The Ewald sphere/focus gradient does not limit the resolution of cryoEM reconstructions

J. Bernard Heymann

National Cryo-EM Program, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD 21701, USA

ARTICLE INFO

ABSTRACT

In our quest to solve biomolecular structures to higher resolutions in cryoEM, care must be taken to deal with all aspects of image formation in the electron microscope. One of these is the Ewald sphere/focus gradient that derives from the scattering geometry in the microscope and its implications for recovering high resolution and handedness information. While several methods to deal with it has been proposed and implemented, there are still questions as to the correct approach. At the high acceleration voltages used for cryoEM, the traditional projection approximation that ignores the Ewald sphere breaks down around 2–3 Å and with large particles. This is likely not crucial for most biologically interesting molecules, but is required to understand detail about catalytic events, molecular orbitals, orientation of bound water molecules, etc. Through simulation I show that integration along the Ewald spheres in frequency space during reconstruction, the "simple insertion method" is adequate to reach resolutions to the Nyquist frequency. Both theory and simulations indicate that the handedness information encoded in such phases is irretrievably lost in the formation of real space images. The conclusion is that correct reconstruction along the Ewald spheres avoids the limitations of the projection approximation.

Introduction

Image formation in the electron microscope occurs through electron scattering and focusing. The geometry of scattering is spherical, where the scattered electron wave front is in phase only on a surface called the Ewald sphere. This is well known in X-ray crystallography, where only reflections on the Ewald sphere are recorded in diffraction patterns (Ewald, 1969). The scattering geometry also means that atoms at different distances along the electron beam deflect electrons with different phases, narrowing the width of the wave front by interference. Each atom thus has its own focus and therefore the ensemble has a focus gradient that is related to the thickness of the specimen (DeRosier, 2000; Downing and Glaeser, 2018; Hoppe, 1970; Spence, 2013). The focusing of the scattered electrons by the objective lens must also be in phase, constituting a second Ewald sphere that is combined with the first to generate the projection image. Because of the high acceleration voltages used in cryoEM, the Ewald sphere is very flat and is often taken to be a plane, representing an infinite depth of focus (Crowther et al., 1970;

DeRosier and Klug, 1968) and known as the projection approximation (Cohen et al., 1984). This is reasonable up to relatively high resolution (2–3 Å) and for typical particles of less than a few hundred ångstrom in diameter.

Several methods to consider the Ewald sphere[1] have been suggested (DeRosier, 2000; Russo and Henderson, 2018), implemented (Leong, 2009; Wolf et al., 2006; Zivanov et al., 2018) and tested (Leong, 2009; Leong et al., 2010; Nakane et al., 2020; Tan et al., 2018; Tegunov et al., 2021; Wolf et al., 2010). It was expected that overcoming the limitation imposed by the projection approximation would contribute significantly to the resolution for thicker specimens. However, the practical results appear to fall short of these expectations, requiring a better analysis to address uncertainties. Much is still being written about the Ewald sphere/focus gradient, indicating that it remains a concern.

Here I examine the issues relating to the Ewald sphere and how that impacts our work in cryoEM. Through simulations, I confirm its equivalence to the focus gradient and how this affects the retrieval of the correct structural information. Most importantly, the phases in the two

Ewald spheres are different, and upon their summation, a fraction of information is lost. I show that the correct treatment is to calculate the reconstruction by integrating the projection image along the Ewald sphere, the "simple insertion method" of Wolf et al. (Wolf et al., 2006). In such reconstructions, even the limited loss of phase information is recovered by integrating projection images with different orientations and revealing that the Ewald sphere/focus gradient does not pose an inherent limit on resolution.

Theory

Calculating the Ewald sphere projection in frequency space

Image formation in the electron microscope involves several contributions from the unscattered beam, and both elastic and inelastic scattering of electrons, one or more times. For cryoEM, the coherent information lies in the electrons scattered elastically only once and we can denote the rest as noise. The Mott formula provides the elastic electron scattering cross sections based on the X-ray cross sections (Mott, 1930). These are derived from quantum mechanical calculations (Coulthard, 1967; Doyle and Turner, 1968) and parameterized by a set of gaussians (Burge and Smith, 1962; Peng et al., 1996; Vand et al., 1957). Because the scattering events are independent, we can represent their interaction with the specimen as a simple summation (Mott, 1930; Peng et al., 1996) (as is done in X-ray crystallography as well). We can therefore accumulate the scattering amplitudes (i.e., the square root of the scattering probabilities) for all the atoms in frequency space. For each atom at a position $\overline{r}_j = \{x_j, y_j, z_j\}$, using its cross section $f_j(\overline{s})$ at a frequency $\overline{s} = \{u, v, w\}$, the sum gives the structure factor:

$$F(\overline{s}) = \frac{\lambda}{V\sqrt{1-\beta^2}}\sum_{j=1}^{n} f_j(\overline{s}) e^{-2\pi i (x_j u + y_j v + z_j w)} \tag{1}$$

over a volume V. The $\frac{\lambda}{\sqrt{1-\beta^2}}$ term is the relativistic electron wavelength, a constant that just affects uniform scaling and is not important in the analysis here.

The Fourier transform of Eq. (1) describes the density in 3D space, but in projection, we only see a 2D image in the direction of the electron beam. In the projection approximation, the density is simply integrated along the beam direction, equivalent to the central section in frequency space (i.e., $w = 0$). This ignores the fact that the scattering has an angular component that results in a spherical rather than planar wave front, the Ewald sphere that is offset from the central section (i.e., w should follow the Ewald sphere and does not lie exactly on the central section: $w = w_{ew} \neq 0$). The focusing of the scattered beam mirrors the scattering geometry, resulting in a second Ewald sphere offset in the opposite direction (i.e., $-w_{ew}$).

The original treatment by Ewald describes the observation of X-ray reflections from a crystal only when they intersect a sphere (Ewald, 1913; Ewald, 1969). Fig. 1A shows how the Ewald sphere relates to the reflections in the Fourier transform of a crystal, with only those falling on the sphere generating a signal. For a single particle the Ewald sphere similarly maps onto the Fourier transform of its density. The red circle in Fig. 1A indicates the first-order Laue zone which corresponds to the inverse of the unit cell length in the electron beam direction. For a single particle map, this is the first layer of pixels from the origin of the Fourier transform (Fig. 1B). In Fig. 1C the spatial frequency distance between the back focal plane and the Ewald sphere for a scattering angle θ is (DeRosier, 2000; Hoppe, 1970):

$$w_{ew} = \frac{2}{\lambda}\sin^2\frac{\theta}{2} \approx \frac{\theta^2}{2\lambda} \approx \frac{\lambda}{2}s^2 \tag{2}$$

where $s \approx \sqrt{u^2 + v^2}$ because $w_{ew}^2 \ll u^2 + v^2$ at the Ewald sphere. This approximation ($\sin\theta \approx \theta$ and $\theta \approx \lambda s$) is the small angle approximation, also known as the parabolic approximation, highly accurate at the small

scattering angles in a high voltage electron microscope (see Supplement Fig. S1). The associated phase shift for an atom j with coordinate z_j along the beam direction to insert into Eq. (1) is (DeRosier, 2000; Hoppe, 1970):

$$\Delta\phi_{ew} = 2\pi z_j w_{ew} \approx \pi\lambda z_j (u^2 + v^2) \approx \pi\lambda z_j s^2 \tag{3}$$

This is the same form as the defocus term in the contrast transfer function (CTF) and thus also represents the difference in focus for an atom at a height z_j. The CTF for an atom j is then:

$$CTF_j(s) = ie^{i\gamma_j(s)} \approx -\sin(\gamma_j(s))$$

$$\gamma_j(s) = \frac{\pi}{2}C_3\lambda^3 s^4 - \pi\lambda(C_1 + z_j)s^2 + \chi \tag{4}$$

where C_1 is the average defocus corresponding to the middle of the particle, λ is the electron wavelength, C_3 is the spherical aberration coefficient and χ is the constant phase shift ("amplitude contrast"). The approximation of the CTF as the sine of the phase shift is valid for the symmetric aberrations.[2] Note that the average defocus is decoupled from the height of the atom, and we need only consider the term containing z_j. The variation in focus also results in partial coherence that dampens the high frequencies (Frank, 1973). However, it is mainly a function of achieving parallel illumination that is well handled in modern microscopes and is beyond the scope of this study.

Because the two Ewald spheres are related in 3D frequency space, the projection can be described as either the addition of the upper and lower spheres at a specific scattering angle (Wolf et al., 2006), or as the addition of the left and right terms with the same scattering angle, just with opposite signs (DeRosier, 2000) (the $F_L(-\overline{s}_{ew}) = F_u(-\overline{s}_{ew})$ and $F_R(\overline{s}_{ew}) = F_u(\overline{s}_{ew})$ terms, respectively, in Fig. 1C). The relationship between the upper and lower spheres adheres to Friedel symmetry in 3D frequency space: $F_u(\overline{s}_{ew}) = F_l^*(-\overline{s}_{ew})$. What is recorded in projection for frequency $\overline{s}$ is then either averaging the upper and lower terms, or averaging the right and left terms:

$$I(u,v) = \frac{1}{2}\left[F_u(\overline{s}_{ew}) + F_l(\overline{s}_{ew})\right] = \frac{1}{2}\left[F_R(\overline{s}_{ew}) + F_L^*(-\overline{s}_{ew})\right] \tag{5}$$

Electron scattering can be viewed as a convolution of the incident wave with the specimen potential, resulting in a spherical wave front with coherent phase. The convolution in real space implies a multiplication in frequency space with the propagation function, $e^{-i\Delta\phi_{ew}}$ (with the phase shift given in Eq. (3)). The second Ewald sphere superimpose on the first as its complex conjugate. The Ewald sphere projection is thus an integration over all the atoms, with the structure factor for each atom multiplied with the average of the two propagation functions[3]:

$$I(u,v) = \frac{1}{2}\sum_{j=1}^{n} f_j(\overline{s}) e^{-2\pi i (x_j u + y_j v + z_j w)}\left[e^{-i2\pi w_{ew}z_j} + e^{i2\pi w_{ew}z_j}\right] \tag{6}$$

This allows us to separate the terms along the electron beam (z and w) and perpendicular to it (x,y and u,v):

$$I(u,v) = \frac{1}{2}\sum_{j=1}^{n} f_j(\overline{s}) e^{-2\pi i (x_j u + y_j v)}\left[e^{-i2\pi(w_{ew}+w)z_j} + e^{i2\pi(w_{ew}-w)z_j}\right] \tag{7}$$

We can approximate the summation over z as an integration over the thickness, t, of the particle, giving the weighting as a sinc function:

$$\frac{1}{t}\int_{-t/2}^{t/2} e^{i2\pi(w_{ew}\pm w)z}dz = \frac{\sin(\pi(w \pm w_{ew})t)}{\pi(w \pm w_{ew})t} \tag{8}$$

This sinc function has maxima on the Ewald spheres, consistent with 3D Fourier transforms of focal series (Kimoto et al., 2012; Taniguchi et al., 1991), and not as sometimes depicted centered on the back focal

---

[2] I omit the other aberration terms for clarity.

[3] I omit any uniform scaling factors such as in equation (1) for clarity.
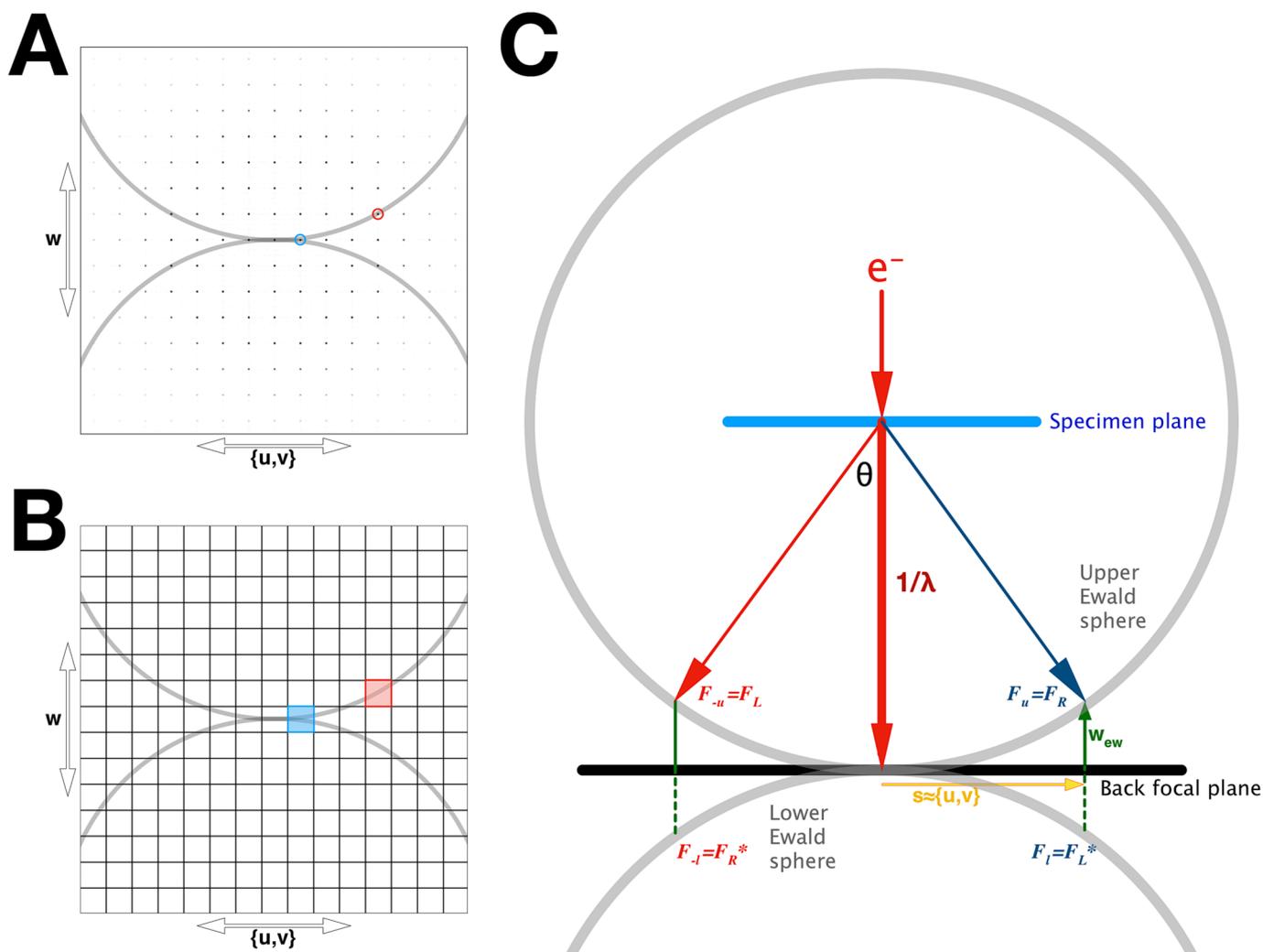
Fig. 1. (A) The Ewald sphere (gray circles) superimposed on a crystal lattice in frequency space. The circles indicate reflections that intersect with the Ewald sphere and would appear in a diffraction image, with the blue circle on the central section plane, while the red circle is in the next layer. (B) The corresponding layout of frequency space voxels of a single particle map where the Ewald sphere intersects with voxels such as the blue one on the central section, progressing towards adjacent layers such as the red voxel. (C) Illustration of the definitions of the different terms describing the Ewald spheres, including the scattering angle, θ, the distance on the back focal plane corresponding to the spatial frequency, s, and the difference between the back focal plane and the Ewald sphere in the direction of the beam, w_{ew}. Also indicated are the structure factors for the upper (F_u) and lower (F_l) Ewald spheres, and the left (F_L) and right (F_R) scattered electrons.
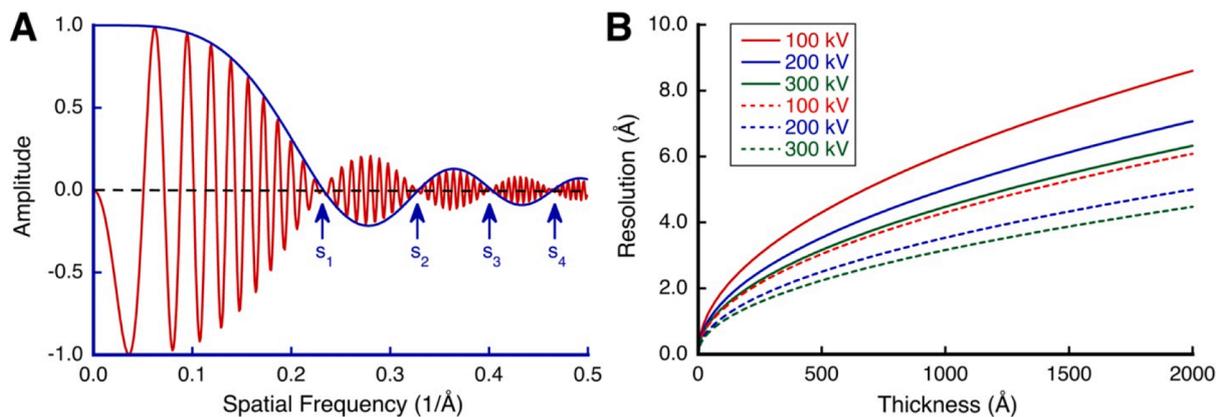

Fig. 2. (A) The sinc function (blue curve) follows the envelope of the CTF (red curve) simulated at 100 kV for a defocus gradient from 1 μm to 1.1 μm (1000 Å thickness). Even if the CTF is corrected to account for the contrast reversals in the red curve, the contrast reversals at the nodes in the sinc envelope remains (arrows). (B) Resolutions where the influence of the Ewald sphere/defocus gradient becomes important (solid curves) and the resolution limits at the first node (s_1) of the sinc function (dashed curves).

plane (Spence, 2013). Note that the effective thickness of the particle in this context, t, is not necessarily its outer diameter, because it depends on the distribution of atoms in the z direction. The projection then incorporates these as weights centered on the two Ewald spheres:

$$I(u,v) \approx \frac{1}{2}\left[\frac{\sin(\pi(w+w_{ew})t)}{\pi(w+w_{ew})t} + \frac{\sin(\pi(w-w_{ew})t)}{\pi(w-w_{ew})t}\right]\sum_{j=1}^{n} f_j(\bar{s})e^{-2\pi i(x_j u + y_j v)} \tag{9}$$

For a single layer of atoms, the thickness is zero, and the projection is equivalent to the central section:

$$I(u,v)|_{t=0} \approx F_{cs}(u,v) = \sum_{j=1}^{n} f_j(\bar{s})e^{-2\pi i(x_j u + y_j v)} \tag{10}$$

For a thick specimen and $w = 0$, we get the thickness-dependent attenuation of the central section:

$$I(u,v)|_{w=0} \approx \frac{\sin(\pi w_{ew}t)}{\pi w_{ew}t}F_{cs}(u,v) \tag{11}$$

Fig. 2A illustrates this attenuation and reversals of contrast at high frequency for a typical average defocus. The phase changes in sign (flips) at the nodes in the sinc function, imposing a resolution limit on the central section at the first node at frequency s_1 (Fig. 2B):

$$r_1 = \frac{1}{s_1} = \sqrt{\frac{\lambda t}{2}} \tag{12}$$

If we set the frequency to one of the Ewald spheres, $w = w_{ew}$, the projection is modified by a shifted sinc function:

$$I(u,v)|_{w=w_{ew}} \approx \frac{1}{2}\left[\frac{\sin(2\pi w_{ew}t)}{2\pi w_{ew}t} + 1\right]F_{cs}(u,v) \tag{13}$$

This indicates that as long as we interpret the projection data as lying along one or the other Ewald sphere, we have partially correct phases. The result is that there is some unavoidable loss of information, but we never get a reversal of contrast, avoiding the limit in Eq. (12).

Eq. (13) implies that we will lose about half of the information at high frequencies. However, the integration in Eq. (8) is an approximation that obscures the correct phases at the Ewald sphere. We can pose comparisons of the Ewald sphere with correct phases, with the projection with the combined Ewald spheres, as Fourier ring correlation (FRC):

$$FRC(s) \approx (1-\kappa)\frac{\sin(2\pi w_{ew}t)}{2\pi w_{ew}t} + \kappa \tag{14}$$

The parameter κ indicates the fraction of information retained at high frequency. As will be shown through simulation, the comparison between one Ewald sphere and the sum of two Ewald spheres gives $\kappa \approx 3/4$. The averaging of the two Ewald spheres (Eq. (6)) is also inherent in the back transformation (averaging of the left and right terms in the Fourier transformation), indicating no opportunity to retrieve Friedel-deviant information from a single projection image. If there is any handedness information encoded at all, it would be if the distribution of atoms in the z direction is skewed. This is likely to be so slight that it is hard to distinguish over a noisy background. It will also change with the direction of projection, indicating a complicated analysis if possible.

Reconstruction considering the Ewald sphere

The appropriate reconstruction in cryoEM is an effort to reverse the projection process, with the integration along both Ewald spheres in frequency space, which also ensures Friedel symmetry in the 3D volume. Adding the projection transform directly to the two Ewald spheres is the simple insertion method of Wolf et al. (Wolf et al., 2006). Given a reconstruction thickness of D_z, the transition to the next layer of voxels occurs where $D_z\frac{\lambda}{2}s^2 > 0.5$. This means that the Ewald sphere already becomes important for reconstruction at a resolution $r < \sqrt{\lambda D_z}$ (solid

curves in Fig. 2B), which also corresponds to the argument of the shifted sinc function in Eq. (14).

For real data in the presence of considerable noise, the drop-off of the sinc function with frequency in Eq. (11) may be confused with a temperature (or B-factor) type effect on the spectral power of the reconstruction:

$$e^{-\frac{B}{2}s^2} \sim sinc\left(\frac{\pi}{2}\lambda t s^2\right) \tag{15}$$

If we set the Gaussian function equal to the sinc function at an argument of $\pi/2$, the apparent B-factor is:

$$B \sim 2\lambda t \ln\frac{\pi}{2} \tag{16}$$

To illustrate the magnitude of this confusion, if we take a specimen at 100 kV with a thickness of 1000 Å, the result is $2 \bullet 0.037\text{Å} \bullet 1000\text{Å} \bullet \ln\frac{\pi}{2} = 33.4\text{Å}^2$. This is toward the low end of reported B-factors for cryoEM reconstructions, usually ascribed to structural heterogeneity (Wlodawer et al., 2017) or errors in the angular alignment of particles (Heymann, 2019; Russo and Passmore, 2014). Taking the Ewald sphere into account doubles the frequency of the sinc function (Eq. (13)) which then translates into doubling the apparent B-factor as well. In a quantitative approach to reconstructions, it is therefore important to understand how the Ewald sphere/focus gradient affects the spectral power at high resolution and not to ascribe it to a temperature factor.

Methods

Implementation

All calculations were done with programs in Bsoft package (Heymann, 2018; Heymann, 2022) with some new additions. (Web page: http://bsoft.ws, Source code available at: https://github.com/CBIIT/Bsoft).

Eq. (1) can be implemented in two equivalent ways. The first is to calculate a 3D map and then extract (with interpolation) the values along the Ewald sphere. Alternatively, the values along the Ewald sphere can be calculated directly from the atomic coordinates (Fig. 3A).

The program bsf was used to calculate 3D maps from atomic coordinates according to Eq. (1), using elemental scattering cross sections (Peng et al., 1996). Because this type of calculation is very computationally expensive for large molecules, the alternative program bgex was also used to calculate a 3D map in real space by placing Gaussian spheres at atom locations.

The program bproject was used to calculate projection images for specified views from a 3D map in frequency space using a kernel-based interpolation method (Lanzavecchia and Bellon, 1995). It was modified to extract values along the Ewald sphere and output either the complex or real projection image. Only the latter corresponds to a projection image recorded in an electron microscope.

A new program, bess, was written to simulate the electron scattering process from individual atoms according to Eq. (1), using the same scattering cross sections as bsf. It further includes options to impose microscope aberrations and to simulate the defocus gradient or Ewald sphere that arises naturally from the distribution of atomic positions in the z direction.

The program bresolve was used to calculate either Fourier ring correlation (FRC in 2D) or Fourier shell correlation (FSC in 3D) to characterize phase differences between comparable images.

The program breconstruct was modified to integrate particle image Fourier transforms either as central sections or along the Ewald spheres in the frequency space reconstruction algorithm. The latter requires specification of the acceleration voltage to set the Ewald sphere radius.

Images with random gaussian distributions of average zero and variance one were generated with the program brandom. Images were masked with the program bedge, using a hard-edge circular or
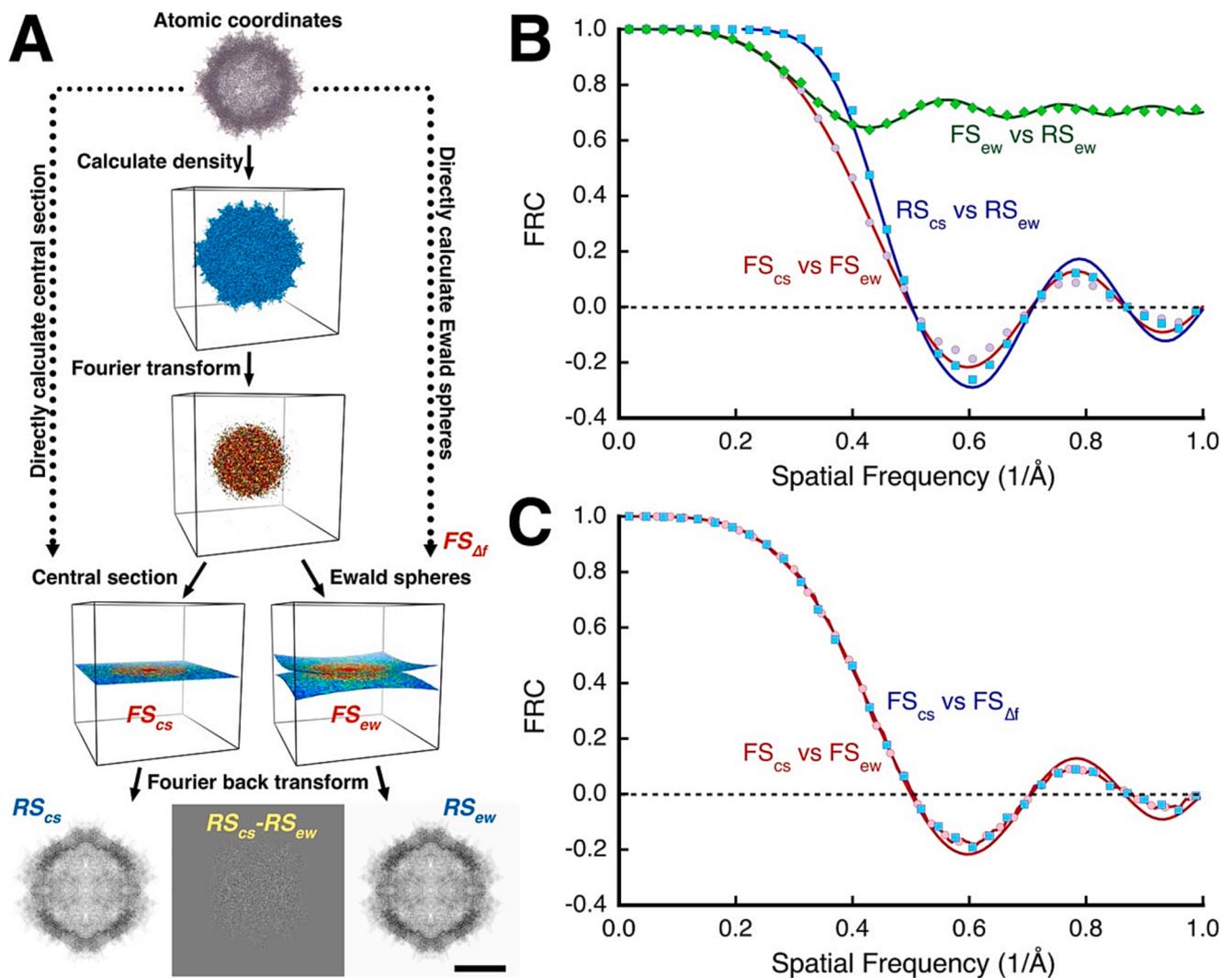
Fig. 3. (A) Flow chart illustrating the calculation of the central section (FS_{cs}) or the Ewald spheres (FS_{ew}) in frequency space. The Ewald spheres can also be directly calculated from the atomic coordinates by imposing a focal shift on each z coordinate ($FS_{\Delta f}$). The frequency space central section and Ewald sphere are then back transformed to produce the real space projections (RS_{cs} and RS_{ew}). The difference image in the bottom center (RS_{cs}-RS_{ew}) emphasizes the subtle difference in the high frequency contributions. (B) The FRC between the central section (FS_{cs}) and the Ewald sphere (FS_{ew}) in frequency space follows a sinc function (red circles) with the first node at 2 Å for the AAV test case (acceleration voltage = 100 kV, defocus Δf = 1 μm). The FRC after transforming both images to real space (RS_{cs} vs RS_{ew}) gives better correlation (blue squares). The FRC between the Ewald sphere and the Ewald sphere projection transformed back to frequency space (green diamonds) indicates that the loss on the conversion to real space is only partial. Fitted equations: $FRC_{red}(s) = sinc(\pi 4.0s^2)$; $FRC_{green}(s) = 0.26sinc(\pi 8.0s^2) + 0.74$; $FRC_{blue}(s) = FRC_{red}(s)/FRC_{green}(s)$. (C) The FRC between the central section and the Ewald sphere (red discs) is the same as for the focus gradient ($FS_{\Delta f}$) imposed on the central section (blue squares).

rectangular shape.

Fits of the comparative data were done with the Levenberg-Marquardt algorithm (Press et al., 1992) as implemented in Kaleidagraph 5 (Synergy Software, Reading, PA) with goodness-of-fit calculated as Pearson's correlation coefficient (R value).

Test cases

The first test case is the structure of an adeno-associated virus (AAV) capsid solved at 2.8 Å (PDB accession number 5UF6) (Xie et al., 2017). Although a better structure at 1.56 Å is available (Xie et al., 2020), it does not change the analysis done here. This particle is roughly spherical, resulting in an almost isotropic radial distribution of atoms. All simulations were done at 100 kV with a box size of 680 × 680 and sampling of 0.5 Å/pixel. Projection images were calculated either from central sections or from the Ewald sphere extracted from the simulated 3D map. For reconstructions, 741 projection images spaced at 1°

intervals in the icosahedral asymmetric unit were calculated and used for reconstruction, either from the central section or Ewald sphere projections, and either by integrating along the central section or the two Ewald spheres.

The second test case is a 70S ribosome dimer of *Thermus thermophilus* solved by X-ray crystallography to 2.6 Å (PDB accession number 1VY4) (Polikanov et al., 2014). Because the map has two molecules of the ribosome, its diameter distribution is highly anisotropic, resulting in various effective thicknesses of the particle. All simulations were done at 100 kV with a box size of 1000 × 1000 and sampling of 0.5 Å/pixel. Projection images were calculated either from central sections or from the Ewald sphere extracted from the simulated 3D map. For reconstructions, 10,326 projection images spaced at 2° intervals were calculated and used for reconstruction, either from the central section or Ewald sphere projections, and either by integrating along the central section or the two Ewald spheres.

The third test case is β-galactosidase of *E. coli* solved by X-ray

crystallography to 1.75 Å (PDB accession number 1JYX) (Juers et al., 2001). All simulations were done at 100 kV with a box size of 400 × 400 and sampling of 0.5 Å/pixel. Projection images were calculated either from central sections or from the Ewald sphere extracted from the simulated 3D map. For reconstructions, 41,258 projection images spaced at 1° intervals were calculated and used for reconstruction, either from the central section or Ewald sphere projections, and either by integrating along the central section or the two Ewald spheres. The inherent D_2 symmetry of the molecule was not used in the calculations.

Results

While the theory appears to be well worked out, some aspects of the Ewald sphere/defocus gradient effect need to be clarified. We therefore turn to simulations to examine the individual effects and their implications, as have been done to some extent previously (Downing and Glaeser, 2018; Wolf et al., 2006).

The problem with the projection approximation is the limitation imposed by the sinc function

The difference between the central section and the Ewald sphere is given by the sinc function in Eq. (11). This can readily be shown by comparing the complex image extracted from the Ewald sphere with the central section in frequency space (Fig. 3A). The red curve in Fig. 3B shows this comparison for the AAV case, closely following a sinc function with the first node at ~2 Å. If we then back transform both images to real space, the FRC now shows a better correlation (blue curve in Fig. 3B), but the nodes in the sinc function remain the same. While there is no loss of information back transforming the central section to real space, the Ewald sphere projection in frequency space deviates from Friedel symmetry and some information is lost on back transformation. The green curve in Fig. 3B indicates that the loss is limited but follows Eq. (14) with double the frequency in the sinc function, converging to a value close to ~3/4 at high frequency. The apparently better correlation of the real space images of the central section and Ewald sphere is a result of normalization, such that $FRC_{blue}(s) = FRC_{red}(s)/FRC_{green}(s)$ (blue curve in Fig. 3B).

The simulation of the Ewald sphere effect is done by either extracting a 2D image from the 3D Fourier transform of the map or by calculating it directly from the atomic structure (Eq. (1) with $w = w_{ew}$) (Fig. 1A). For comparison, the focus gradient is simulated by adopting an individual defocus for each atom based on its z coordinate in calculating the 2D projection (Eq. (4)). Fig. 3C shows that these two approaches produce

the same sinc function in the FRC comparison with the central section.

The sinc function fits in Fig. 3B,C have the first node at a resolution of 2 Å, corresponding to an effective thickness of 216 Å. This is less than the ~290 Å diameter of the particle as well as less than the simulation box edge at 340 Å. As stated in the Theory section, the effective thickness is a function of the distribution of the z-coordinates of the atoms, and not the maximum diameter.

Images from random data also show the same Ewald sphere effects

The nature of the content of the maps does not matter as far as the Ewald sphere/focus gradient effects are concerned. To illustrate this, I generated a volume of 1000^3 with random values according to a gaussian distribution with average zero and variance one, and the sampling set to 0.5 Å/voxel. The FRC between the central section and Ewald sphere projection images for 100 kV follows the same patterns as for simulations from atomic structures (red curve in Fig. 4A). Masking this volume with either a cubic or spherical mask limited the extent of the data present in the z direction (blue and green curves in Fig. 4A). The importance here is that if the FRC comparison is done on projections from a cryoEM reconstruction, the effective thickness will reflect the extent of data in the z direction. It will be the z size of the map unless it is masked, in which case it will be the effective distribution of the mask interior in the z direction.

The effective thickness may vary depending on orientation

Based on the analysis in the previous section, a particle that varies in thickness in different directions is expected to give different FRC curves. I chose an example of the ribosome with two copies in the crystallographic unit cell (Polikanov et al., 2014) giving a decidedly anisotropic distribution of atoms. Fig. 4B shows the FRC curves for comparisons of real space projections from the atomic coordinates in orthogonal directions calculated as central sections and along the Ewald sphere. Clearly evident is the variation of the FRC curves as a function of the orientation. The effective thicknesses are also smaller than the corresponding maximum dimensions in each direction (~332 Å versus ~440 Å for the longest axis, and ~162 Å versus ~200 Å for the shortest axis). The result is that the phase reversals occur at different frequencies and thus may contribute to anisotropy in the resolution. For a reconstruction where the noise extends to the end of the volume, the effective thickness will be similar to the depth of the volume, just as it is for the noise image in the previous section. If a strongly shaped mask is imposed, the effective thickness will correspond to the mask dimension in the
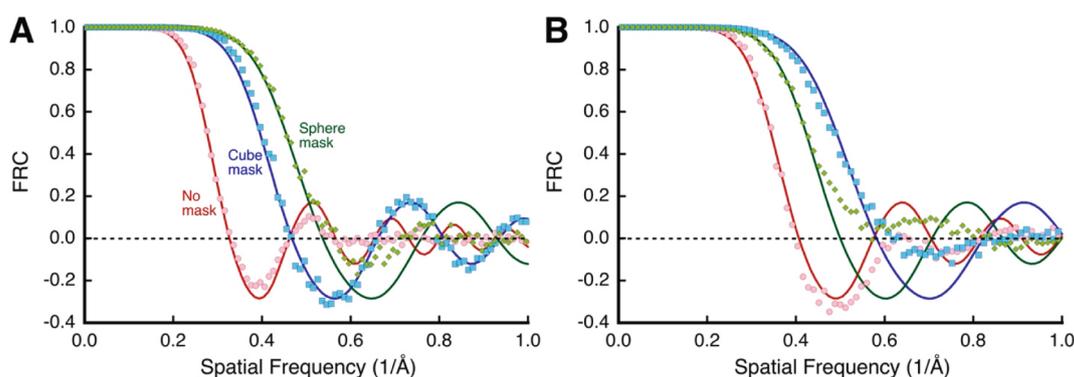

Fig. 4. The influence of effective thickness on comparisons between projection images calculated from the central section and Ewald sphere. (A) Projection images at 100 kV from a volume of 1000^3 with gaussian noise, sampled at 0.5 Å/voxel. The red curve is for the whole volume without masking, while the blue curve is after applying a rectangular mask of 500^3 voxels, and the green curve is after applying a spherical mask of radius 250 voxels. The sinc function fits indicate effective thicknesses of 500 Å, 250 Å and 189 Å respectively. The corresponding first nodes (at the dashed line) are at resolutions of 3.05 Å, 2.15 Å and 1.87 Å, respectively. (B) Projection images calculated from atomic coordinates of a non-spherical particle. The coordinates of the two copies of a 70S ribosome (PDB accession number 1VY4) were used to calculate projections from orthogonal views at 100 kV. The curves are sinc functions for thicknesses of ~332 Å (red), ~215 Å (green) and ~162 Å (blue), compared to the longest (~440 Å) and shortest (~200 Å) axes of the particle.

direction of projection and would be different for different views.

Reconstructing along the Ewald sphere

Reconstructions in cryoEM based on the projection approximation suffer from the resolution limitation indicated in Eq. (11) because of the mismatch between the Ewald sphere and the central section. In a typical simulation that has been commonly used, the central section in a particular orientation is extracted from the 3D map transform and back transformed to real space. If many such projections are integrated into a 3D volume as central sections with the correct orientations, the original map is recovered with the only loss related to the number of projections (Heymann, 2019). It serves as a control to represent an exact reconstruction from a given number of projection images. To examine the effect of the mismatch, central section or Ewald sphere projections were generated and used to calculate reconstructions, either along the central sections or Ewald spheres. These are essentially reversing the flow of processing in Fig. 3A to recover the 3D data from many projection images. Fig. 5 shows the results in terms of the FSC between the original map and the 3D reconstructions, as well as the radial power spectra of the reconstructions. Calculating reconstructions inappropriately (central section to Ewald, cs2ew, or Ewald to central section, ew2cs) causes a severe loss of information beyond the first node in the sinc function. It therefore does not matter in which direction the mismatch is. An Ewald sphere reconstruction from Ewald sphere projections (ew2ew) shows minor loss of phase information, indicating the integration of the

coherent information from one Ewald sphere while averaging out the incoherent information from the other Ewald sphere. These reconstructions have suppressed amplitudes consistent with Eq. (14). At high resolution (1 Å) the power of the ew2ew reconstructions for AAV
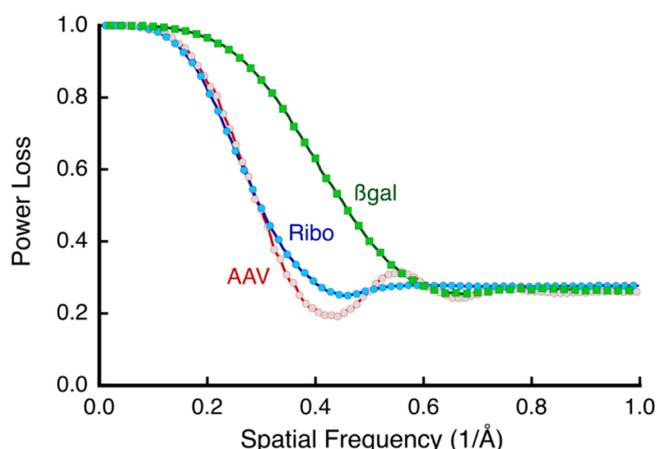

Fig. 6. Loss of power in the reconstructions using the simple insertion method. The curves indicate the ratios of the ew2ew reconstruction powers to the corresponding cs2cs reconstruction in Fig. 5.
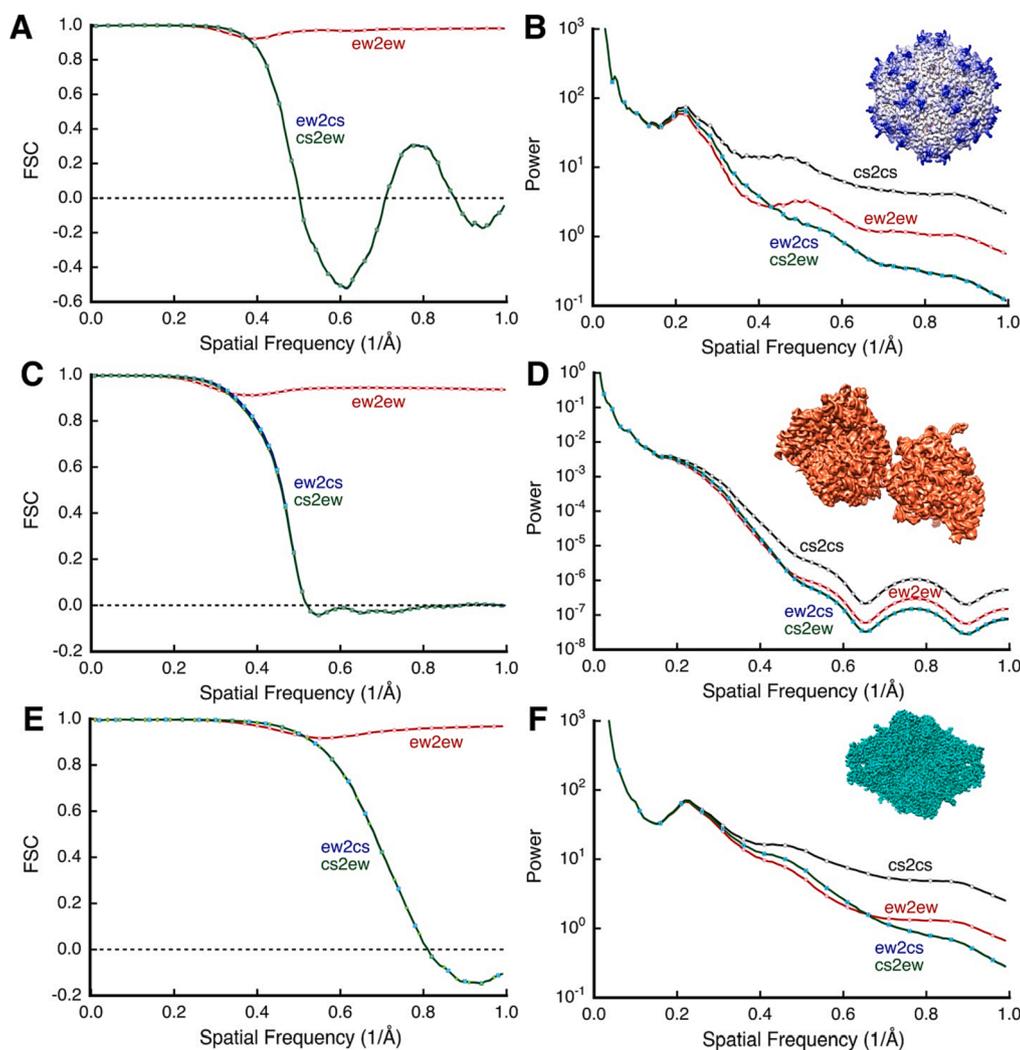

Fig. 5. Reconstructions from simulated projection images calculated from central sections or the Ewald sphere, compared by FSC (A,C,E) and assessed by their radial power spectra (B,D,F), for the cases of AAV (A,B), the ribosome dimer (C,D) and β-galactosidase (E,F). Reconstruction from Ewald sphere projections along the Ewald spheres in frequency space recovers phase information to high resolution (A,C,E: ew2ew, red discs). Reconstruction from central sections along the Ewald spheres (A,C,E: cs2ew, green diamonds) or reconstruction from Ewald sphere projections along the central sections (A,C,E: ew2cs, blue squares) severely suppresses phase information at high resolution. Radial power spectra of the reconstructions coincide at low resolution (B,D,F). Using the central section-to-central section reconstruction (cs2cs, black circles) as reference, the spectral power for the other reconstructions is suppressed.

and β-galactosidase is 0.26 of the reference cs2cs reconstructions, while for the ribosome dimer case, it is 0.28 (Fig. 6). The difference in the latter is likely owing to the way the original map was calculated (gaussian spheres rather than proper atomic cross sections). Also note that the power loss for the spherical AAV looks like a sinc function, while those for the ribosome dimer and β-galactosidase are smoothed out beyond the first node because of their anisotropic diameters.

Discussion

The effect of the Ewald sphere/focus gradient has been known for many years (DeRosier, 2000; Hoppe, 1970). It is consistent with the geometry of near-field or Fresnel diffraction, the propagation function in multi-slice theory and is related to the fractional Fourier transform. However, before the realization of direct electron detectors, we did not have a compelling reason to look beyond the projection approximation. Quite high resolution reconstructions have been reported without considering the Ewald sphere effect. Simulations and reconstructions from real data based on the projection approximation are consistent with the resolution limit of Eq. (12) (Zhu et al., 2018). One of the best cryoEM reconstructions to date also conform to this: apoferritin at 1.22 Å (Nakane et al., 2020) with a diameter (thickness) of ~120 Å and imaged at 300 kV gives a resolution limit of 1.1 Å. The interest in considering the Ewald sphere in reconstruction is therefore to visualize finer details at the atomic and molecular orbital level. In addition, the prospect of quantitatively accurate reconstructions to very high resolution requires attention to every aspect of image formation.

Doing a reconstruction correctly has no limit in resolution related to the Ewald sphere

DeRosier (DeRosier, 2000) suggested two methods to deal with the Ewald sphere to recover the correct data from projections: (i) combining images at different defocus values as also proposed by Schiske (Schiske, 1968), and (ii) combining images from different orientations. In our reconstructions, we accomplish both by integrating particle images from different orientations and at different defoci. It should therefore not be surprising that the "simple insertion" method of Wolf et al. (Wolf et al., 2006) produces the correct phases by averaging in the same way any coherent averaging works. This study shows this simple integration suffices to overcome the limit imposed by the projection approximation (Fig. 5). The loss of information at high resolution is a result of the averaging of the two Ewald spheres in each micrograph, leading to an approximately fourfold suppression of the spectral power in reconstructions (Fig. 6). Any improvement on the simple insertion method must therefore reflect a more quantitative recovery of the power. Because the phases carry most of the relevant structural information, such an exercise may not be necessary to solve structures to high resolution.

What constitutes an improvement in resolution on considering the Ewald sphere?

The influence of the Ewald sphere is at high frequencies, complicating an assessment of the any improvement in resolution on considering it. The hard limit in the projection approximation is at the first node of the sinc function (Eq. (12)). However, as indicated in Fig. 2B, the influence should also be evident at lower frequencies. We can take the resolution at which we should expect an improvement as $\sqrt{\lambda D_z}$ and approximate the effective thickness D_z as 80 % of the outer particle diameter. In the first attempt to use the simple insertion method, the reconstruction of the bovine papillomavirus achieved a resolution of 3.6 Å, with marginal improvement on considering the Ewald sphere (Wolf et al., 2010). Taking the outer diameter for the virus as ~550 Å imaged at 300 kV, the resolution at which they would be expected to obtain an

improvement is ~3 Å. Similarly for the two cases in Tan et al. (Tan et al., 2018), with outer diameters of 250 Å for the AAV2 capsid and 700 Å for the DLP particle, the expected resolutions to see an improvement are ~2 Å and ~3.3 Å, respectively. They reported improvement in resolution of 1.9 Å to 1.8 Å and 2.6 Å to 2.3 Å. This demonstrates that the projection approximation is sufficient up to quite high resolution, and any assessment in considering the Ewald sphere should take into account at what resolution it would be impactful.

Recovering the individual Ewald sphere phases may be impossible

Because the Ewald spheres deviate from Friedel symmetry and this information is lost on conversion to real space, it was thought that there may be some ways to recover it. As shown in the Theory section, the information from the Ewald spheres is already incorporated in the projection image, suggesting it is unlikely that we will be able to deconvolve them from a single image. The method proposed by DeRosier (DeRosier, 2000) to use images taken at different focus values unfortunately suffers from ill-conditioned matrices (the determinants are actually zero) (see the Supplement for more detail).

Both Wolf et al. (Wolf et al., 2006) and Leong (Leong, 2009) developed iterative methods to attempt to refine the individual Ewald sphere contributions. In these a previous reconstruction is used to adjust for the phase difference between $F_u(s)$ and $F_l(s)$ and to iterate the process until convergence. The implementation by Leong (Leong, 2009) has a quirk in that the combined CTF and Ewald sphere correction at low resolution is unstable, suggesting a problem with the applicability of the algorithm. The solution is to correct for the CTF at low resolution assuming the projection approximation, while applying the Ewald sphere algorithm only to high resolution terms. It appeared to perform the expected enhancements in simulations. However, in real cases, it did not yield any improvement (Leong, 2009; Leong et al., 2010). The lesson here is that any algorithm that aims at resolving the two Ewald sphere terms must do so while adhering closely to the projection approximation at low resolution.

In a different approach, Russo et al. (Russo and Henderson, 2018) proposed to use single sideband images that contribute to the full CTF (also shown in Downing and Glaeser (Downing and Glaeser, 2008)). Their idea is that the left and right shifts imposed by defocusing can be separated to represent the left and right terms in the Ewald spheres. However, defocusing causes a lateral shift while the Ewald sphere offset is in the electron beam direction. It is therefore unclear how the single sideband terms relate to the Ewald spheres. Eq. (4) indicates that the focus variation can be decoupled from the average focus, so that the specific lateral shifts are unrelated to the Ewald sphere. The use of the single sideband images also has the problem that it is direction dependent. This means there is a discontinuity at the seam perpendicular to the direction, where the CTF switches between the two terms. Their solution is to apply the single sideband terms in sectors to avoid the seams. This problem is also evident at low frequencies where it should have approached the projection approximation and not present such discontinuities.

The single sideband method (Russo and Henderson, 2018) was implemented in Relion 3.1 (Zivanov et al., 2018). Nakane et al. (Nakane et al., 2020) reported an improvement in resolution from 1.3 to 1.22 Å for apoferritin at 300 kV. The program M (Tegunov et al., 2021) appears to implement the Ewald sphere handling based on the same method. It purports to generate the Ewald sphere projections convolved with the CTF and its conjugate and use these as reference during orientation refinement. This is claimed as an improvement over the Relion implementation, but again to a modest extent (1.41 to 1.34 Å). It is likely that the improvement could simply be attributed to integration along the Ewald sphere together with an additional degree of freedom in particle alignment. Where the attempted Ewald sphere correction does not improve the resolution, it is likely never reported.

Handedness information is completely lost on projection

The theory indicates that the back transformation of the Ewald sphere projection to real space eliminates any imaginary part that carries handedness information. The two Ewald spheres are related, and their combination cancels out any directional information. The only potential way handedness information persists, is if the distribution of atomic z coordinates is sufficiently skewed to produce a focus offset from the center of the particle. This would be orientation-dependent and requiring an analysis in every view direction. In noise-dominated micrographs and considering that this information is only available at high resolution, it is unlikely to be easily retrieved.

The influence of reconstruction thickness

In using the projection approximation, the thickness of the reconstruction is critical in that it imposes a resolution limit (Eq. (12)). Therefore, a possible remedy is to decrease the thickness, effectively increasing the central section voxel depth and thus accessing higher frequency information along the Ewald sphere. Zhu et al. (Zhu et al., 2018) developed a method to reconstruct a very large virus capsid (1900 Å) piece by piece, applying the appropriate defocus for each small reconstruction individually, based on its distance in z from the particle center. This approach yielded an improvement from 4.4 (the theoretical limit based on Eq. (12)) to 3.5 Å at 300 kV (Fang et al., 2019d). The reconstruction was done in EMAN2, which adheres to the projection approximation and is thus subject to its limitations. While this approach showed some success, it is laborious and likely benefit from the icosahedral symmetry of the capsid.

In contrast, reconstruction along the Ewald sphere eliminates the thickness dependence in terms of achievable resolution. However, the thickness of the reconstruction still determines at what frequency the power decreases as in Fig. 6. While this does not limit the resolution, it still needs to be considered in assessing the decrease in the radial power spectrum of the reconstruction as opposed to a gaussian decay that is typically ascribed to alignment inaccuracy and structural heterogeneity.

Conclusions

It is now clear that the simple insertion method yields the correct reconstruction to Nyquist frequency. It incurs a cost in averaging out incorrect phases in each of the two Ewald spheres, suppressing the spectral power of the reconstruction by fourfold at high resolution. Any improvement on the simple insertion method will have to show a recuperation of the phases at each Ewald sphere and the power at high resolution. Because the individual phases of the Ewald spheres are lost on summation and back transformation, it is unlikely to be retrievable from single images.

The correct handling of the Ewald sphere/focus gradient is therefore in generating the proper projection images from a reference map and the proper integration during reconstruction, both along the Ewald spheres. With this approach, the Ewald sphere/focus gradient does not impose a hard limit on the achievable resolution and should never present a problem in cryoEM.

Funding information

CRediT authorship contribution statement

J. Bernard Heymann: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yjsbx.2022.100083.

References

Burge, R.E., Smith, G.H., 1962. New Calculation of Electron Scattering Cross Sections and a Theoretical Discussion of Image Contrast in Electron Microscope. Proc. Phys. Soc. Lond. 79, 673–1000.

Cohen, H.A., Schmid, M.F., Chiu, W., 1984. Estimates of validity of projection approximation for three-dimensional reconstructions at high resolution. Ultramicr. 14, 219–226.

Coulthard, M.A., 1967. A Relativistic Hartree-Fock Atomic Field Calculation. Proc. Phys. Soc. Lond. 91, 44–49.

Crowther, R.A., DeRosier, D.J., Klug, A., 1970. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. Proc. R. Soc. Lond. 317, 319–340.

DeRosier, D.J., 2000. Correction of high-resolution data for curvature of the Ewald sphere. Ultramicr. 81, 83–98.

DeRosier, D.J., Klug, A., 1968. Reconstruction of three dimensional structures from electron micrographs. Nature 217, 130–134.

Downing, K.H., Glaeser, R.M., 2008. Restoration of weak phase-contrast images recorded with a high degree of defocus: the "twin image" problem associated with CTF correction. Ultramicr. 108, 921–928.

Downing, K.H., Glaeser, R.M., 2018. Estimating the effect of finite depth of field in single-particle cryo-EM. Ultramicr. 184, 94–99.

Doyle, P.A., Turner, P.S., 1968. Relativistic Hartree-Fock X-Ray and Electron Scattering Factors. Acta Crystallogr. Section A 24, 390–1000.

Ewald, P.P., 1913. The theory of the interference of X-rays in crystals. Physikalische Zeitschrift 14, 465–472.

Ewald, P.P., 1969. Introduction to Dynamical Theory of X-Ray Diffraction. Acta Crystall. A 25, 103–108.

Fang, Q., Zhu, D., Agarkova, I., Adhikari, J., Klose, T., Liu, Y., Chen, Z., Sun, Y., Gross, M. L., Van Etten, J.L., Zhang, X., Rossmann, M.G., 2019. Near-atomic structure of a giant virus. Nat. Commun. 10, 388.

Frank, J., 1973. Envelope of Electron-Microscopic Transfer-Functions for Partially Coherent Illumination. Optik 38, 519–536.

Heymann, J.B., 2018. Guidelines for using Bsoft for high resolution reconstruction and validation of biomolecular structures from electron micrographs. Protein Sci. 27, 159–171.

Heymann, J.B., 2019. Single-particle reconstruction statistics: a diagnostic tool in solving biomolecular structures by cryo-EM. Acta Crystallogr. F Struct. Biol. Commun. 75, 33–44.

Heymann, J.B., 2022. Bsoft: Image Processing for Structural Biology. Bio-protocol 12, e4393.

Hoppe, W., 1970. Principles of electron structure research at atomic resolution using conventional electron microscopes for the measurement of amplitudes and phases. Acta Crystallographica Section A 26, 414–426.

Juers, D.H., Heightman, T.D., Vasella, A., McCarter, J.D., Mackenzie, L., Withers, S.G., Matthews, B.W., 2001. A structural view of the action of Escherichia coli (lacZ) beta-galactosidase. Biochemistry 40, 14781–14794.

Kimoto, K., Kurashima, K., Nagai, T., Ohwada, M., Ishizuka, K., 2012. Assessment of lower-voltage TEM performance using 3D Fourier transform of through-focus series. Ultramicr. 121, 31–37.

Lanzavecchia, S., Bellon, P.L., 1995. A Bevy of Novel Interpolating Kernels for the Shannon Reconstruction of High-Bandpass Images. J. Vis. Commun. Image Repr. 6, 122–131.

Leong, P.A., 2009. Computational challenges in high-resolution cryo-electron microscopy. California Institute of Technology. Dissertation (Ph.D.).

Leong, P.A., Yu, X., Zhou, Z.H., Jensen, G.J., 2010. Correcting for the ewald sphere in high-resolution single-particle reconstructions. Methods Enzymol. 482, 369–380.

Mott, N.F., 1930. The scattering of electrons by atoms. Proceedings of the Royal Society of London Series a-Containing Papers of a Mathematical and Physical Character 127, 658-665.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P., Grigoras, I.T., Malinauskaite, L., Malinauskas, T., Miehling, J., Uchanski, T., Yu, L., Karia, D., Pechnikova, E.V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S.W., Chirgadze, D.Y., Murshudov, G., Aricescu, A.R., Scheres, S.H.W., 2020. Single-particle cryo-EM at atomic resolution. Nature 587, 152–156.

Peng, L.-M., Ren, G., Dudarev, S.L., Whelan, M.J., 1996. Robust parameterization of elastic and absorptive electron atomic scattering factors. Acta Crystall. A52, 257–276.

Polikanov, Y.S., Steitz, T.A., Innis, C.A., 2014. A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. Nat. Struct. Mol. Biol. 21, 787–793.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C, second ed. Cambridge University Press, Cambridge.

Russo, C.J., Henderson, R., 2018. Ewald sphere correction using a single side-band image processing algorithm. Ultramicr. 187, 26–33.

Russo, C.J., Passmore, L.A., 2014. Robust evaluation of 3D electron cryomicroscopy data using tilt-pairs. J. Struct. Biol. 187, 112–118.

Schiske, P., 1968. Image reconstruction by means of focus series, pp. 145-146 Proceedings of the Fourth Regional Congress on Electron Microscopy, Vol. 1, Rome.

Spence, J.C.H., 2013. High-Resolution Electron Microscopy OUP Oxford.

Tan, Y.Z., Aiyer, S., Mietzsch, M., Hull, J.A., McKenna, R., Grieger, J., Samulski, R.J., Baker, T.S., Agbandje-McKenna, M., Lyumkis, D., 2018. Sub-2 Å Ewald curvature corrected structure of an AAV2 capsid variant. Nat. Commun. 9, 3628.

Taniguchi, Y., Ikuta, T., Shimizu, R., 1991. Assessment of Image-Formation by 3-Dimensional Power Spectrum in Transmission Electron-Microscopy. J. Electron Microsc. 40, 5–10.

Tegunov, D., Xue, L., Dienemann, C., Cramer, P., Mahamid, J., 2021. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 A in cells. Nat. Methods 18, 186–193.

Vand, V., Eiland, P.F., Pepinsky, R., 1957. Analytical Representation of Atomic Scattering Factors. Acta Crystall. 10, 303–306.

Wlodawer, A., Li, M., Dauter, Z., 2017. High-Resolution Cryo-EM Maps and Models: A Crystallographer's Perspective. Structure 25 (1589–1597), e1581.

Wolf, M., DeRosier, D.J., Grigorieff, N., 2006. Ewald sphere correction for single-particle electron microscopy. Ultramicr. 106, 376–382.

Wolf, M., Garcea, R.L., Grigorieff, N., Harrison, S.C., 2010. Subunit interactions in bovine papillomavirus. PNAS 107, 6298–6303.

Xie, Q., Spear, J.M., Noble, A.J., Sousa, D.R., Meyer, N.L., Davulcu, O., Zhang, F., Linhardt, R.J., Stagg, S.M., Chapman, M.S., 2017. The 2.8 A Electron Microscopy Structure of Adeno-Associated Virus-DJ Bound by a Heparinoid Pentasaccharide. Mol. Ther. Methods Clin. Dev. 5, 1–12.

Xie, Q., Yoshioka, C.K., Chapman, M.S., 2020. Adeno-Associated Virus (AAV-DJ)-Cryo-EM Structure at 1.56 A Resolution. Viruses 12, 1–15.

Zhu, D., Wang, X., Fang, Q., Van Etten, J.L., Rossmann, M.G., Rao, Z., Zhang, X., 2018. Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. Nat. Commun. 9, 1552.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., Scheres, S.H., 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7, 1–22.