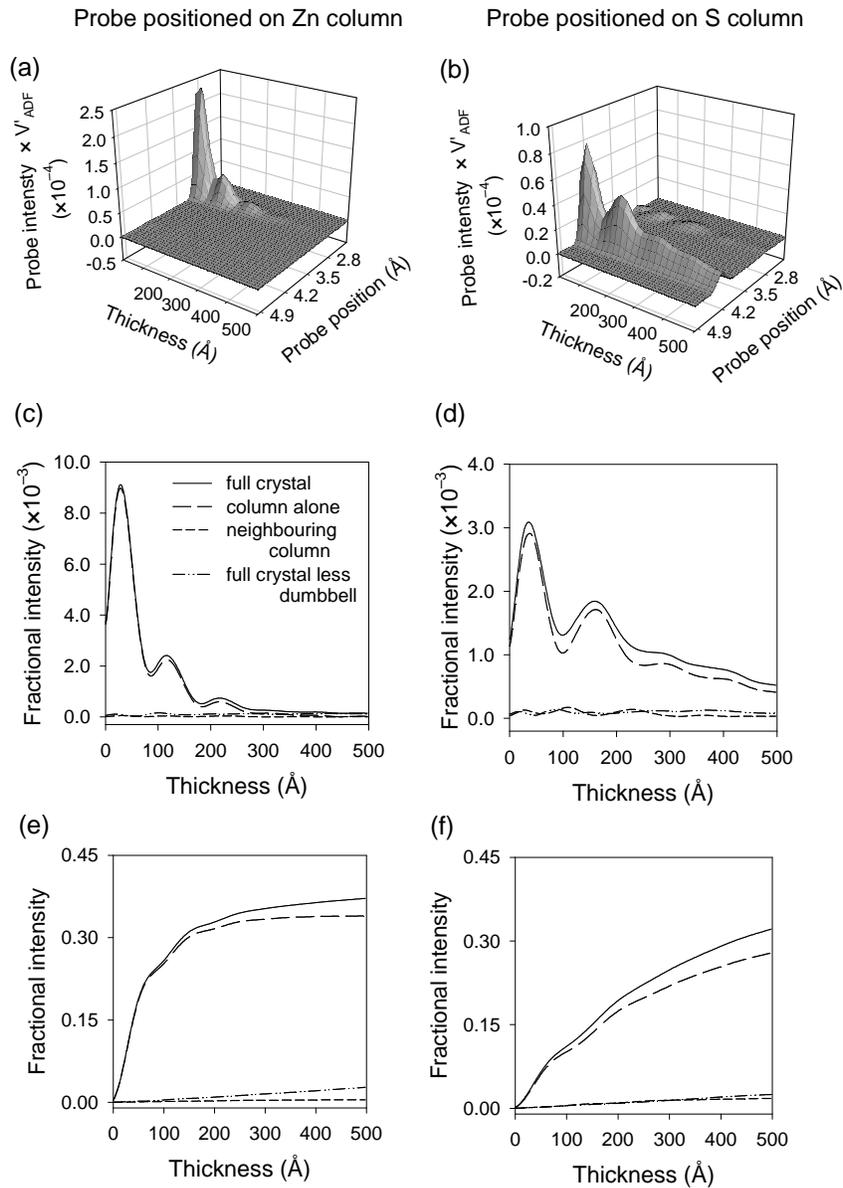


Figure 3.2: ADF simulations using the 100 keV, aberration-balanced probe on ZnS along [011]. The product of the probe intensity and the ADF potential as a function of thickness along a single, horizontal line intersecting the dumbbells with the probe situated on (a) the Zn column and (b) the S column. For the probe situated on (c) the Zn column and (d) the S column, the contribution to the fractional intensity from the previous 2 Å thick slice is shown for the ADF signal using the form factor for: the full crystal (solid line), the column beneath the probe (dashed line), the neighbouring column in the dumbbell (dash-dotted line), and the target excluding the dumbbell in question (dotted line). The fractional intensity is shown in (e) and (f) as a function of sample thickness.

bouring column in the dumbbell; and (iv) taken from the full crystal, but with contributions from the local dumbbell excised. Contributions (ii)-(iv) naturally add up to give contribution (i), the total ADF signal. Crosstalk plays a more significant role when the probe is focused onto a sulfur column. The integrated total intensity for a given thickness is shown in figures 3.2(e) and (f). The cumulative effect of cross-talk is relatively small for the probe on the zinc column. Cross-talk becomes more

significant when positioned on the sulfur column, though it is still a very small effect for the case shown here. It is interesting to note that for a thick enough specimen the contributions from the full crystal less the dumbbell exceed those from the neighbouring column.

3.5 The role of absorption

The possibility for the evolution of the probe to affect features in the image, in particular to give ambiguous results by providing signal from an admixture of columns if the probe spreads out onto them, is well covered in the literature, particularly under the heading of cross-talk [30, 57–61]. Two other mechanisms have recently been noted which might significantly alter the interpretation of the images. One is the imaging-at-a-distance result of Oxley et al. [35], to be discussed in the next section. The other is the result of significant absorption on a finely focused probe [33], which we shall briefly review here.

The absorption potential for thermal diffuse scattering is closely related to the inelastic scattering coefficients given in Eq. (3.18) when the integral is taken over the full solid angle. In this case the local approximation is excellent. Plots of the effective absorption potential for thermal scattering show it to be highly peaked at the atomic columns [62]. Thus absorption due to thermal scattering occurs predominantly on the columns. The more tightly focused a probe is, the smaller the diameter in which the majority of the intensity is enclosed, the greater the proportion of the electron density which may be absorbed by thermal scattering if the probe is located upon a column (see Ref. [33] for a discussion about how this may be quantified). The reduction in electron density in the elastic wave function with increasing depth into the sample suggests that the contribution from each depth will attenuate with depth even in the absence of probe spreading. This gives rise to a circumstance where the signal from some other scattering event, inner-shell ionization say, is significantly reduced if there is a high degree of absorption.

Consider a silver crystal, 100 Å thick, viewed along the [001] zone axis. We shall take EELS line scans along the [100] direction, using a 100 keV, aberration-balanced probe ($C_s = -0.05$ mm, $C_5 = 63$ mm, $\Delta f = 62$ Å) with a probe-forming aperture angle of 20 mrad. Assuming a 40 eV energy window above the appropriate threshold, figure 3.3(a) shows single atom EELS images for the *L*-shell and the *M*-shell. The FWHM is about 1.1 Å for the *L*-shell and 1.6 Å for the more loosely bound *M*-shell. In projection, the distance between silver atoms along the [100] direction is 2 Å. So, by the separation measure alone, we might expect these atoms to be resolvable, even in the *M*-shell signal, as shown in figure 3.3(b). Consider an absorptive model, in which an absorption due to thermal scattering is used in calculating the elastic wave function [30], to calculate the STEM image from the crystal. We then see from figure 3.3(c) that for the *M*-shell case the maximum peak is between the atomic columns rather than on them. Figure 3.3 (d) shows the *L*-shell result to better correspond to the structure. Again in the absorptive model, figures 3.3(e) and (f) show the contribution per slice to the signals in figures 3.3(c) and (d). It is seen that while the signal from the first few depths is reminiscent of the single atom case, cf. figure 3.3(b) for the *M*-shell, the signal on the columns drops off very quickly. For the *L*-shell this does not lead to reversal of contrast because this more tightly bound shell contributes little to the signal when the probe is between the columns. But, for the more loosely bound *M*-shell, the contribution between the columns is significant and does not drop off nearly so rapidly as when the probe is on the column. This difference can be explained simply by the differing amounts of absorption resulting for these different probe positions. The result, for the *M*-shell signal, is that the cumulative signal favours the gap between columns rather than the columns

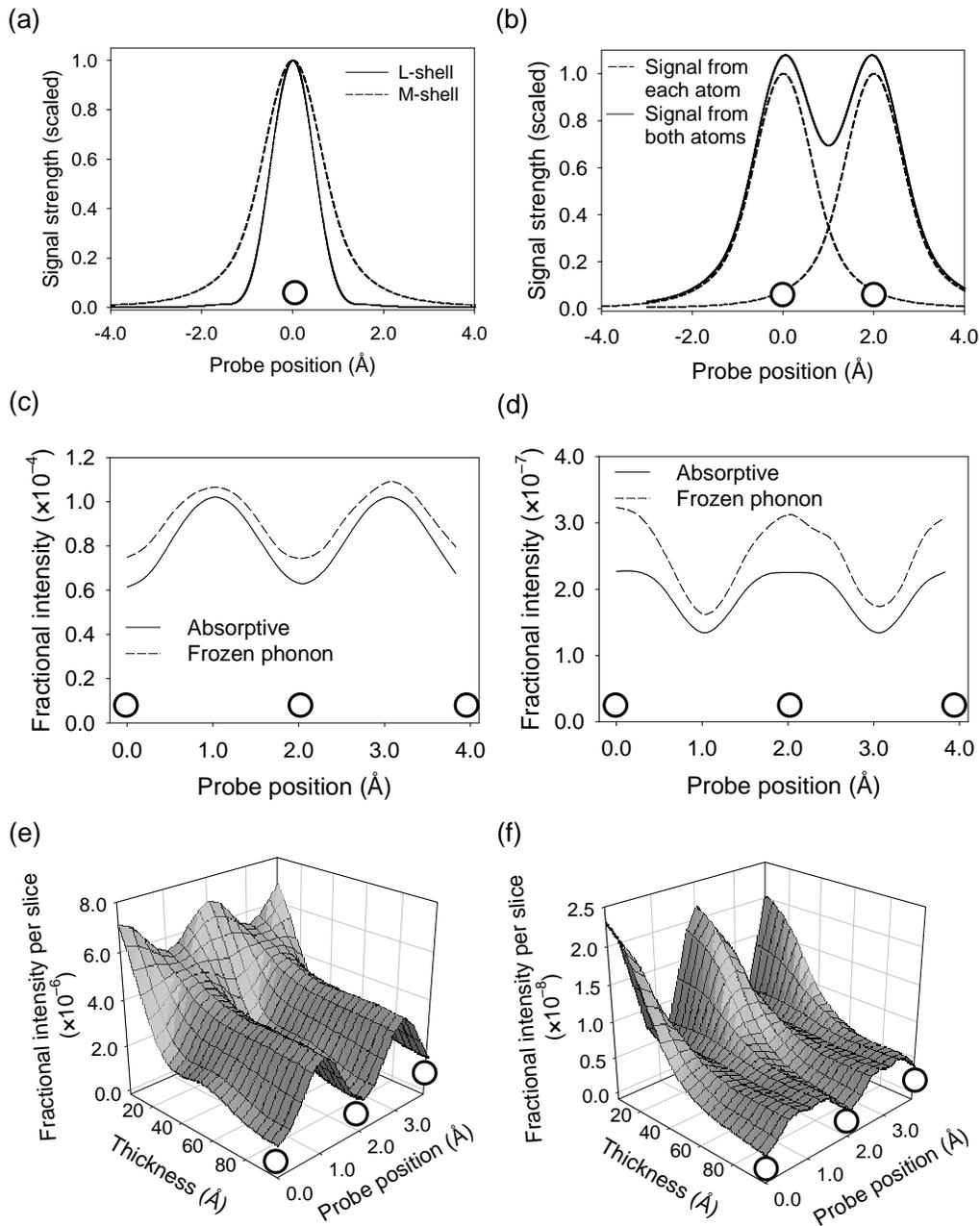


Figure 3.3: (a) Single atom EELS images for L -shell and M -shell ionization of silver. (b) Individual and total M -shell signals for two silver atoms separated by 2 Å. (c) M -shell and (d) L -shell in Ag [001] simulated with the absorptive model and the frozen phonon model. (e) and (f) show the contributions per slice for the M -shell and L -shell results respectively using the absorptive model. The positions of columns of silver atoms are indicated by the circles.

themselves.

One may object that, while the reduction in electron density within the elastic beams necessarily reduces the contribution from the elastic beams, the thermally absorbed electrons are still in the crystal and still capable of causing detectable ionization events. Findlay *et al.* [33] presented two

models to incorporate the contribution from these electrons into the total EELS signal. One of these involved a synthesis between the MDFF method and the frozen phonon method [63,64], and these results are also shown in figure 3.3(c) and (d). While in both figures the inclusion of the contribution from thermally scattered electrons leads to a quantitative correction, over 30% for the L -shell signal, the shape of the line scans and correspondingly the visual interpretations which would be made are not much changed. In particular, the additional contribution is not sufficient to give positive contrast back to the column in the M -shell case.

3.6 Nonlocality and “imaging at a distance”

Direct experimental evidence has recently been seen for the effects of nonlocality [65,66], and the possibility of counter-intuitive results due to nonlocal effects has been discussed [35] and we shall review that result here.

Let us now consider STEM using 200 keV electrons incident along the [011] zone axis of a slab of SiC, which we are illuminating with an aberration free probe formed using a 50 mrad aperture (planned for the next generation of microscopes). We calculate, using a frozen phonon model [33,67], the variation with probe position in the number of electrons which have ionized an electron in the C K -shell and detected in an axial EELS detector with acceptance semi-angle 10 mrad and an energy window of 40 eV. In figure 3.4(a) we show a line scan along the [100] direction. The locations of Si columns are indicated by grey circles and of C columns by black circles.

When the probe is situated above the Si column then one is obtaining a signal which would usually be interpreted either as the presence of C atoms in that column or a strong cross talk signal from the adjacent C column. The former is not true and by the latter we mean that the probe has spread *significantly* onto the C column as it channels through the crystal. However, plotting the probe wave function within the crystal as a function of depth we find that this is simply not the case, as can be seen in figure 3.4(b). The integrated intensity for the probe on both the Si and C columns shown in figure 3.4(c) confirms this. We are in fact obtaining a larger signal on the Si column, when there is negligible (integrated) intensity on the C column, than when the probe is on the C column itself and there is considerable flux on that column.

This non-intuitive connection is one possible consequence of a nonlocal effective interaction. Getting a feel for the nonlocal potential is difficult because, even working with the projected approximation to $W(\mathbf{r}, \mathbf{r}')$, $W(\mathbf{r}_\perp, \mathbf{r}'_\perp)$ is a four dimensional function and as such is difficult to visualize. The terms localized and delocalized are misleading because they are also applied to local potentials and we shall not use them as descriptors for the extent of nonlocality. However, such descriptors can usefully be applied to the features in the STEM images. To this end Cosgriff *et al.* [68] systematically explored the widths of single atom STEM images for a range of elements, which approach has the added advantage of removing effects due to channelling.

Figure 3.5(a) shows real space intensity profiles for two 300 keV probes. The FWHM of the probe with an aperture of 10 mrad is approximately 1.0 Å and represents a probe size routinely achievable in a modern STEM machine fitted with an aberration corrector. The probe formed using an aperture of 20 mrad has a FWHM of approximately 0.5 Å and represents the probable probe size we can reasonably hope to attain in future machines.

Typical images are shown in figure 3.5(b) for K -shell ionization of carbon for various probe forming apertures. Results are normalized to a common maximum value of unity for ease of comparison.

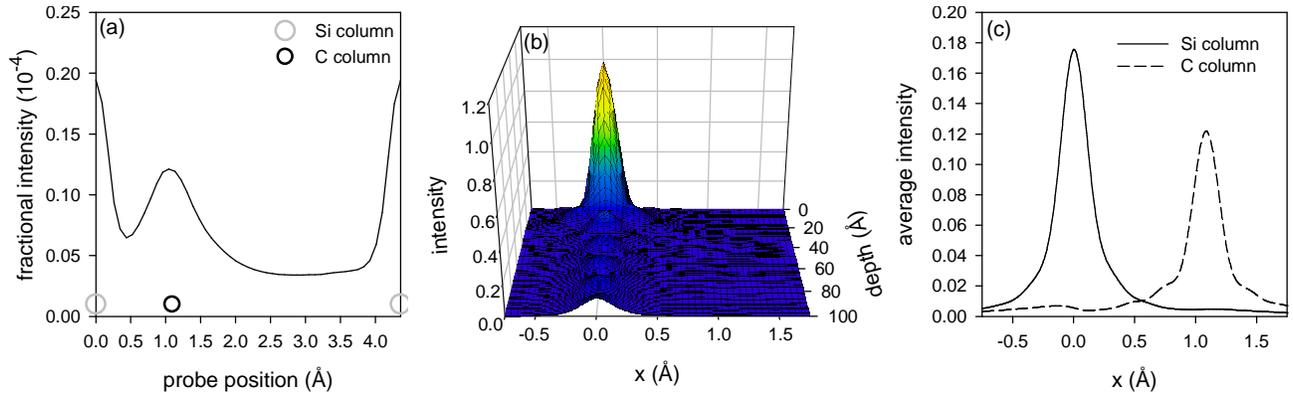


Figure 3.4: (a) Simulated STEM image obtained by scanning along $[100]$ with a 200 keV probe formed with a 50 mrad aperture and incident along the $[011]$ zone axis of a 100 Å thick slab of SiC. Electrons that have ionized a K -shell electron in C are detected in an axial detector with acceptance semi-angle of 10 mrad and an energy window of 40 eV. The result is expressed as a fraction of the incident electron flux. (b) Evolution of the intensity of the incident probe as a function of depth when the probe is positioned above the Si column. (c) The integral along z of the intensity shown in (b) (presented as an average) is compared with the integrated result obtained when the probe is above the C column.

The development of a “volcano” as a function of increasing α is evident. For $\alpha = 15$ mrad the STEM image is sharply peaked. As α is increased to 20 mrad the top of the image flattens out, leading to a small increase in the FWHM. By the time $\alpha = 30$ mrad the volcano-like structure is well defined and the FWHM is reduced. The FWHM of a STEM EELS image is dependent not only on the probe size, but also on the size of the EELS detector, the dependence of $f_{h,g}$ on collection angle being explicit in equation (3.20). To investigate this dependence, the FWHM for carbon and calcium K -shell ionization has been plotted as a function of both the probe forming aperture size and the detector collection angle in figures 3.5(c) and 3.5(d) respectively.

Two interesting features can be seen in the carbon FWHM plot shown in figure 3.5(c). The first is the diagonal ridge highlighted by the thick black line. This ridge represents the point at which volcano-like structures form in the STEM images. These structures occur on the large α side of the ridge. As β increases in size, electrons scattered through larger angles are detected. This leads to a more “localized” interaction and we expect a correspondingly reduced FWHM for the STEM image. For small values of α this effect is clear. In the case of the carbon K shell this is somewhat distorted by the volcano-formation ridge. However for $\alpha = 40$ mrad, all images have volcano-like structure, and there is a small but obvious increase in the FWHM with increasing collection angle.

For the more tightly bound K -shell electrons in calcium (with less volcano formation) the increase in the FWHM with increasing β for large α is not evident. We shall now explore the role of detector size more fully.

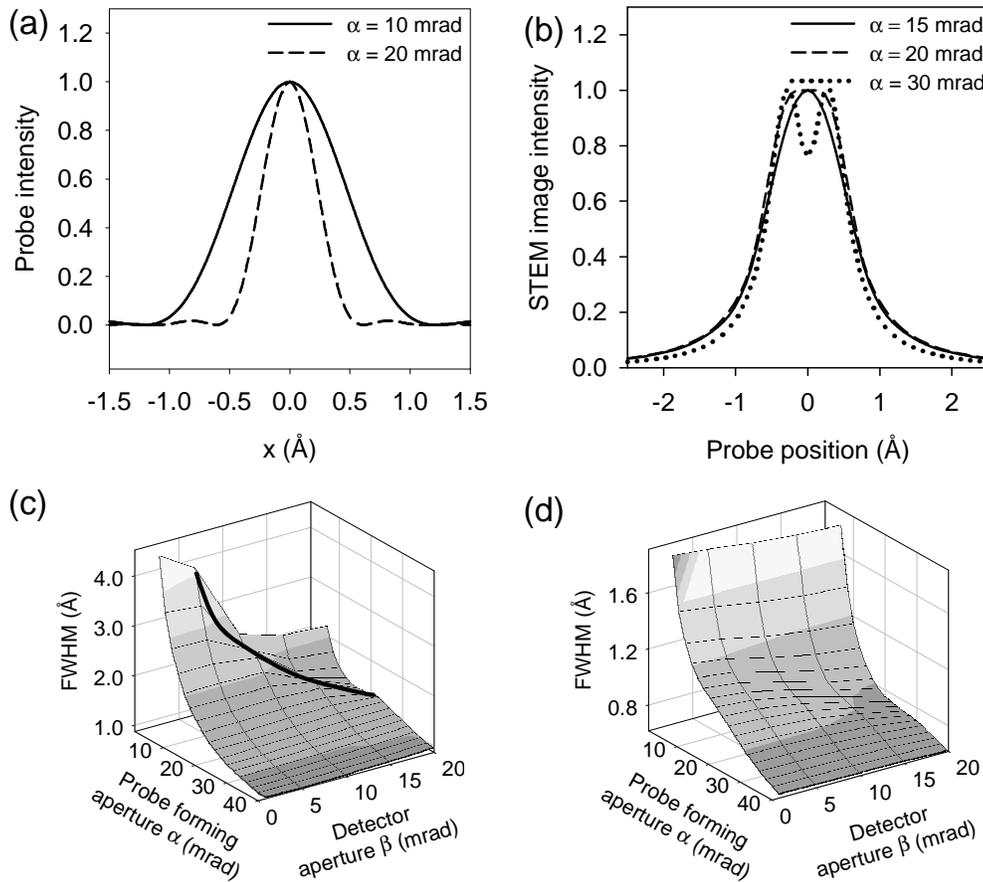


Figure 3.5: (a) Aberration free STEM probe intensities for 300 keV probes with probe-forming aperture semiangles of 10 and 20 mrad, normalized to unity in each case for ease of comparison. (b) STEM images for the K -shell ionization of a single carbon atom showing the appearance of a “volcano” for the larger probe forming aperture with an EELS detector semi-angle of 20 mrad. (c) Variation of the FWHM of STEM images with probe forming aperture and EELS detector semi-angle for the K -shell ionization of a carbon atom. (d) As (c) but for calcium.

3.7 Nonlocality and the importance of detector size

Let us restate equation (1.49) here for convenience:

$$W(\mathbf{r}_\perp, \mathbf{r}'_\perp) \approx \frac{2\pi m}{\hbar^2 t} \sum_{n \neq 0} \frac{1}{k_n} \sum_{\alpha} H_{\alpha, n0}^*(\mathbf{r}_\perp) H_{\alpha, n0}(\mathbf{r}'_\perp) \int e^{2\pi i \mathbf{K}'_\perp \cdot (\mathbf{r}_\perp - \mathbf{r}'_\perp)} d^2 \mathbf{K}'_\perp. \quad (3.22)$$

The local approximation would require that equation (3.22) be approximately $2V'(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}')$. This delta-function type behaviour could arise in equation (3.22) in two ways. First, if $H_{n0}(\mathbf{r})$ has a single, highly localized peak, then the product $H_{0n}(\mathbf{r})H_{n0}(\mathbf{r}')$ may be delta-function like. Second, the integral over the detector conditions could lead to delta-function like behaviour from cancellations in the integration over the terms $e^{2\pi i \mathbf{K}'_\perp \cdot (\mathbf{r}_\perp - \mathbf{r}'_\perp)}$. These possibilities can, of course, work in tandem.

Figure 3.6 shows z -projected transition matrix elements $H_{n0}(r_\perp)$ [69] from (a) the Si K -shell and (b) the C K -shell to continuum final states with energy 5 eV above the ionization threshold and defined

by final angular momentum quantum numbers $l' = 0$, $l' = 1$ and $l' = 2$, with $m_l' = 0$ in each case. These states have partly been chosen because they have axial symmetry, though states with larger values of l' are unimportant in this case [69]. The localization of $H_{n0}(r_\perp)$ changes notably for the different final states. In particular the transitions to the final state $l' = 1$ and $m_l' = 0$, a dipole favoured transition, show significant extent away from the atomic site. The extent of nonlocality of the term $\sum_n H_{0n}(\mathbf{r})H_{n0}(\mathbf{r}')$ will depend on that of the individual terms, and provided that some significant contributions $H_{n0}(\mathbf{r})$ are delocalized we may expect $\sum_n H_{0n}(\mathbf{r})H_{n0}(\mathbf{r}')$ to have an appreciable nonlocal extent. From Figure 3.6 we note that each interaction potential for the C K -shell is more delocalized than its counterpart for the Si K -shell, which is consistent with the differences in binding of the initial state. Though this intuitive correlation between orbital size and locality is not always maintained [66], it holds here and as such we shall use the carbon K -shell in the depictions of the extent of nonlocality in the remainder of this section.

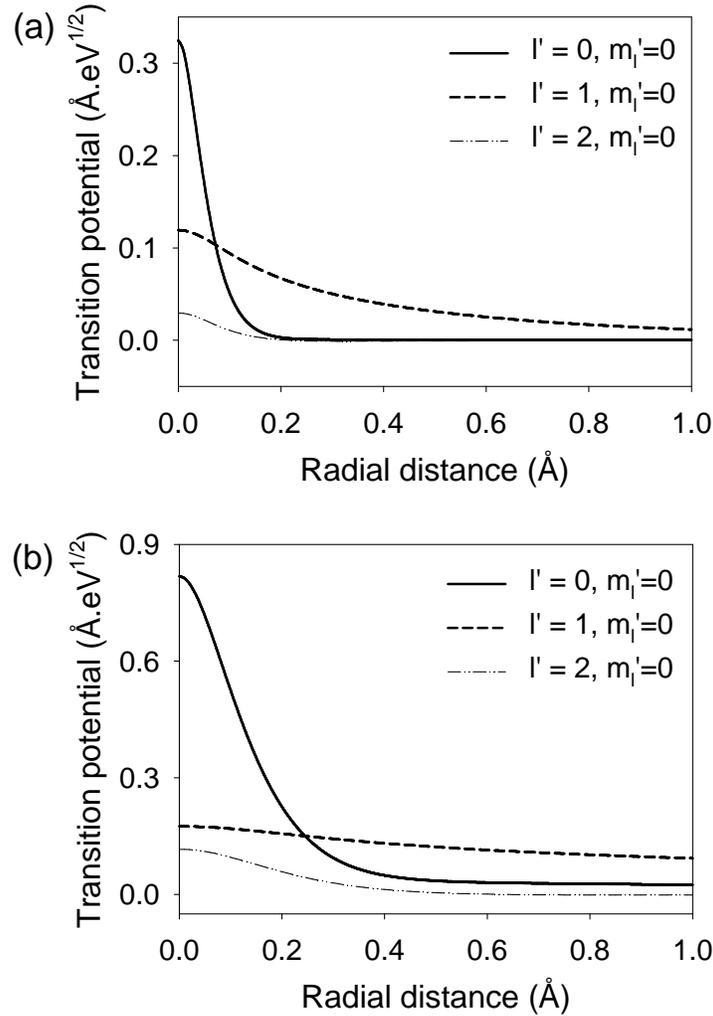


Figure 3.6: Transition matrix elements $H_{n0}(r_\perp)$ (i.e. z -projected) for (a) the Si K -shell and (b) the C K -shell to continuum final states with energy 5 eV above the ionization threshold and defined by final angular momentum quantum numbers listed in the legends.

Consider now the role of the detector. As the detector size increases, the range of the integral over the plane becomes larger. In the limit that the detector size tends to infinity (recall that we have

made the flat Ewald sphere approximation)

$$\int e^{2\pi i \mathbf{K}'_{\perp} \cdot (\mathbf{r}_{\perp} - \mathbf{r}'_{\perp})} d\mathbf{K}_{\perp} \rightarrow \delta(\mathbf{r}_{\perp} - \mathbf{r}'_{\perp}). \quad (3.23)$$

This limit cannot be exactly reached since clearly $|\mathbf{K}'_{\perp}| > |\mathbf{K}'|$ is unphysical. We demonstrate below that the regime in which it is sufficiently accurate to validate the local approximation can be practically achieved.

We present a visual demonstration and measure of this using the technique described by Oxley *et al.* [35]. Assume that the potential varies only along a single direction, such as is assumed in a systematic row orientation. Further assume that the probe varies in only one direction (this assumption does not have a physical analogue in conventional microscope geometry but constitutes only a small sacrifice in generality given the visual insight it provides). With these assumptions we write an analogue of equation (3.1) as

$$\sigma(R_x) \propto \int_0^t \int \int \phi_0^*(R_x, x, z) W(x, x') \phi_0(R_x, x', z) dx dx' dz. \quad (3.24)$$

Figures 3.7(a)-(e) show the effective potential $W(x, x')$ for K -shell ionization of a single carbon atom, assuming an incident energy of 200 keV, a detector energy window of 40 eV above the ionization threshold and detector semi-angles 10 mrad, 20 mrad, 30 mrad, 40 mrad and 50 mrad respectively. Fig. 3.7 (f) assumes an EDX type integration over the full solid angle though again only a 40 eV energy window. It is seen that as the detector aperture angle increases the significant values in $W(x, x')$ tend towards the $x = x'$ diagonal. The local approximation is justified in the limit that $W(x, x') \rightarrow 2V'(x)\delta(x - x')$, and increasing detector angle corresponds to approaching this delta-function type behaviour, to “becoming more local”.

If we are to refer to the effective potential becoming “more local” as the detector size increases and the effective potential collapses towards the $x = x'$ diagonal, what ought we compare relative degrees of nonlocality to? The cross-section expression in equation (3.18) provides the answer, since the integral is over the product of $W(x, x')$ with $\phi_0^*(R_x, x, z)\phi_0(R_x, x', z)$. If the extent of $\phi_0^*(R_x, x, z)\phi_0(R_x, x', z)$ is broad with respect to that of $W(x, x')$ then the integral may be less sensitive to the precise extent of $W(x, x')$. In particular, if we can replace $W(x, x')$ with a function lying on the diagonal and still get a good approximation to the overall integral then the local approximation is valid. (How to best choose such a function is also an issue and we shall return to this question presently.) When the extent of the $\phi_0^*(R_x, x, z)\phi_0(R_x, x', z)$ is less than that of $W(x, x')$ then the integral may be particularly sensitive to variations in $\phi_0^*(R_x, x, z)\phi_0(R_x, x', z)$ with both x and x' , the relative phases for instance, and as such the local approximation will likely be poor, the full nonlocal model being required instead.

Figure 3.6(g), (h) and (i) show $\Re[\phi_0^*(R_x, x, z)\phi_0(R_x, x', z)]$ (the imaginary part integrates to zero) for the probe distribution on the surface assuming a 10 mrad, 20 mrad and 30 mrad probe-forming aperture semi-angle respectively, for a single position of the probe (above the atom). Shifting the probe corresponds to translating this figure along the $x = x'$ diagonal. Using the spread off the $x = x'$ diagonal as a guide, we can say that the potentials in figure 3.7(c) and (d), 30 mrad and 50 mrad detector semi-angles respectively, are reasonably local in comparison with the extent of the probe products in figure 3.7 (e) and (f). Conversely, the extent of the potentials in figure 3.7 (a) and (b) off the diagonal is commensurate with that of probe products. Thus we conclude that what constitutes a large detector for the purposes of justifying a local approximation for the interpretation of a given

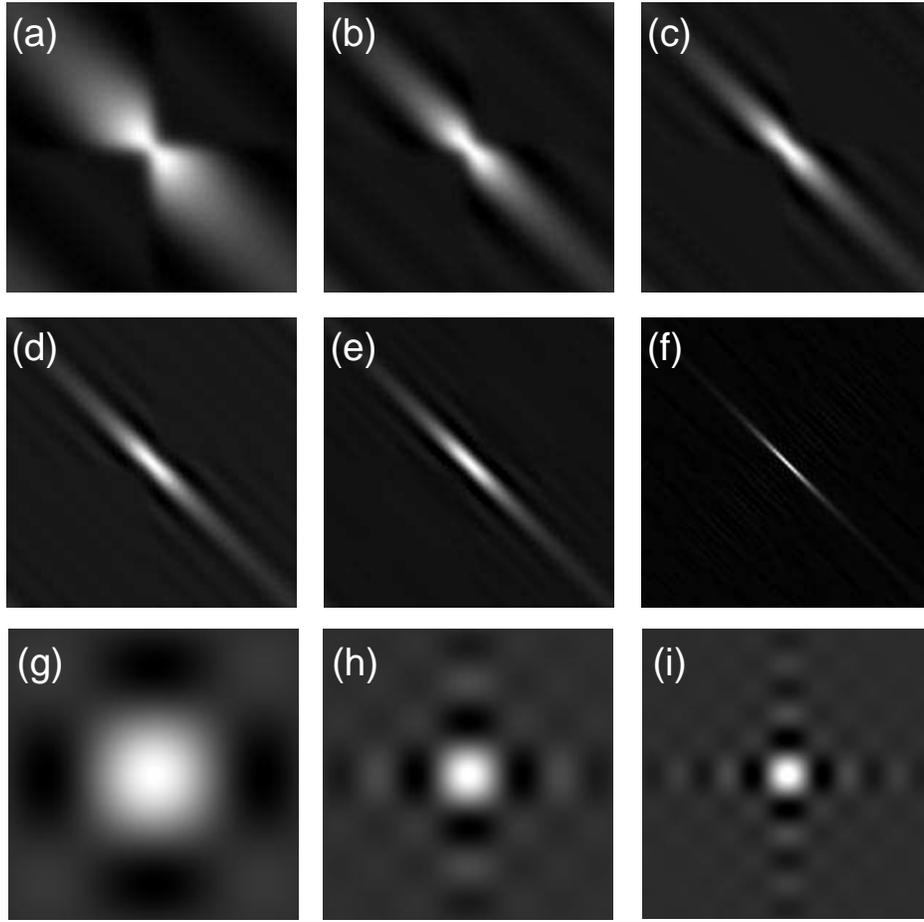


Figure 3.7: Nonlocal effective potentials $W(x, x')$, assuming a “systematic row” condition, for C K -shell ionization of a single atom with a 40 eV energy window above threshold and a detector aperture semi-angle of: (a) 10 mrad, (b) 20 mrad, (c) 30 mrad, (d) 40 mrad, (e) 50 mrad, and (f) for the full solid angle. Probe product $\Re[\phi_0(x)\phi_0^*(x')]$ for a “one dimensional” probe with probe-forming semi-angle: (g) 10 mrad, (h) 20 mrad, and (i) 30 mrad. A 200 keV probe is assumed. Each cell displayed is 5 Å on each side, with the atom in (a)-(f) located half way along the axes.

experiment must be defined relative to the probe size used. This is consistent with previous findings [65,68].

To put this on a more quantitative footing and to get a feel for when the local approximation is adequate, let us consider the interaction in a single plane, like the single atom calculations. We assume an aberration-free probe with probe-forming aperture p_{\max} (in \AA^{-1} , the semi-angle in radians being obtainable via $\alpha \approx p_{\max}/k_0$). We consider some simplified forms for the inelastic component $\sum_n H_{0n}(\mathbf{r})H_{n0}(\mathbf{r}')$ in equation (3.15). We describe a completely delocalized interaction by setting $\sum_n H_{0n}(\mathbf{r})H_{n0}(\mathbf{r}') = 1$. In this case the signal is independent of probe position and the ratio between the nonlocal and local intensities can (with some effort) be determined analytically: $\sigma_{\text{nonloc}}/\sigma_{\text{loc}} = [\min(p_{\max}, K_{\max})/p_{\max}]^2$. Thus if the detector aperture is equal to or larger than that of the probe, the local approximation agrees exactly with the nonlocal model. If the detector aperture is smaller than the probe forming aperture then the nonlocal signal is smaller than the local approximation predicts. This is shown graphically in figure 3.8(a).

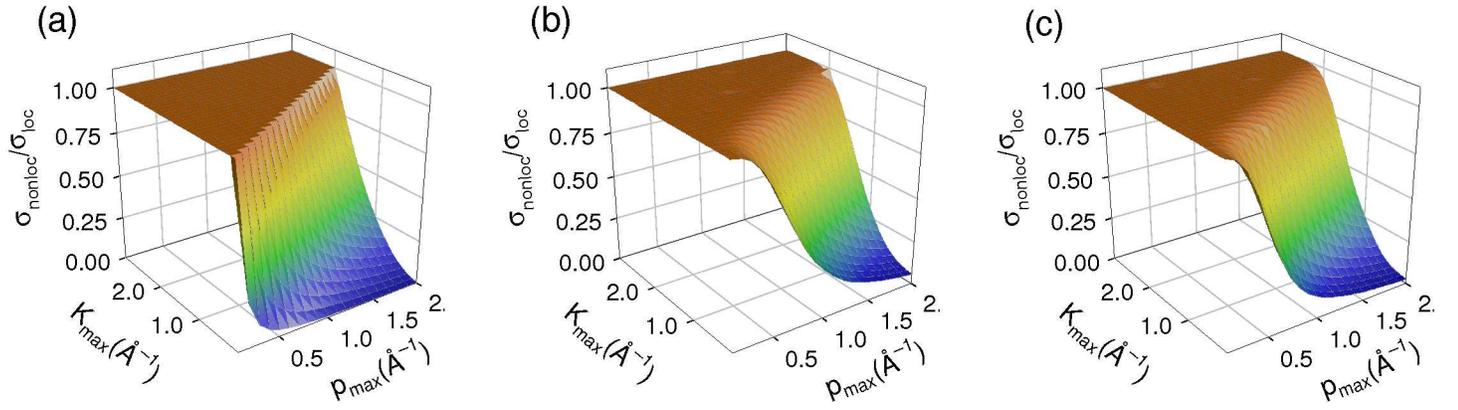


Figure 3.8: Ratio of signal in the nonlocal model to that in the local approximation as a function of the size of the probe-forming aperture, p_{max} , and the detector aperture, K_{max} , for: (a) a completely delocalized interaction, $\sum_n H_{0n}(\mathbf{r})H_{n0}(\mathbf{r}') = 1$; (b) $\sum_n H_{0n}(\mathbf{r})H_{n0}(\mathbf{r}') = e^{-r^2/2\rho^2}e^{-r'^2/2\rho^2}$ with $\rho = 0.5$ Å and the probe situated above the potential centre; (c) $\sum_n H_{0n}(\mathbf{r})H_{n0}(\mathbf{r}') = e^{-r^2/2\rho^2}e^{-r'^2/2\rho^2}$ with $\rho = 0.5$ Å and the probe situated 0.59 Å away from the potential centre.

Another simple ansatz for the interaction is to assume a single final state with a Gaussian interaction potential: $\sum_n H_{0n}(\mathbf{r})H_{n0}(\mathbf{r}') = e^{-r^2/2\rho^2}e^{-r'^2/2\rho^2}$. Setting $\rho = 0.5$ Å, figure 3.8(b) plots $\sigma_{\text{nonloc}}/\sigma_{\text{loc}}$ with the probe situated above the centre of the Gaussian interaction potential, and figure 3.8(c) plots $\sigma_{\text{nonloc}}/\sigma_{\text{loc}}$ with the probe displaced by 0.59 Å from the centre of the Gaussian interaction potential, a distance which corresponds to the half-width-at-half-maximum of $e^{-r^2/2\rho^2}$. Thus we conclude that, as a rule of thumb, that the detector aperture size should be about equal to or above that of the probe aperture. This is still only a guide, since these simple simulations do not involve a probe with any phase variation and do not account for any scattering in the specimen (the creation via scattering in the specimen of wave function components with higher transverse momenta components than in the initial probe might be thought of as an effective increase in the probe-forming aperture size for the purposes of our rule of thumb).

3.8 Simulations of STEM images

In this section we suggest a number of simulations that can be done in the Workshop using the computer program STEM Playpen.

SrTiO₃: Fig. 3.1 shows simulations for ADF, EELS and EDX images for ZnS. Using STEM Playpen we can make similar calculations for SrTiO₃. For an incident energy of 100 keV the relevant file containing the crystal data is

SrTiO3_001_100.xtl.

As also indicated in the file name, the images will be simulated for a $\langle 001 \rangle$ zone axis. Let us firstly consider ADF imaging. There are two ways to simulate images, namely the Bloch wave and multislice approaches. Using the Bloch wave method calculate ADF images (both 2D and a line scan) for a few thicknesses. Repeat the line scan simulation (2D will take too long) using the multislice method for your smallest thickness (once again to save time). You can do similar calculations for energy-loss signals from the Ti *L*-shell using the inelastic scattering coefficients in the *.eel files provided (these are computationally expensive and have therefore been calculated earlier). An energy window of 40 eV was assumed for the detector energy window. The file

SrTiO3_001_100_TiL1(2s)_nonlocal.eel

contains the inelastic scattering coefficients calculated in the nonlocal model for ionization from the 2s orbital (binding energy 590 eV). The file

SrTiO3_001_100_TiL23(2p)_nonlocal.eel

contains the inelastic scattering coefficients for the 2p orbital (binding energy 486 eV). Some of the SrTiO₃ calculations can be repeated at a higher energy (300 keV) using the files

SrTiO3_001_300.xtl and SrTiO3_001_300_TiL23(2p)_nonlocal.eel.

SiC: There are files given for SiC which will allow you to reproduce the results in Fig. 3.4(a) for the C *K*-shell (binding energy 309 eV), namely

SiC_011.xtl and SiC_011_CK_10mrad_40ev_nonlocal.eel.

You will need the vectors $(11\bar{1})$ and $(\bar{1}\bar{1}1)$ to generate your supercell. Note that if you want to repeat this calculation using the multislice approach you should use the files

SiC_011_ms.xtl and SiC_011_CK_10mrad_40ev_nonlocal_ms.eel

to obtain a converged result (kinematically forbidden beams which are handled internally in the Bloch wave code need to be included). There are also *.eel files for SiC which will allow you to investigate the validity of the local approximation as a function of detector size:

SiC_011_CK_10mrad_40ev_local.eel,

SiC_011_CK_30mrad_40ev_local.eel and SiC_011_CK_30mrad_40ev_nonlocal.eel,

SiC_011_CK_50mrad_40ev_local.eel and SiC_011_CK_50mrad_40ev_nonlocal.eel.

Further calculations: There are several other parameters that you can vary in the above calculations:

- Dependence on probe forming aperture
- Dependence on defocus Δf
- Dependence on third-order spherical aberration C_s .

Have fun!

Chapter 4

Overview

The interaction between the fast probing electrons and the specimen in electron microscopy is a quantum many body problem. By expanding in a basis of eigenstates of the target specimen the many body Schrödinger equation can be reduced to a set of coupled equations for the elastic wave function ψ_0 and the wave functions ψ_n which represent inelastic scattering exciting the sample from its ground state to the n 'th excited state. The coupled channel equations were collapsed to give an effective single particle Schrödinger equation for the ground state, though in the process a nonlocal effective potential was introduced. Using conservation of electrons and the assumption that inelastic scattering from different atoms is incoherent, a cross-section expression was derived from which contributions to various inelastic signals could be computed from a knowledge of the elastic wave function and an appropriate effective scattering potential. For high energy electrons, the paraxial approximation can be made. The projected potential approximation was also made, slicewise for a sample which is significantly nonperiodic along the direction of the electron beam.

For a perfect crystal the periodicity is such that the projected potential of one cell is that of the whole crystal. Using the lateral periodicity in such a case the Bloch wave theory was developed. It was shown that the Bloch wave method can be applied to STEM as well as to the case of plane wave incidence by using different excitation amplitudes appropriate to the different boundary conditions. The cross-section expression was evaluated using the Bloch wave ansatz for the wave function. This approach allows for fast and efficient calculation of the cross section for crystalline specimens. The multislice method can also be used for the evaluation of the cross section, and is particularly useful in application to samples where a lack of periodicity makes the Bloch wave method unsuitable.

Forms for the inelastic scattering coefficients, a Fourier space representation of the effective scattering potentials used in the cross-section expression, were given for thermal scattering (applicable to ADF imaging) and inner shell ionization (appropriate to EELS and EDX). Examples were presented exploring the dynamics of the simulations and examining questions of practical importance. The spreading of the probe via dynamical scattering and the possibility of cross talk, the role of thermal scattering, the counter-intuitive results which may follow from the nonlocal effective scattering potentials were considered. EELS imaging from core-loss events may fall in this last category, and the extent of the effective nonlocality is primarily governed by the detector geometry. Experiments may thus be designed to enable interpretation based on a local model by suitable choice of detector size: as a rough guide, the detector semi-angle should be comparable to, indeed preferably larger than, the probe-forming semi-angle.

Chapter 5

Appendix - sign conventions

Normal QM convention

Free space plane wave propagating in the direction of \mathbf{k} (time independent wave eq.):

$$\psi(\mathbf{r}) = \exp(+2\pi i \mathbf{k} \cdot \mathbf{r})$$

and $|\mathbf{k}| = 1/\lambda$ where λ is the wavelength.

Fourier transform from real space to reciprocal space:

$$\mathcal{F}[f(\mathbf{r})] \equiv F(\mathbf{q}) = \int f(\mathbf{r}) \exp(-2\pi i \mathbf{q} \cdot \mathbf{r}) d\mathbf{r}$$

Fourier transform from reciprocal space to real space:

$$\mathcal{F}^{-1}[F(\mathbf{q})] \equiv f(\mathbf{r}) = \int F(\mathbf{q}) \exp(+2\pi i \mathbf{q} \cdot \mathbf{r}) d\mathbf{q}$$

Transmission function:

$$q(\mathbf{r}_\perp) = \exp[+i\sigma\phi(\mathbf{r}_\perp)]$$

where the projected potential $\phi(\mathbf{r}_\perp) = \int \phi(\mathbf{r}_\perp, z) dz$ is positive and attracts electrons.

Propagation function:

$$\exp[-i\pi\lambda\Delta z q^2]$$

Contrast transfer function:

$$T(\mathbf{q}) = A(\mathbf{q}) \exp[-i\pi\lambda q^2(\Delta f + C_s\lambda^2 q^2/2)]$$

where $\Delta f > 0$ for overfocus.

Alternative convention

Free space plane wave propagating in the direction of \mathbf{k} (time independent wave eq.):

$$\psi(\mathbf{r}) = \exp(-2\pi i \mathbf{k} \cdot \mathbf{r})$$

and $|\mathbf{k}| = 1/\lambda$ where λ is the wavelength.

Fourier transform from real space to reciprocal space:

$$\mathcal{F}[f(\mathbf{r})] \equiv F(\mathbf{q}) = \int f(\mathbf{r}) \exp(+2\pi i \mathbf{q} \cdot \mathbf{r}) d\mathbf{r}$$

Fourier transform from reciprocal space to real space:

$$\mathcal{F}^{-1}[F(\mathbf{q})] \equiv f(\mathbf{r}) = \int F(\mathbf{q}) \exp(-2\pi i \mathbf{q} \cdot \mathbf{r}) d\mathbf{q}$$

Transmission function:

$$q(\mathbf{r}_\perp) = \exp[-i\sigma\phi(\mathbf{r}_\perp)]$$

where the projected potential $\phi(\mathbf{r}_\perp) = \int \phi(\mathbf{r}_\perp, z) dz$ is positive and attracts electrons.

Propagation function:

$$\exp[+i\pi\lambda\Delta z q^2]$$

Contrast transfer function:

$$T(\mathbf{q}) = A(\mathbf{q}) \exp[+i\pi\lambda q^2(\Delta f + C_s\lambda^2 q^2/2)]$$

where $\Delta f > 0$ for overfocus.

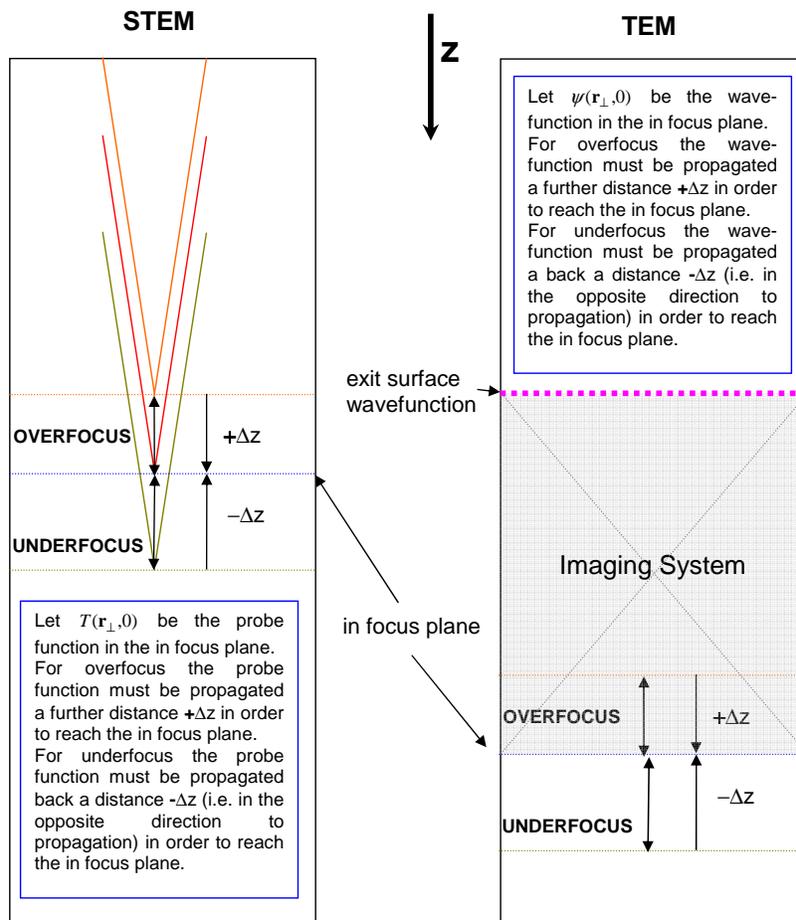


Figure 5.1: Comparison of defocus Δf and propagation Δz . If *overfocus* is positive then positive Δf corresponds to positive Δz , i.e. Overfocus corresponds to additional propagation of the wave function in the z -direction.

Notes:

- 1) The correct signs can be found noting that a plane wave in the normal QM convention increases in phase as it moves in the direction of \mathbf{k} along the positive z -direction (the converse is true for the alternative convention). Transmission functions, propagation functions and contrast transfer functions (CTFs) must be consistent with this. This is illustrated above for defocus term in the CTF. Note that C_s is an “effective overfocus”. For the free space propagator a beam scattered in the direction \mathbf{g} undergoes less phase change than the forward scattered beam along the z -direction.
- 2) P. G. Self *et al.* [70], who addressed matters similar to those in this appendix, should have the Fourier transform and structure factor definitions swapped between the columns.
- 3) The follow up paper by W. O. Saxton *et al.* [71] has the following errors: In Table 1 in the transmission function a ϕ is missing in the first column (before the subscript) and a negative sign is missing in the exponent in the second column. The wave aberration function in the first column has the wrong sign in front of the term containing C_s . It should be positive.

Bibliography

- [1] H. Yoshioka, *J. Phys. Soc. Jpn.* 12 (1957) 618.
- [2] Y. H. Ohtsuki, *Charged Beam Interactions with Solids*, Taylor & Francis, London (1983).
- [3] K. Kambe and K. Molière, in *Advances in Structure Research by Diffraction Methods*, edited by R. Brill and R. Mason, Pergamon, Oxford (1970) Vol. 3, p. 53.
- [4] C. J. Humphreys and M. J. Whelan, *Philos. Mag.* 20 (1969) 165.
- [5] Z. L. Wang, *Acta Crystallogr.* A46 (1990) 366.
- [6] T. Fujikawa and L. Hedin, *Phys. Rev. B* 40 (1989) 11507.
- [7] Y. H. Ohtsuki and S. Yanagawa, *J. Phys. Soc. Jpn.* 21 (1996) 326.
- [8] P. H. Dederichs, *Phys. Kondens. Mater.* 5 (1966) 347.
- [9] P. Rez, *Acta Crystallogr.* A34 (1978) 48.
- [10] T. W. Josefsson and A. E. Smith, *Phys. Rev. B* 50 (1994) 7322.
- [11] E. Merzbacher, *Quantum Mechanics, Third Edition*, John Wiley & Sons, New York (1998).
- [12] L. J. Allen and C. J. Rossouw, *Phys. Rev. B* 39 (1989) 8313.
- [13] L. J. Allen, T. W. Josefsson, *Phys. Rev. B* 53 (1996) 11285.
- [14] C. J. Humphreys, *Rep. Prog. Phys.* 42 (1979) 1825.
- [15] D. M. Bird and Q. A. King, *Acta Cryst.* A46 (1990) 202.
- [16] L. J. Allen, H. M. L. Faulkner and H. Leeb, *Acta Crystallogr.* A56 (2000) 119.
- [17] L. J. Allen, H. Leeb and A. E. C. Spargo, *Acta Crystallogr.* A55 (1999) 105.
- [18] L. J. Allen, T. W. Josefsson and H. Leeb, *Acta Crystallogr.* A54 (1998) 388.
- [19] L. J. Allen and T. W. Josefsson, *Phys. Rev. B* 52 (1995) 3184.
- [20] J. C. H. Spence and J. M. Cowley, *Optik* 50 (1978) 129.
- [21] A. R. Wilson and A. E. C. Spargo, *Philos. Mag.* A 46 (1982) 435.
- [22] S. D. Findlay, L. J. Allen, M. P. Oxley and C. J. Rossouw, *Ultramicroscopy* 96 (2003) 65.

- [23] S. J. Pennycook and D. K. Jesson, *Ultramicroscopy* 37 (1990) 14.
- [24] P. D. Nellist and S. J. Pennycook, *Ultramicroscopy* 78 (1999) 111.
- [25] P. D. Nellist and S. J. Pennycook, *Advances in Imaging and Electron Physics* 113 (2000) 147.
- [26] T. Yamazaki, K. Watanabe, A. Recnik, M. Ceh, M. Kawasaki and M. Shiojiri, *J. Electron Microsc.* 49 (2000) 753.
- [27] J. Broeckx, M. Op de Beeck and D. Van Dyck, *Ultramicroscopy* 60 (1995) 71.
- [28] S. D. Findlay, *Theoretical aspects of scanning transmission electron microscopy* (Doctoral thesis), The University of Melbourne (2005). May be downloaded from <http://eprints.unimelb.edu.au/archive/00001057/>
- [29] B. F. Buxton, J. E. Loveluck and J. W. Steeds, *Philos. Mag. A* 38 (1978) 259.
- [30] L. J. Allen, S. D. Findlay, M. P. Oxley and C. J. Rossouw, *Ultramicroscopy* 96 (2003) 47.
- [31] G. R. Anstis, D. Q. Cai and D. J. H. Cockayne, *Ultramicroscopy* 94 (2003) 309.
- [32] M. P. Oxley, L. J. Allen and C. J. Rossouw, *Ultramicroscopy* 80 (1999) 109.
- [33] S. D. Findlay, M. P. Oxley, S. J. Pennycook and L. J. Allen, *Ultramicroscopy* 104 (2005) 126.
- [34] J. C. H. Spence, *Mat. Sci. Eng.* 26 (1999) 1-49.
- [35] M. P. Oxley, E. C. Cosgriff and L. J. Allen, *Phys. Rev. Lett.* 94 (2005) 203906.
- [36] H. Kohl and H. Rose, *Adv. Electron. Electron Phys.* 65 (1985) 173.
- [37] D. A. Muller and J. Silcox, *Ultramicroscopy* 59 (1995) 195.
- [38] P. Schattschneider, M. Nelhiebel and B. Jouffrey, *Phys. Rev. B* 59 (1999) 10959.
- [39] P. Schattschneider, M. Nelhiebel and H. Souchay, B. Jouffrey, *Micron* 31 (2000) 333.
- [40] L. J. Allen, C. J. Rossouw, *Phys. Rev. B* 42 (1990) 11644.
- [41] D. Cherns, A. Howie and M. H. Jacobs, *Zeitschrift für Naturforsch. A*28 (1973) 565.
- [42] V. W. Maslen and C. J. Rossouw, *Philos. Mag.* A47 (1983) 119.
- [43] V. W. Maslen, C. J. Rossouw, *Philos. Mag.* A49 (1984) 735-742.
- [44] C. J. Rossouw, V. W. Maslen, *Philos. Mag.* A49 (1984) 743-758.
- [45] M. P. Oxley and L. J. Allen, *Acta Cryst.* A56 (2000) 470.
- [46] M. P. Oxley and L. J. Allen, *Phys. Rev. B* 57 (1998) 3273.
- [47] M. P. Oxley and L. J. Allen, *Acta Cryst.* A57 (2001) 713.
- [48] D. K. Saldin and P. Rez, *Phil. Mag.* B55 (1987) 481.

- [49] J. S. Reid, *Acta Cryst.* A39 (1983) 1.
- [50] P. M. Fields and J. M. Cowley, *Acta Cryst.* A34 (1978) 103.
- [51] J. M. Cowley and P. M. Fields, *Acta Cryst.* A35 (1979) 28.
- [52] T. Yamazaki, M. Kawasaki, K. Watanabe, I. Hashimoto and M. Shiojiri, *J. Electron Microscopy* 50 (2001) 517.
- [53] A. J. McGibbon, S. J. Pennycook and D. E. Jesson, *J. Microscopy* 195 (1999) 44.
- [54] K. Watanabe, Y. Kotaka, N. Nakanishi, T. Yamazaki, I. Hashimoto and M. Shiojiri, *Ultramicroscopy* 92 (2002) 191.
- [55] T. W. Josefsson, L. J. Allen, *Phys. Rev. B* 53 (1996) 2277-2285.
- [56] S. Hillyard, R. F. Loane and J. Silcox, *Ultramicroscopy* 49 (1993) 14.
- [57] T. Plamann and M. J. Hytch, *Ultramicroscopy* 78 (1999) 153.
- [58] C. Dwyer and J. Etheridge, *Ultramicroscopy* 96 (2003) 343.
- [59] K. Ishizuka, *J. Electron Microscopy* 50 (2001) 291.
- [60] K. Ishizuka, *Ultramicroscopy* 90 (2002) 71.
- [61] C. J. Rossouw, L. J. Allen, S. D. Findlay and M. P. Oxley, *Ultramicroscopy* 96 (2003) 299.
- [62] L. J. Allen and C. J. Rossouw, *Phys. Rev. B* 42 (1990) 11644.
- [63] R. F. Loane, P. Xu and J. Silcox, *Acta Crystallogr. A* 47 (1991) 267.
- [64] R. F. Loane, P. Xu and J. Silcox, *Ultramicroscopy* 40 (1992) 121.
- [65] L. J. Allen, S. D. Findlay, M. P. Oxley, C. Witte and N. J. Zaluzec, *Ultramicroscopy*, in press (2006).
- [66] M. P. Oxley, M. Varela, T. J. Pennycook, K. van Benthem, S. D. Findlay, L. J. Allen and S. J. Pennycook, submitted for publication (2006).
- [67] E. J. Kirkland, *Advanced Computing in Electron Microscopy*, Plenum Press, New York (1998).
- [68] E. C. Cosgriff, M. P. Oxley, L. J. Allen and S. J. Pennycook, *Ultramicroscopy* **102**, 317 (2005).
- [69] C. Dwyer, *Ultramicroscopy* 104 (2005) 141.
- [70] P. G. Self, M. A. O'Keefe, P. R. Buseck and A. E. C. Spargo, *Ultramicroscopy* **11**, 35 (1983).
- [71] W. O. Saxton, M. A. O'Keefe, D. J. H. Cockayne and M. Wilkens, *Ultramicroscopy* **12**, 75 (1983).