Ultramicroscopy 47 (1992) 266–281
North-Holland

ultramicroscopy

# Ultimate resolution and information in electron microscopy: general principles

D Van Dyck
*University of Antwerp (RUCA), Groenenborgerlaan 171, B-2020 Antwerpen, Belgium*

and

A F de Jong
*Philips Research Laboratories, P O Box 80000, NL-5600 JA Eindhoven, The Netherlands*

Received at Editorial Office 29 June 1992

A new concept of resolution is introduced based on the idea that the electron microscope can be considered as a communication channel that transfers information from the object to the observer The channel consists of three subchannels in series (i) the transfer through the object, (ii) the transfer through the microscope and (iii) the recording of the image For each of these subchannels the transfer function and resolution are discussed and from this the total information capacity of the whole channel can be estimated

Two different regimes of resolution exist If the resolution is insufficient to discriminate the individual atoms of the object in projection, the necessary information exceeds the capacity of the channel so that a priori knowledge is needed from other techniques and only limited new information can be obtained with high-resolution electron microscopy (HREM) If the instrumental resolution is sufficient to discriminate the individual atoms, all atom positions can in principle be determined with relatively high precision and without a priori knowledge provided the information can be retrieved directly from the images With the recent developments, the instrumental resolution approaches 0 1 nm which is close to the ultimate resolution which is limited by the object rather than by the instrument

## 1. Introduction

The microstructure of matter can be studied by interaction with particles, which carry information from the object to the observer For this purpose, electrons are extremely useful since, as compared to other particles, they are easy to generate, easy to accelerate, easy to deflect and easy to detect

One can either use *internal* electrons which are extracted (tunnelled) from the object by an external voltage or *external* electrons which are accelerated and scattered by the object

With these electrons one can form images in two ways in *sequential* imaging, the object is scanned with a probe and in *parallel* imaging the image is formed as a whole

In combination four ways of electron imaging are possible In practice the four methods exist and provide structural information about a material, at comparable resolution (see table 1)

In field ion microscopy (FIM) and scanning tunnelling microscopy (STM) electrons are tunnelled from the surface of the object and hence only provide information from the surface area These techniques will not be discussed here

Information from the bulk of the object can be obtained by scattering with high-energy electrons ($\geq 100$ keV) as is the case in high-resolution electron microscopy (HREM) and scanning transmission electron microscopy (STEM)

In an ideal scattering experiment, the state of the electron is carefully determined immediately before and after the interaction with the object

Table 1
Survey of different interaction and imaging modes using electrons

| | Internal electrons | External electrons |
|---|---|---|
| sequential imaging | STM | STEM |
| parallel imaging | FIM | HREM |

(i e in the planes A and B in fig 1) From the change in the electron state one can deduce information about the interaction and hence about the object itself The plane A characterizes the illumination condition and the plane B the detecting condition The states of the electron can be determined in either real space or reciprocal space The following extreme situations can now be envisaged

parallel illumination delta function in the reciprocal plane of A,
focused illumination delta function in the real plane of A,
diffraction mode detection in the reciprocal plane of B,
imaging mode detection in the real plane of B

Full flexibility for converting the electron states in the planes A and B from real to reciprocal space or vice versa can be obtained by placing two lenses (or two lens combinations), one at each side of the object, i e the condensor and objective lens (fig 1) All existing combinations of illumination and detection are listed in table 2 Due to reciprocity (symmetry of time reversal) HREM and STEM are equivalent In a sense both techniques have the same configuration (fig 1) if the $z$-axis is inverted

The electron microscope can thus be described as a communication channel in which each elec-

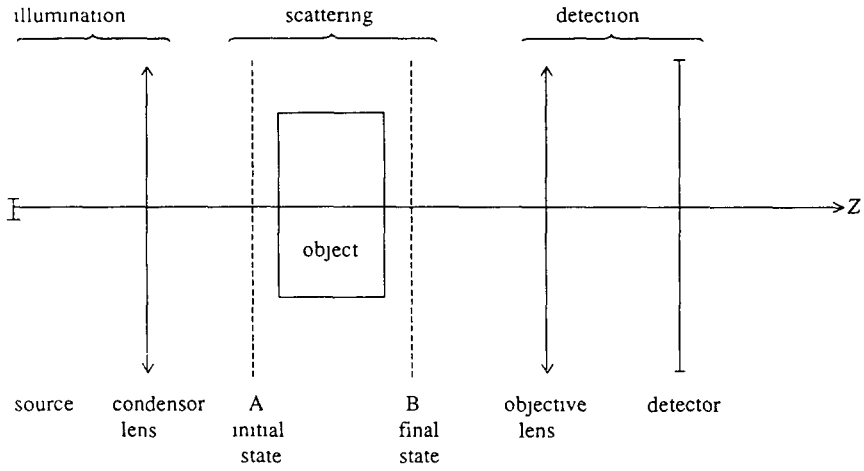


Fig 1 Schematic representation of an ideal scattering experiment

Table 2
Survey of different illumination and detection modes

| Illumination | Detection | Technique |
|---|---|---|
| Parallel | Image | High-resolution electron microscopy (HREM) |
| Parallel | Diffraction | Diffraction |
| Focused | Image | CBIM [1] [a] |
| Focused fixed probe | Diffraction | CBED |
| Focused scanned probe | Diffraction (selected) | STEM |
| Focused scanned probe | Diffraction | Phtychography [2] |

[a] Convergent beam imaging

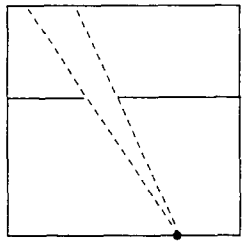

Fig 2 Projection box

tron carries two degrees of spatial information In ptychography, 2D diffraction information is obtained while scanning the probe over the object so that a total 4D information is obtained, which can be used to retrieve the structure of the object [2] Due to inelastic scattering, the energy of the electron also carries information about the energy states of the object These spectroscopic information channels can in principle be separated using an energy filter

An ideal electron microscope, in which all illumination detection and filter modes are possible, should contain a field emission source, a symmetric twin-type condensor-objective lens of high quality, an energy filter before and after the object and a CCD detector, all operated under full computer control

In the following we focus attention on elastic high-resolution electron microscopy and especially on the information content of the images

## 2. Image formation

### 2 1 *Basic principles*

Let us first consider, as an illustrative example, the simplest imaging device the camera obscura

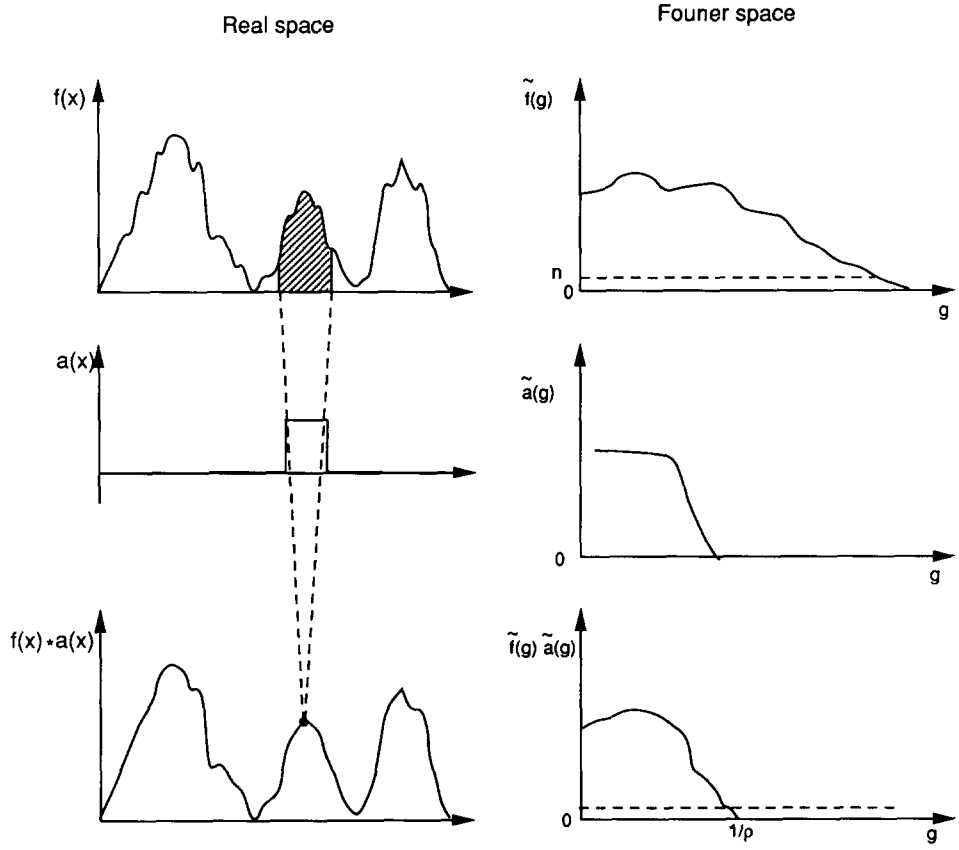


Fig 3 Schematical representation of the image formation in a projection box in real space (left) and reciprocal space (right)

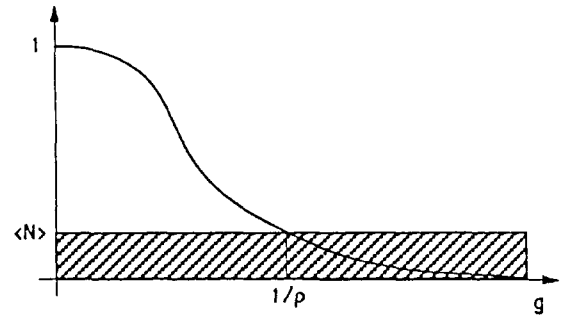


Fig 4 Transfer function

This is a black box with a pinhole (fig 2) Each part of the one-dimensional object, represented by the function $f(x)$, is transmitted through the pinhole (aperture) to the image (for simplicity we take the function and the camera to be one-dimensional) Calling $a(x)$ the aperture function, which is equal to one in the aperture and zero elsewhere, then the image is obtained as

$$f_{\mathrm{im}}(x) = \int a(x'-x) f(x')\, \mathrm{d}x' \qquad (1)$$

which is the definition of a convolution product

$$f_{\mathrm{im}}(x) = a(x) * f(x) \qquad (2)$$

In Fourier space, the Fourier transform of the convolution product yields a direct product of the Fourier transforms of the corresponding functions, i e

$$\tilde{f}_{\mathrm{im}}(g) = \tilde{a}(g)\ \tilde{f}(g) \qquad (3)$$

with $g$ the spatial frequency This is illustrated in the right-hand side of fig 3 $\tilde{a}(g)$ is usually called the (modulation) transfer function of MTF of the imaging device It is shown in more detail in fig 4

Every imaging device can be characterized by its transfer function (band filter), which describes the magnitude with which information of a spatial frequency $g$ is transferred through the device $N$ is the noise

## 2 2 Resolution

Usually, the resolution of the instrument $\rho$ is defined from the cut-off $1/\rho$ between signal and noise beyond which no spatial information is transferred This is the type of resolution in the sense defined by Rayleigh [3] The Fourier transform of the transfer function to real space is usually called the impulse response function (IRF) It is the generalisation of the aperture function of the camera obscura It is a sharply peaked function which represents the image of a point object The width of the IRF is also related to the Rayleigh resolution The sharper the IRF, the better the resolution If the transfer function is known, the original image can be restored up to the resolution $\rho$ by dividing by $\tilde{a}(g)$ This is called image restoration or deblurring

If a communication channel consists of a series of subchannels, the total transfer function is the product of the transfer functions of the subchannels

## 2 3 Resolution and information theory

According to the theory of Shannon [4] the maximal information rate of a communication channel is given by

$$C = B \log_2(1 + S/N) \text{ bits/s}, \qquad (4)$$

where $B$ is the bandwidth of the channel, $S$ is the average signal power and $N$ the noise power at the output of the channel

Eq (4) can be applied to microscopy in the following way The bandwidth of a channel is defined as the number of independent degrees of freedom that the channel can transmit per unit time In microscopy the bandwidth can be defined as the number of independent degrees of freedom that the microscope can transmit per unit area Consider, for example, a two-dimensional square unit cell of size $a \times a$ The reciprocal space (i e the diffraction pattern) consists of delta functions (reflections) at the positions of the reciprocal lattice with mesh $1/a \times 1/a$

Each reciprocal node contains one independent degree of freedom (In practice, each node contains two (real and imaginary) numbers which, for a real object, are symmetry-related by Friedel's law ) The total number of degrees of freedom,

transmitted within the band filter up to the resolution $\rho$ of the microscope, is then equal to

$$\frac{\pi(1/\rho)^2}{(1/a)^2} = \frac{\pi a^2}{(\rho)^2} \tag{5}$$

The number of degrees of freedom per unit area is then

$$B = \pi/(\rho)^2, \tag{6}$$

which is independent of the choice of $a$

Eq (4) can now be interpreted as follows The microscope transmits about three independent degrees of freedom per unit $\rho^2$ If we consider the noise level $N$ as the smallest significant piece of information, the signal + noise can be expressed in units $N$ as $(S+N)/N$, which in binary code equals $\log_2(1+S/N)$ bits Hence, each degree of freedom carries, on the average, $\log_2(1+S/N)$ bits of information The "image" is thus only to be considered as an intermediate data plane that carries the information Strictly speaking one is not interested in the image as such but rather in its information content

## 3. Electron microscope as communication channel

An electron microscope can be considered as a series of three communication channels which act successively on the incident electrons (1) interaction electron–object, (2) transfer in the microscope, (3) recording of the image

In fact, the channels (1) and (2) act on the wavefunction of the electrons so that the transfer functions are complex functions with an amplitude and a phase component We will now discuss each of the different communication channels separately and their effect on the total information transfer

### *3 1 Interaction electron–object*

#### *3 1 1 Two-dimensional object*

As is well known, a thin object acts as a phase object which multiplies the wavefunction of the incident electrons with a phase factor

$$e^{i\sigma V(\boldsymbol{R})} \tag{7}$$

$V(\boldsymbol{R})$ is the projected electrostatic potential of the object, $\boldsymbol{R}$ is a two-dimensional vector perpendicular to the incident beam and $\sigma$ is a proportionality constant Eq (7) can thus be considered as the transfer function, which, in this case, only affects the phase If the object is very thin and consists of light atoms, $V(\boldsymbol{R})$ is small and (7) can be expanded

$$e^{i\sigma V(\boldsymbol{R})} \approx 1 + i\sigma V(\boldsymbol{R}) \tag{8}$$

The real part of the transfer function is then constant (unity) and the imaginary part is proportional to the projected electrostatic potential

Each object consists of atoms, so in fact $V(\boldsymbol{R})$ is the superposition of the electrostatic potential of the individual atoms In this way the incident plane wave is modified so as to carry the information about the structure of the object

In the case of one atom, the projected potential $V(\boldsymbol{R})$ is a 2D Gaussian-like function The atom thus serves as a channel, the impulse response function of which is the projected potential of that atom The transfer function is then the Fourier transform of $V(\boldsymbol{R})$, which is also Gaussian-like in the sense of fig 3, as schematically shown in fig 9 (top), and which in scattering theory corresponds with the scattering factor $f(g)$

The resolution of this one-atom channel is then related to the width of $V(\boldsymbol{R})$ and is typically of the order of 0 05 to 0 1 nm This means that, within the atom, no smaller relevant details are present A further resolution-limiting effect is the thermal fluctuation of the atom position This can be accounted for by convoluting $V(\boldsymbol{R})$ with the probability distribution function of the atom positions, which usually is Gaussian-like In this way, the scattering factor $f(g)$ (a transfer function) is multiplied with a Gaussian function, which further limits the resolution, especially at high temperatures This is the well known Debye–Waller factor

It is important to note that for each known type of atom, the potential $V(\boldsymbol{R})$ is known Hence,

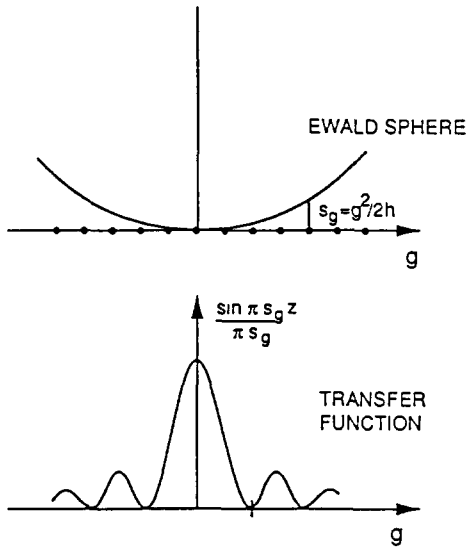


Fig 5 Ewald sphere and corresponding transfer function (TEM)

the only information that one needs to deduce is the position of the atom (in projection) In this way, each atom has only two parameters to be determined (two degrees of freedom)

*3 1 2 Three-dimensional object kinematical approximation*

The resolution is limited by the diffraction of the electron in the object This can be seen as follows consider the kinematical approximation for the amplitude of the beam $g$ diffracted at a crystalline object

$$a_g = f(g)\frac{\sin \pi s_g z}{\pi s_g}, \tag{9}$$

with $f(g)$ the scattering factor (structure factor), $z$ the thickness of the object and $s_g$ the excitation error, i e the distance between the reciprocal node $g$ and the Ewald sphere measured along $z$ (fig 5) In a zone-axis orientation, the excitation error is given by

$$s_g = g^2/2k, \tag{10}$$

with $k$ the wavenumber of the incident electron

From eq (9) it is clear that the amplitudes of the outermost beams are dampened with increasing distance from the Ewald sphere resulting in an effective transfer function which is schematised in fig 5 This is a consequence of the fact that $g$ cannot obey at the same time the law of energy conservation (Ewald sphere) and the projection approximation (zone plane) It is thus a conflict between the scattering in the 3D object and the 2D imaging which limits the resolution An estimation can be obtained from the first zero of eq (9) given by

$$s_g z = 1, \tag{11}$$

from which the resolution is

$$\rho = \frac{1}{g} = \left(\frac{z}{2k}\right)^{1/2} \tag{12}$$

In fact eq (11) expresses Heisenberg's relation, which couples the uncertainty in the position of the scattering, i e the object thickness $z$, with the uncertainty in the wavevector $s_g$ The same result (12) can be derived from STEM in a crude way as follows in order to make a spot size of dimension $x$, the apex angle of the incident beam cone has to be such that, from Heisenberg,

$$x\Delta = 1, \tag{13}$$

with $\Delta$ the spread on the incident wavevector (fig 6) The crossover of the projection of the cone over the object thickness can be estimated as

$$\frac{z\alpha}{2} = \frac{z\Delta}{2k} \tag{14}$$

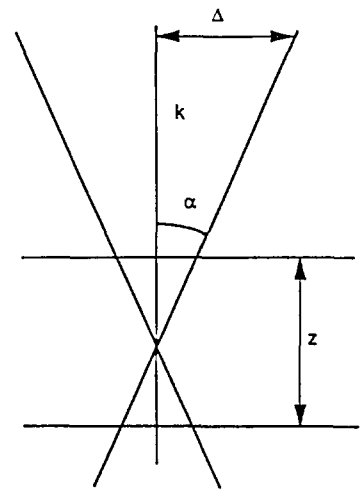


Fig 6 Resolution-limiting effects in STEM

The final resolution is a compromise between the spot size and the crossover which can roughly be estimated as

$$\rho \approx \sqrt{\left(\frac{1}{2\Delta}\right)^2 + \left(\frac{z\Delta}{2k}\right)^2} \tag{15}$$

The optimal resolution can be obtained for

$$\Delta^2 = k/z, \tag{16}$$

from which

$$\rho = \left(\frac{z}{2k}\right)^{1/2}, \tag{17}$$

which is essentially the same result as (12), as could be expected, from the reciprocity between TEM and STEM The physical reason behind (12) and (17) is also that for 2D imaging, the ideal object should be 2D The influence of the third dimension of the object deteriorates the resolution

For a given foil thickness $z$, the resolution can be improved by increasing $k$, i e by increasing the accelerating voltage However, the maximum useful voltage is limited to the threshold energy for displacement radiation damage [#1] which can be estimated as follows the maximum transferred energy $T$ in a head-on collision with a particle of mass $M$ is related to the wavevector $k$ of the incident particle by the (relativistic) expression

$$k = \left[MT/2\hbar^2\right]^{1/2} \tag{18}$$

The displacement threshold energy is somewhat larger than the binding energy of the atom in the crystal and can be expressed as

$$E = \beta E_0, \tag{19}$$

with

$$E_0 = e^2/4\pi\epsilon_0 a \approx 27 \text{ eV} \tag{20}$$

the electrostatic binding energy of two elementary charges separated by the Bohr radius (i e twice the ionisation energy of the hydrogen atom) $a = \hbar^2\epsilon_0/\pi e^2 m_e = 0\,529 \times 10^{-10}$ m, $\beta$ is a dimensionless constant (cf Madelung constant) which for most materials (ionic compounds, metals, alloys, semiconductors) is situated between 0 5 and 1 [6] Requiring now that the transferred energy $T$ equals the threshold $E$ one finally obtains, using (17), (18) and (20) [7],

$$\rho_D = Ca, \quad \text{with } C = 0\,32 z^{1/2}/[M\beta]^{1/4}, \tag{21}$$

where the foil thickness $z$ is expressed in units of the Bohr radius and $M$ is expressed in atomic mass units

$$1 \text{ AMU} = 1\,825 m_e \tag{22}$$

For most experimental situations the resolution in the Rayleigh sense is limited to the order of 0 1 nm and increases with the square root of the crystal thickness

It is interesting to notice that these results are independent of the type of scattering particle and apply equally well for protons, etc

*3 1 3 Electron channelling*

If the crystal object is perfectly oriented along a zone axis, the incident electrons are trapped in the positive potential of the columns The columns then, in a sense, act as channels for the electrons [8,9] If the distance between the electrons is not too small, a one-to-one correspondence between the the wavefunction at the exit face and the column structure of the crystal is established Within the columns, the electrons oscillate as a function of depth without, however, leaving the column (fig 7) Hence the classical picture of electrons traversing the crystal as plane-like waves in the direction of the Bragg beams, which historically stems from X-ray diffraction, is in fact misleading It is important to note that channelling is not a property of a crystal, but occurs even in an isolated column and is not much affected by the neighbouring columns provided the distance is not too close

The channelling can best be understood in real space as follows [9] Assuming normal incidence and taking the $z$ axis perpendicular to the specimen foil, the high-energy equation describing the dynamical electron scattering in real space is equivalent to the time-dependent Schrodinger equation [18]

$$-\frac{\hbar}{1}\frac{\partial\phi}{\partial t}(\boldsymbol{R}, t) = H\phi(\boldsymbol{R}, t), \tag{23}$$

[#1] The radiation damage due to ionisation decreases with increasing voltage

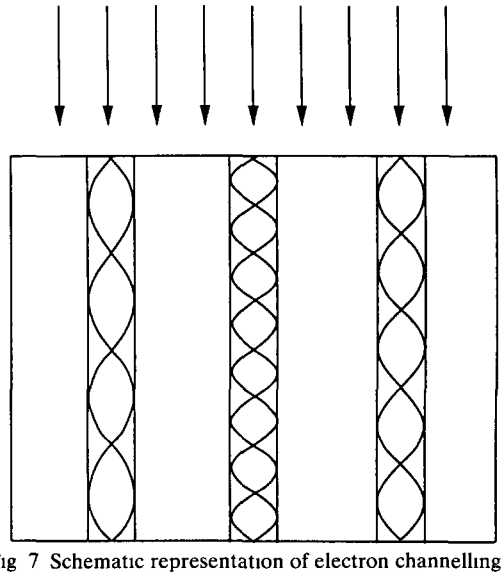

Fig 7 Schematic representation of electron channelling

in which the time is replaced by the depth $z$ using $t = mz/hk$ and in which the Hamiltonian is given by

$$H = -\frac{\hbar^2}{2m}\Delta - eU(\boldsymbol{R}, t), \tag{24}$$

with $U(\boldsymbol{R}, t)$ the electrostatic crystal potential, $m$ and $h$ the relativistic electron mass and wavevector This can be understood by assuming that in the direction of propagation ($z$ axis) the high-energy electron behaves as a classical particle with a constant velocity equal to $hk/m$ In this way the $z$ axis plays the role of a time axis We will from now on use $t$ instead of $z$

It is easy to verify that the solution of (23) which obeys the boundary condition $\phi(\boldsymbol{R}, 0)$ is now given by

$$\phi(\boldsymbol{R}, t) = 1 + \sum_n C_n \phi_n(\boldsymbol{R}) \left[\exp\left(-\frac{1}{\hbar}E_n t\right) - 1\right], \tag{25}$$

with $\phi_n(\boldsymbol{R}) = 1$ the bounded eigenstates of the Hamiltonian and $E_n$ its energy ($E_n < 0$) obeying

$$H\phi_n(\boldsymbol{R}) = E_n \phi_n(\boldsymbol{R}) \tag{26}$$

A somewhat more general expression for (25) can be found in ref [9] If the atoms are not too heavy and the accelerating potential is not too high only one bound state appears, so that

$$\phi(\boldsymbol{R}, t) = 1 + C\phi(\boldsymbol{R}) \exp\left(\frac{-\mathrm{i}Et}{\hbar} - 1\right) \tag{27}$$

From this it is clear that the electron wavefunction varies perfectly periodically with depth, the periodicity being determined by $E$, which is related to the mass of the column From (27) it is clear that $\phi(\boldsymbol{R})$ represents a kind of impulse response function for that particular column Its Fourier transform can then be considered as the maximum scattering factor for that column The scattering factor varies periodically between zero and this maximum This effect is known as "dynamical extinction" In a sense, the resolution limited by the object then also varies periodically with depth The best resolution is obtained for those values for which (27) becomes maximal However, the variation is different for different types of columns The optimal resolution can be estimated from the width of $\phi(\boldsymbol{R})$ which is of the same order of magnitude as the width of the projected $U(\boldsymbol{R})$ and hence of the atom The width of $\phi(\boldsymbol{R})$, i e the resolution, increases with increasing projected potential of the column (which is proportional to its "weight") and with increasing accelerating voltage but is always of the order of 0 1 nm

### *3 1 4 Inelastic scattering*

The electron can be scattered inelastically by an object, either by exciting an atom or molecule, a phonon or a plasmon or by emitting a photon (bremsstrahlung) Inelastic scattering destroys the coherence with respect to the incident electrons so that in fact each different inelastic scattering event can be considered as a different parallel communication channel These channels can only be disentangled by energy filtering However, the present energy filters usually do not combine high-energy resolution (order of 1 eV) with high spatial resolution and are thus not yet used in HREM The inelastically scattered electrons then mainly contribute to the noise

Here inelastic scattering also puts fundamental limits on the resolution that can be obtained This has already been stated by De Broglie and Gabor [16] Along the same lines of thought we will try to estimate a limit for the resolution

From classical scattering theory, the energy transfer from the incident electron, with mass $m$ to an atom with mass $M$ ($M > m$) is given by

$$\Delta E = (4m/M)E \sin^2(\theta/2), \tag{28}$$

with

$$E = h^2k^2/2m \tag{29}$$

the energy of the incident electron and $k$ its wavevector, $\theta$ is the diffraction angle, which is related to the resolution by

$$\theta = 1/k\rho \tag{30}$$

In principle the atom is not free but bounded between its neighbours Hence the smallest possible energy $\Delta E$ that can be transferred to the atom is given to the vibrational energy Considering the atom motion as anharmonic oscillation and estimating the force constant from the limiting energy of the atom (19) one obtains

$$\Delta E = \gamma E_0 \sqrt{m/M}, \tag{31}$$

with $E_0$ given by (20) $\gamma$ is a configuration constant, typically of the order of 0 1 For instance, for an atom of average mass $\Delta E$ is of the order of 0 01 eV Substitution of (20), (29), (30) and (31) into (28) then finally yields

$$\rho \approx 0\ 1 Z^{-1/4}\ \text{nm}, \tag{32}$$

with $Z$ the atom number Beyond the value given by (32), inelastic scattering starts dominating the diffraction process so that (32) can be considered as a rough limit for the ultimate resolution

### 3 2 Transfer in the electron microscope

#### 3 2 1 Phase transfer function

The wavefunction in the image plane of an electron microscope is given (without incoherent aberrations) by

$$\psi = \psi_0 * p, \tag{33}$$

where $\psi_0$ is the wavefunction at the exit face of the object given by (8) or (25) and $t$ is the impulse response function (point spread function) of the electron microscope and is given by the Fourier transform of the transfer function

$$T(\boldsymbol{g}) = \exp[-\imath X(\boldsymbol{g})] \tag{34}$$

with

$$X(\boldsymbol{g}) = \tfrac{1}{2}\pi\left[C_s\lambda^3 g^4 + 2\epsilon\lambda g^2\right] \tag{35}$$

$C_s$ is the spherical aberration constant, $\lambda$ is the wavelength and $\epsilon$ is the defocus

In reciprocal space, the wavefunction (diffraction amplitude) is the Fourier transform of (33)

$$\phi(\boldsymbol{g}) = \phi_0(\boldsymbol{g})\ T(\boldsymbol{g}) \tag{36}$$

#### 3 2 2 Incoherent effects

*(i) Chromatic aberration*

Chromatic aberration results from fluctuations of the focus due to voltage and lens current fluctuation

If the defocus distribution is Gaussian with spread $\Delta$, i e

$$f(\epsilon) = C \exp\left[-\epsilon^2/2\Delta^2\right], \tag{37}$$

it causes, in the coherent approximation (weak object), a damping factor (envelope function) for the transfer function given by

$$E_c(g) = \exp\left[-\pi^2\lambda^2\Delta^2 g^4/2\right] \tag{38}$$

The defocus spread $\Delta$ is related to the voltage spread $\Delta V$ the thermal energy spread $\Delta E$ of the incident electron and the lens current spread $\Delta I$, by the relation

$$\Delta = C_c\left[\frac{(\Delta V^2) + (\Delta E)^2}{V^2} + 4\left(\frac{\Delta I}{I}\right)^2\right]^{1/2}, \tag{39}$$

with $C_c$ the chromatic aberration constant (typically $10^{-3}$ nm) $\Delta V$ is the fluctuation in the incident voltage and $\Delta E$ the thermal energy

spread of the electrons and $\Delta I/I$ is the relative fluctuation of the lens current

*(ii) Beam convergence*

The effect of beam convergence can be introduced as follows Assume a conical incident beam with a Gaussian angular profile

$$\exp\left[-\Theta/2\alpha^2\right], \tag{40}$$

with $\alpha$ the spread on the apex angle

The effect of this convergence results also in an envelope function for the transfer function, given (in the coherent approximation) by

$$E_s(g) = \exp\left\{-\left[2\pi\alpha g\left(C_s\lambda^2 g^2 + \epsilon\right)\right]^2/2\right\} \tag{41}$$

*(iii) Other effects*

The detailed form of these and other damping envelope functions (drift, vibration) will be discussed in ref [10]

*3 2 3 Instrumental resolution*

*General considerations*

In principle the characteristics of an electron microscope can be completely defined by its transfer function, i e by the parameters $C_s$, $\Delta f$, $\Delta$ and $\alpha$ A clear definition of resolution is not easily given for an electron microscope For instance, for thick specimens, there is not necessarily a one-to-one correspondence between the projected structure of the object and the wavefunction at the exit face of the object, so that the image usually does not show a simple relationship

If one wants to determine a "resolution" number, this can only be meaningful for thin objects Furthermore one has to distinguish between *structural resolution* as the finest detail that can be interpreted in terms of the structure and the *information resolution* or *information limit* which is the finest detail that can be resolved by the instrument, irrespective of a possible interpretation The information resolution may be better than the structural resolution With the present electron microscopes, individual atoms cannot yet be resolved within the structural resolution

*Structural resolution (point resolution)*

The electron microscope in the phase-contrast mode at optimum focus directly reveals the projected potential, i e the structure, of the object provided the object is very thin All spatial frequencies $g$ with a nearly constant phase shift are transferred forward from object to image Hence the resolution can be obtained from the first zero of the transfer function (35) as

$$\rho_S = 1/g \approx 0\ 65 C_s^{1/4}\lambda^{3/4} = 0\ 65\ \mathrm{Gl}, \tag{42}$$

with $\mathrm{Gl} = C_s^{1/4}\lambda^{3/4}$ the Glaser unit This value is generally accepted as the standard definition of the structural resolution of an electron microscope It is typically between 0 15 and 0 2 nm for modern instruments and it can be improved by using higher voltages It is often also called the point resolution It is also related to the width of the impulse response function The information beyond $\rho_S$ is transferred with a non-constant phase and, as a consequence, is redistributed over a larger image area

*Information resolution*

The information resolution can be defined as the finest detail that can be resolved by the instrument It corresponds to the maximum diffracted beam angle that is still transmitted with appreciable intensity, i e the transfer function of the microscope is a spatial band filter which cuts all information beyond the information resolution For a thin specimen, this limit is mainly determined by the envelope of chromatic aberration (temporal incoherence) and beam convergence (spatial incoherence) In principle beam convergence can be reduced using a smaller illuminating aperture and a longer exposure time If chromatic aberration is predominant, the damping envelope function is given by (38) from which the resolution can be estimated as

$$\rho_I = \frac{1}{g} = \left(\frac{\pi\lambda\Delta}{2}\right)^{1/2}, \tag{43}$$

with the defocus spread given by (39) For the best microscopes (with a standard electron source) the information resolution is of the order of 0 15 nm

*Ultimate instrumental resolution*

The information between $\rho_S$ and $\rho_I$ is present in the image, albeit with the wrong phase Hence this information is redistributed over the image

However, it can be restored by means of holographic methods In that case $\rho_I$ is the ultimate instrumental resolution Using a field emission gun (FEG) the spatial as well as the temporal incoherence can be reduced so as to push the information resolution towards 0 1 nm A detailed description of the ultimate instrumental resolution limit is given in ref [10] Fig 8 shows the phase transfer function and corresponding impulse response function (IRS) of a modern 300 keV instrument with FEG Here the information limit extends to 0 1 nm but a large amount of information with the wrong phase is present between $\rho_S$ and $\rho_I$, i e in the tails of the IRS, and has to be restored by holographic methods combined with image processing

### 3 3 *Image recording*

In practice, the image is captured by a photoplate or electronic detector, the transfer of which is characterized by a point spread function (in case of a CCD detector the spread is mainly caused by the scintillator (YAG or phosphor) which converts the electrons into photons) [11]

$$p(\boldsymbol{R}) = C \exp(-2R^2D^2), \tag{44}$$

with $D$ the effective pixel size, i e the image intensity is convoluted with $p(\boldsymbol{R})$

$$|\psi|^2 * p \tag{45}$$

Fourier-transformation of (45) then shows that the transfer function should be multiplied with the Fourier transform of $p(\boldsymbol{R})$, i e the modulation transfer function (MTF) of the camera which, from (44), is

$$C \exp(-\pi^2 g^2 D^2/2) \tag{46}$$

The resolution of the recording instrument is not essential since it can be adapted by changing the magnification of the microscope What is more important is the number of pixels Indeed on the one hand the pixel size has to be much smaller than the resolution of the microscope On the other hand the image field has to be large enough to collect all redistributed information from the tails of the impulse response function (fig 8)

In terms of the transfer function (fig 7) the sampling in reciprocal space has to be small enough to sample the rapid oscillations and at the same time the spatial frequency range has to be large enough to gather all information in the oscillating part of the transfer function Since the

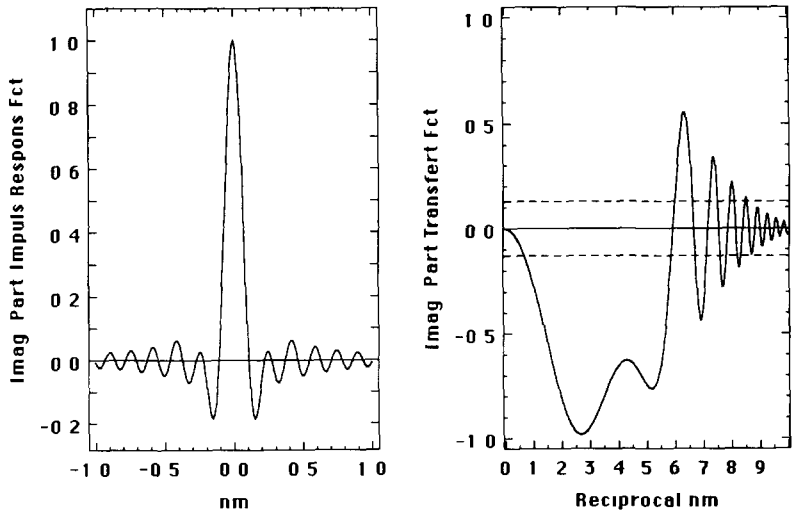


Fig 8 Phase transfer function, left, and corresponding impulse response function, right, for a modern 300 keV instrument ($C_s$ = 0 7 nm, $C_c$ = 1 3 nm, $\Delta E$ = 0 8 eV)

sampling in reciprocal space is the inverse of the image field in real space and the largest spatial frequency is the inverse of the sampling in real space, this puts the same restrictions on the minimal number of sampling points The situation can be improved somewhat by choosing a focus value of the order of $-300$ nm, called the "Lichte" focus [19], for which the oscillations are minimized in the whole frequency range

As a rule of thumb one can state that in this situation the number of pixels $N$ has to be larger than

$$N > (2\rho_S/\rho_I)^4, \tag{47}$$

with $\rho_S$ and $\rho_I$ respectively given by (42) and (43) For $\rho_S = 0\,2$ nm and $\rho_I = 0\,1$ nm one has $N > 256$ which is just within reach with modern CCD cameras For electron holography, where extra fringes have to be sampled, this requirement is strengthened by a factor 3 It should be noted that the maximum number of pixels is not limited by the CCD detector itself but rather by the scintillator (YAG or phosphor) Indeed, in order to obtain sufficient detection efficiency the scintillator should be sufficiently thick (e g 50 $\mu$m) and hence causes a spread on the incident electron which might be comparable to the thickness The CCD on the other hand has no spreading effect on the electron and, in a sense, has no transfer function

It is also interesting to note that this type of recording is also close to its limits Suppose, for instance, that a $(4096)^2$ CCD is within reach and that the resolution of the scintillator can be improved to 25 $\mu$m Then the total field of view is about $10 \times 10$ cm, which covers the whole image area of an electron microscope

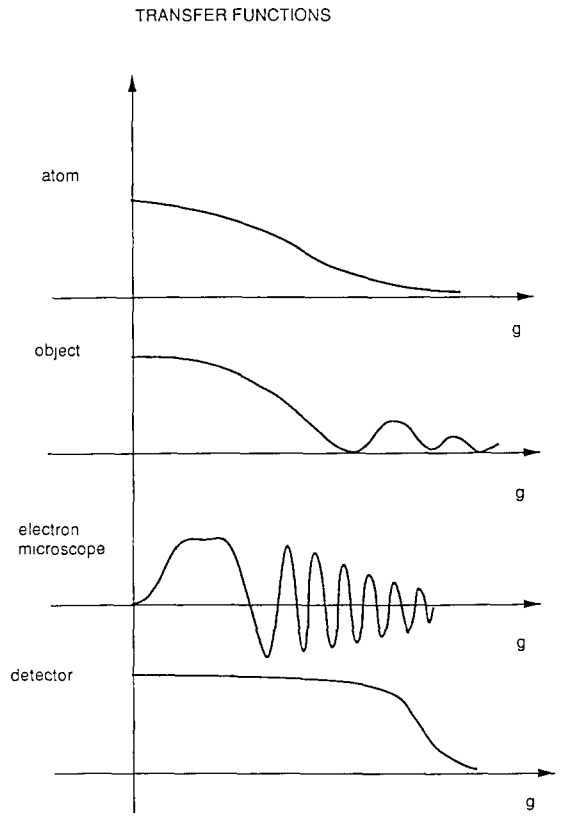


Fig 9 Transfer functions of the different subchannels (schematic)

### 3 4 Transfer of the whole communication channel

#### 3 4 1 Transfer function

As already stated above, the whole transfer function of the electron microscope is the product of the transfer functions of the respective subchannels A schematic representation is given in fig 9 The whole imaging process is schematised in fig 10 The object structure is determined by the atom coordinates This information is spread out through a complex impulse response function Finally the image intensity is recorded

#### 3 4 2 Ultimate resolution

The ultimate resolution is determined by the subchannel with the worst resolution Thus far, the weakest part has been the electron microscope itself

The interpretable resolution $\rho_S$ can be improved by reducing the spherical aberration coefficient $C_s$ and/or by increasing the voltage However, since $C_s$ depends mainly on the pole-piece dimension and the magnetic materials used, not much improvement can be expected Hence, at present, all high-resolution electron microscopes yield comparable values for $C_s$ for comparable situations (voltage, tilt, ) Furthermore, the effect of $C_s$ on the resolution is rather limited In

the far future, a major improvement can be expected by using superconducting lenses

Another way of increasing the resolution is by correcting the third-order spherical aberration by means of a system of quadrupole, hextapole and/or octopole lenses

Increasing the voltage is another way of increasing the resolution However, increasing the voltage also increases the displacive radiation damage of the object (although the ionisation damage is reduced) At present the optimum value, depending on the material, is situated between 200 and 500 keV In our view the tendency in the future will be towards lower rather than towards higher voltages

A much more promising way of increasing the resolution is by restoring the information that is present between $\rho_S$ and $\rho_I$ and that is still present in the image, albeit with the wrong phase For this purpose, image processing will be indispensable In that case, the resolution will be determined by $\rho_I$ $\rho_I$ can be improved drastically by using a field emission gun (FEG) which reduces the spatial as well as the temporal incoherence However, this puts severe demands on the number of pixels in the detector The newest generation of CCD cameras with YAG scintillator and tapered fibres might be the solution to this problem Furthermore, these cameras, when cooled, are able to detect nearly all single electrons Taking all these considerations into account, an instrumental resolution of the electron microscope of 0 1 nm is within reach

The ultimate resolution, however, will be determined by the object itself This resolution can be optimized by using very thin objects although thin films are not always representative for bulk specimens (e g elastic relaxation) The ultimate probe is the atom potential, the width of which is of the order of 0 05 to 0 1 nm

Since resolution is a trade-off between signal and noise, some improvement can still be expected by reducing the noise Thus far only little attention has been paid to this idea The recording noise can be improved by using CCD cameras Specimen noise (inelastic scattering) can be reduced by energy filtering Nevertheless, phonon inelastic scattering with energy transfers of the order of 0 01 eV will not easily be eliminated in the near future However, if we assume that the total transfer function is Gaussian an improvement in the signal-to-noise ratio from 20 to 100 only results in an improvement of the resolution with 25% Considering all possible resolution-limiting effects it can be expected that the ultimate resolution attainable with electron microscopy can hardly be pushed below 0 05 nm

*3 4 3 A new concept of resolution*

We will now introduce a new concept of resolution based on information theory Gabor [5] stated that Shannon's information theory is of

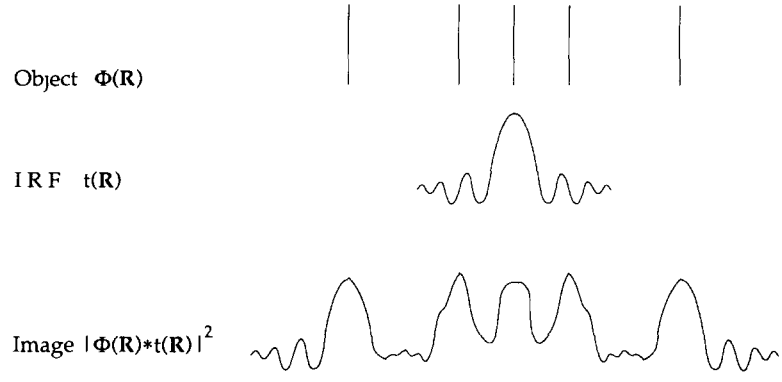


Fig 10 Scheme of the imaging process

little use for electron microscopy, since it was developed for situations where two persons communicate by means of a vocabulary of commonly agreed messages, whereas in electron microscopy, the sender is an object which is usually unknown

As a consequence, electron microscopy can only resolve structures for which sufficient information is obtained from other techniques (e g X-ray diffraction) so that the structure model only contains a small number of unknown parameters, which is much less than the information capacity of the electron microscope Examples are the characterisation of defects (type, Burgers vector), the structure of building-block structures (e g mixed layer compounds) which are determined uniquely by their stacking sequence, the structure of binary alloys ordering on a known lattice, etc Most of these structures can be determined unambiguously even at low resolution On the other hand, it is amazing that most high-resolution images are only interpreted by visual comparison with computer simulations Indeed, this is a very poor technique which allows one only to discriminate between a limited number of plausible structure models and which therefore requires considerable prior information

However, HREM is now able to resolve individual atom columns This is a completely new situation Since all possible atom types are known, a structure can then be characterised completely by the positions of its constituent atoms The atoms can thus be considered as the messages of Shannon and the argument of Gabor does not hold In this way a structure could be completely resolved by HREM without prior knowledge However, the requirement is that the number of unknowns (e g atom coordinates) is less than the capacity of the microscope, i e 3 per unit $(\rho_1)^2$ (section 2 3)

In this way resolution gets a completely new meaning If the structure (in projection) contains fewer than about 1 5 atoms per $(\rho_1)^2$, the position of each atom can in principle be determined with an average precision of $\log_2(1 + S/N)$ bits This opens quite new perspectives comparable to X-ray crystallography where, using comparable information (diffracted beams), the atom positions can be determined with high precision If, on the other hand, the resolution is insufficient to determine the individual atoms, i e the number of atoms exceeds 1 5 per $(\rho_1)^2$, the required information exceeds the capacity of the microscope channel In a sense the channel is then blocked and no information can be obtained without much a priori knowledge

In a real object the first electron "sees" the projected structure of the object Hence it is important to notice that the requirement of fewer than 1 5 atoms per unit $(\rho_1)^2$ has to be fulfilled for the projected object This requirement can most easily be met when studying a crystal along a simple zone axis in which the atoms are aligned along columns parallel to the beam direction However, for more complicated zone axes, the number of atoms in projection increases and the channel may be blocked Also in amorphous objects the number of different atoms in projection increases with depth, so that, except for very thin amorphous objects (a few nm), the information channel is blocked and the images only reveal information about the imaging characteristics of the microscope rather than about the object [12] For amorphous objects the information can be increased using a tilt series (tomography)

Concluding, we propose to define the *resolving capacity* of the electron microscope as the *number of independent degrees of freedom (parameters) that can be determined per unit area* (per Å$^2$ or nm$^2$) (In this way the inconsistency is avoided that exists in the terminology high resolution = small detail ) It should be noted that resolution can also be studied within the framework of catastrophe theory [17] Here also two regimes are considered in which the atom positions can or cannot be determined and which are limited by the errors (noise) in the measurement

In order to determine a structure completely without prior knowledge it is essential that the number of atom coordinates in projection does not exceed the resolving capacity From sections 2 3 and 3 4 the ultimate resolving capacity of electron microscopy is of the order of 5 degrees of freedom per Å$^2$ which allows one to determine the coordinates of about 2–3 atoms per Å$^2$

However, it is equally important that this information can be retrieved from the images in a

direct unambiguous way For this purpose, direct methods are needed Only recently major progress in this field has been achieved This is discussed in section 4

## 4. Direct retrieval of information

A direct method should consist of three stages First the wavefunction in the image plane has to be reconstructed (phase problem) Then the wavefunction at the exit face of the object has to be calculated Then finally from this the structure of the object has to be retrieved The phase problem can be solved in different ways Promising methods are electron holography [14] or focus variation [15] In electron holography, the beam is split by an electrostatic biprism into a reference beam and a beam that traverses the object Interference of both beams in the image plane then yields fringes, the positions of which yield the phase information [14] In order to assess this information one needs a very high-resolution camera (CCD), a powerful image processor, and a field emission gun to provide the necessary spatial coherence The CCD camera needs a sufficient number of pixels so as to gather all the distributed information and must be sufficiently sensitive to detect individual electrons In the focus variation method, the focus is used as a controllable parameter [15] Images are captured at very close focus values so as to collect all information in the three-dimensional image space By Fourier-transforming to 3D reciprocal space the linear information (i e the wavefunction) can be separated from the nonlinear information Once the wave function in the image plane is known, the second step, i e returning to the object, is straightforward and consists of using the inverse phase transfer function

The final step consists of retrieving the projected structure of the object from the wavefunction at the exit face If the object is thin enough to act as a phase object, the phase is proportional to the electrostatic potential of the structure, projected along the beam direction so that the retrieval is straightforward If the object is thicker, the problem is much more complicated However, in a zone-axis orientation the electrons are channelled and the wavefunction at the exit face of the object can be expressed in a simple analytical form and still shows a one-to-one correspondence with the structure, which allows one to retrieve the structure [9]

In order to put this method into practice one needs a medium-voltage high-resolution electron microscope with large-specimen-tilt possibilities, equipped with a field emission gun (FEG) and a high-resolution CCD camera with a high detection quantum efficiency (DQE), directly coupled to a fast image-processing system The microscope should be aligned in an automatic way and all imaging parameters have to be under computer control In principle, for a perfect retrieval, the microscope has not to be aligned perfectly but its imaging parameters ($C_s$, focus, tilt ) have to be known with very high accuracy ($< 1\%$) This is a difficult task Usually the imaging parameters are determined from the images of a known object However, for a full automatic retrieval of the structure of an unknown object, the influence of microscope and object cannot be disentangled and the final retrieval should contain the determination of the optical parameters in a self-consistent way, based on general principles (real potential, atomic structure, ) Recently a European Brite-Euram project has been set up, funded by the European community, in which the ultimate goal is to obtain direct 1 Å structural information using holography and focus variation using a 300 keV instrument The first experimental results are shown in fig 11 for the high-$T_c$ superconductor $YBa_2Cu_4O_8$ obtained with a 200 keV FEG/CCD microscope The results are quite spectacular since the resolution has been lowered from 0 24 nm (point resolution of CM20 Supertwin) to less than 0 15 nm Hence most of the oxygen columns are resolved Furthermore, the FEG allows the use of all illumination angles whereas the CCD collects all electrons either in image space or in diffraction space In the future it would be desirable to equip such an instrument with an energy filter above and below the specimen In this way nearly all information that can be obtained with electrons can be assessed

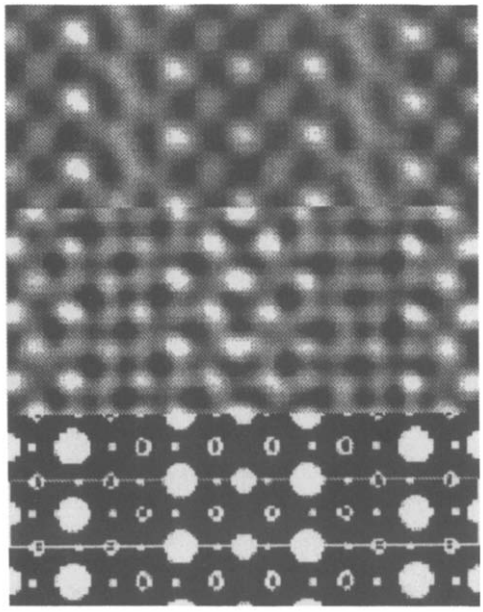

Fig 11 Experimentally retrieved wavefunction at the exit face of the high-$T_c$ superconductor $Y_1Ba_2Cu_4O_8$ From top to bottom high-resolution image at optimum focus, phase of the wavefunction at the object, structure model (Courtesy M Op de Beeck, W Coene [15])

## 5. Conclusion

Recent developments in high-resolution electron microscopy have pushed the resolution, as defined by Rayleigh, down to the 0 1 nm level This value is close to the physical limits of resolution which can be obtained with real materials What is more important, however, is that this resolution just enables one to discriminate individual atoms, the building blocks of nature Hence, it enables us to describe the imaging process in terms of communication theory in which the atoms are as messages and a structure can be uniquely determined by its atom coordinates The electron microscope can then be considered as a communication channel and the concept of resolution can be put into another perspective From this we propose the term "resolving capacity" as the number of independent degrees of freedom that can be determined per unit area However, in order to exploit this information, one needs a direct method that reverses the complete imaging process Recent developments in this direction are very encouraging

## Acknowledgment

The authors wish to thank M Op de Beeck, W Coene and A van den Bos for stimulating discussions and H Lichte for bringing ref [16] to our attention

## References

[1] C J Humphries, D J Eaglesham, D M Maher and H L Fraser, Ultramicroscopy 26 (1988) 13
[2] W Hoppe, Acta Cryst A 25 (1969) 495, 502, 508, J Rodenburg, in EUREM '92, Granada, 1992
[3] Lord Rayleigh, Phil Mag 8 (1879) 261, 403, 477, 9 (1880) 40
[4] C E Shannon, Proc IRE 37 (1949) 10, E H Linfoot, J Opt Soc Am 45 (1955) 808
[5] D Gabor, Lab Invest 14 (1965) 2
[6] D S Billington and J H Crawford, Radiation Damage in Solids (Princeton, NJ, 1961)
[7] D Van Dyck, in EUREM '88, York, Inst Phys Conf Ser 93, Vol 2 (1988) 349
[8] J Lindhard, K Dan Vidensk Selsk Mat Fys Medd 34 (1965) 1, G Lempfuhl, Z Naturforsch A 27 (1972) 425
[9] D Van Dyck, in Proc 12th Int Congr for Electron Microscopy, Seattle, 1990) (San Francisco Press, San Francisco, 1990) p 64, in Electron Diffraction Techniques, Ed J Cowley (Oxford University Press, Oxford, 1992)
[10] A F de Jong and D Van Dyck, Ultramicroscopy, in press
[11] K K Hermann, in Proc 12th Int Congr for Electron Microscopy, Seattle, 1990 (San Francisco Press, San Francisco, 1990) p 112
[12] G Y Fan and J M Cowley, Ultramicroscopy 21 (1987) 125
[13] H Lichte, Adv Opt Electron Microsc 12 (1991) 25, Ultramicroscopy 20 (1986) 293
[14] D Van Dyck, in Proc 12th Int Congr for Electron Microscopy, Seattle, 1990 (San Francisco Press, San Francisco, 1990) p 24
[15] M Op de Beeck, W Coene and D Van Dyck, Ultramicroscopy, in preparation
[16] D Gabor, Rev Mod Phys 28 No 3 (1956) 210
[17] A Van den Bos, J Opt Soc Am 4 (1987) 1402
[18] D Van Dyck, in Adv Electron Electron Phys 65 (1985) 295
[19] H Lichte, Ultramicroscopy 38 (1991) 13