Phase Problems for Multidimensional and Undersampled Data

Rick Millane

Purdue University

Workshop on New Approaches to the Phase Problem

Lawrence Berkeley National Laboratory

17–19 May 2001

OUTLINE

Preliminaries

- Nyquist sampling
- "Oversampling"
- Aliasing, sampling, solvent volume

Undersampled Data

- Effect of size & shape of support
- Effect of symmetry

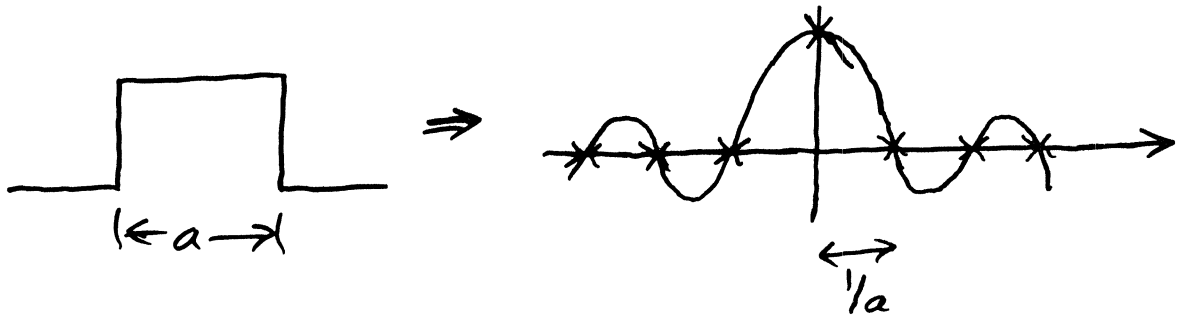
Multidimensional (3D) case

- Uniqueness
- Sampling requirements
- Stability & convergence
- Fiber diffraction

Application of the HIO algorithm to protein crystallography

NYQUIST SAMPLING

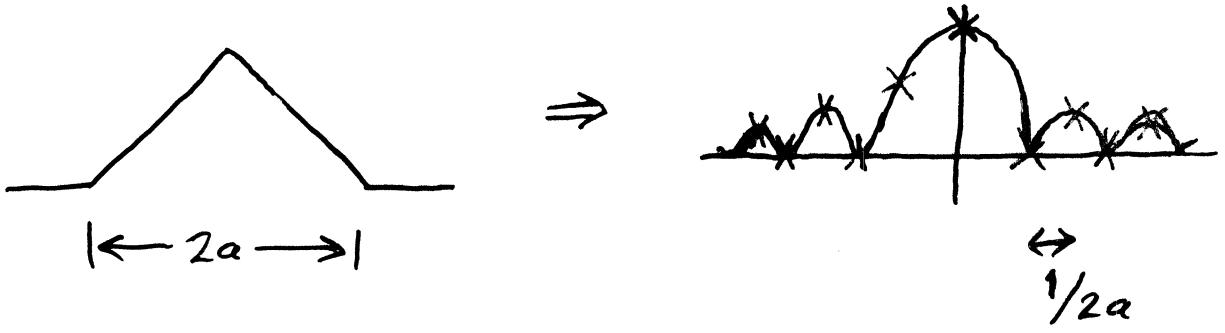
Minimum sample spacing to reconstruct a band-limited signal.



Support of image = a $\Rightarrow$ Nyquist spacing = $1/a$

Autocorrelation

Intensity



Use of term "oversampling"

Oversampling usually means < Nyquist spacing

NOT < twice Nyquist spacing

If the spacing is > Nyquist we say that the signal is "undersampled".

FOR PHASE RETRIEVAL

Is the continuous amplitude sufficient to uniquely recover the phase?

1D - no

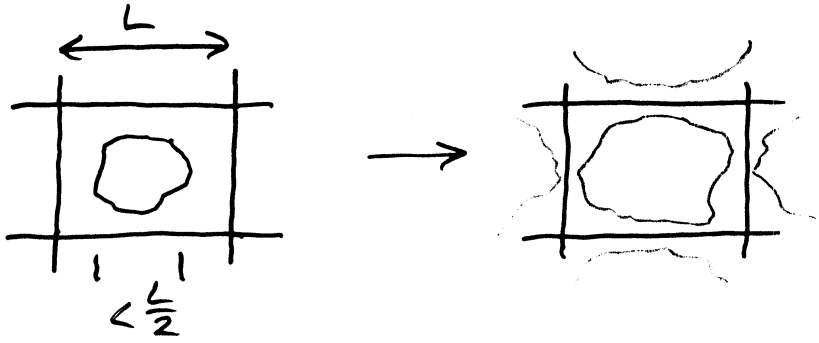
2D - yes

3D - more than sufficient by a factor of two

ALIASING, SAMPLING & SOLVENT VOLUME

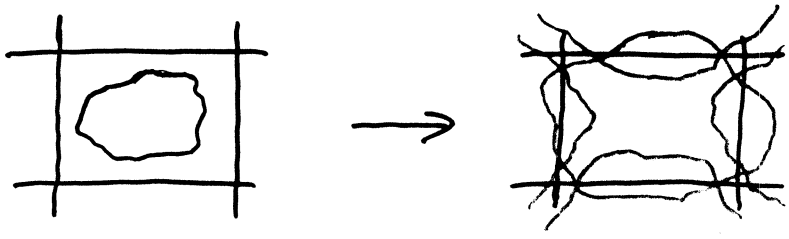
Aliasing $\Rightarrow$ overlap of supports if the spectrum is undersampled.

Two dimensions



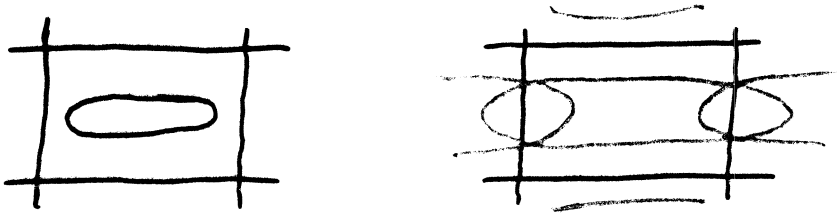
Sampling intensity at $\frac{1}{L}$ is ok

$$\frac{V_{molecule}}{V_{cell}} < \frac{1}{4}$$



Sampling at $\frac{1}{L}$ not ok

$$\frac{V_{molecule}}{V_{cell}} > \frac{1}{4}$$



$$\frac{V_{molecule}}{V_{cell}} < \frac{1}{4}$$, but Sampling at $\frac{1}{L}$ is not ok

Error due to support being "too large"
OR sample spacing being too large

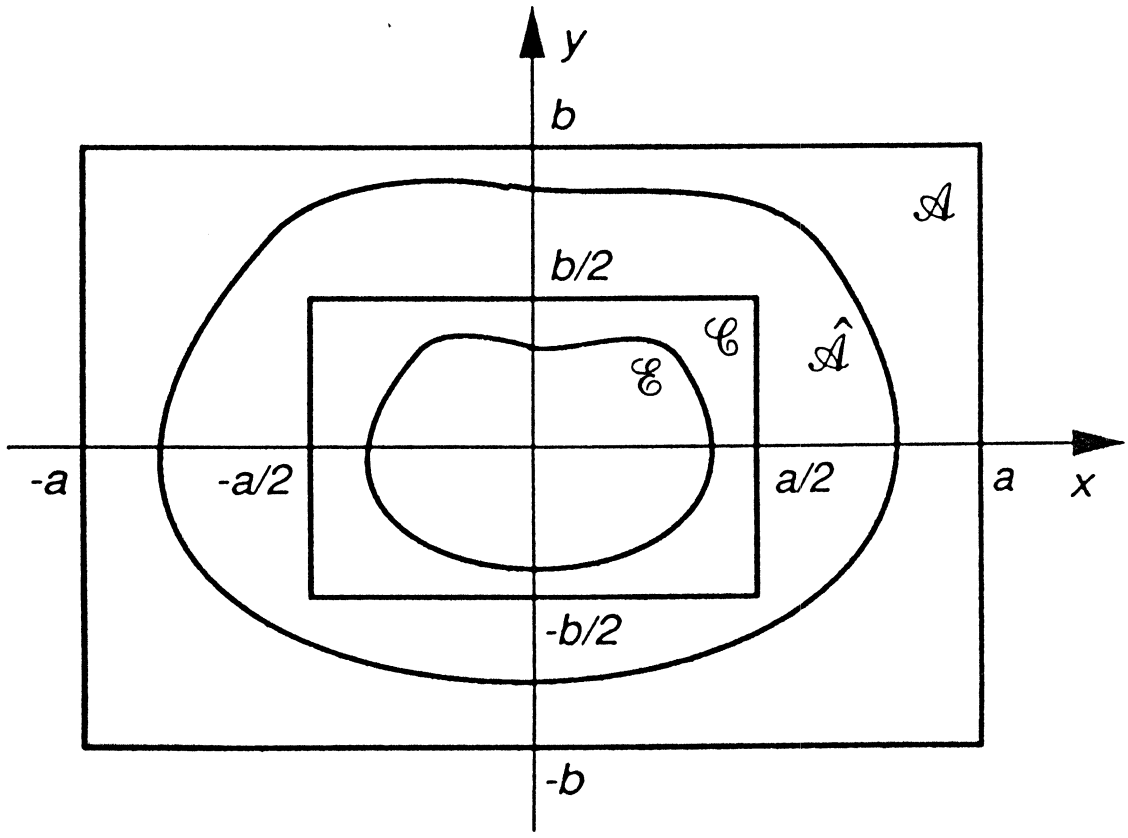
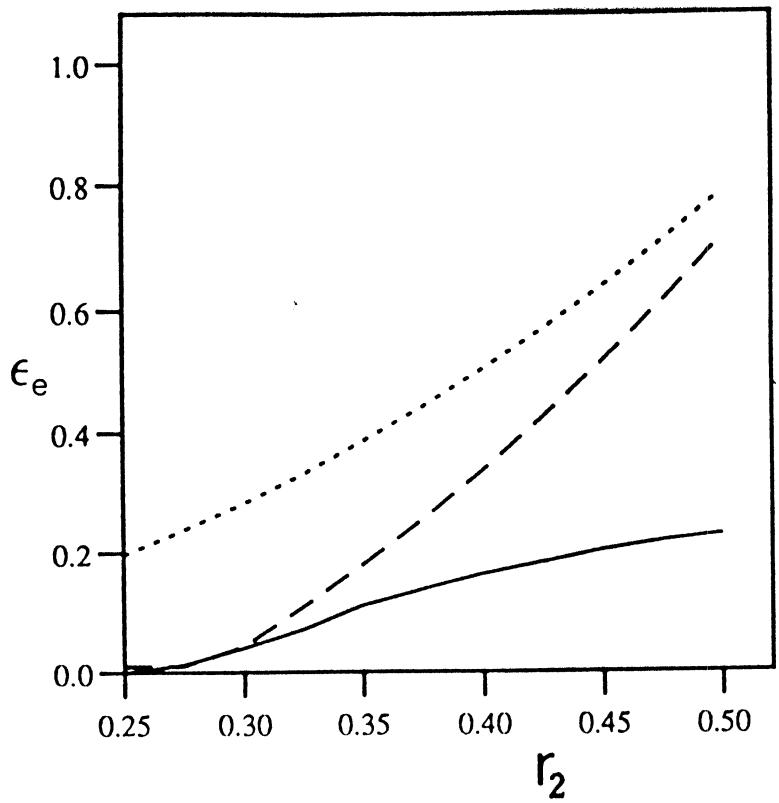


Fig. 1. Image space showing the regions $\mathcal{C}$, $\mathcal{A}$, $\mathcal{E}$ and $\hat{\mathcal{A}}$ defined in the text.

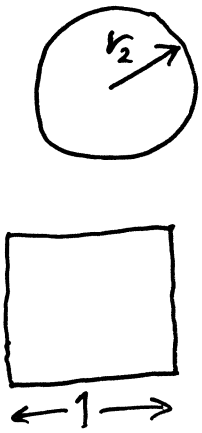
Millane, J. Opt. Soc. Am. A, 10, 1037-1045 (1993).

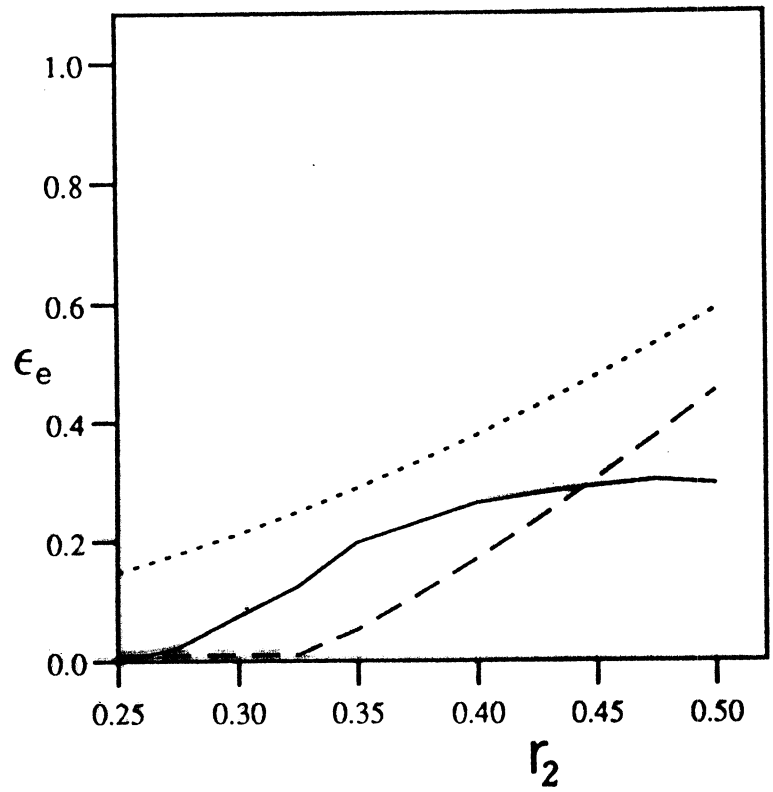
Aliasing error



(a)

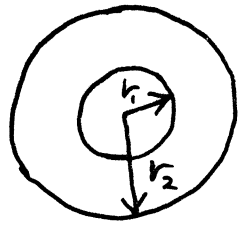
Circular support





(b)

Annulus



$r_1 = \frac{r_2}{2}$

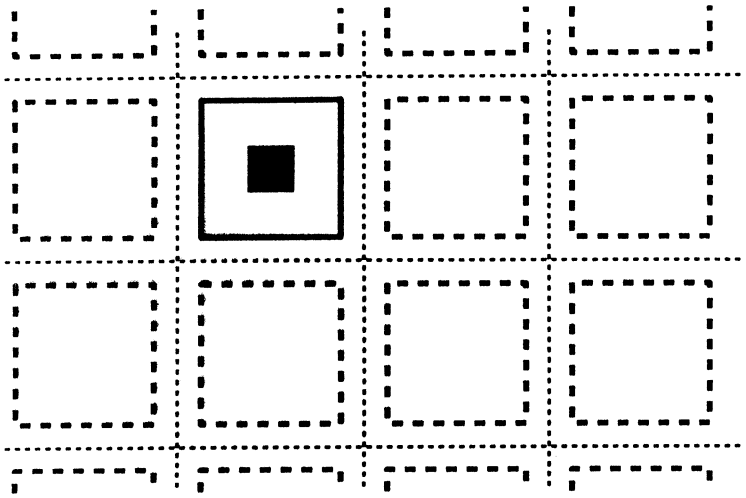
Fig. 4. Error metrics $\epsilon_e^{(1)}$ (– – –), $\epsilon_e^{(2)}$ (——) and $\epsilon_e^{(3)}$ (· · ·) for annular envelopes as a function of r_2, for (a) $r_1 = 0$ and (b) $r_1 = r_2/2$.

Constraints: symmetry

There are two type of symmetry: crystallographic and non−crystallographic symmetry.

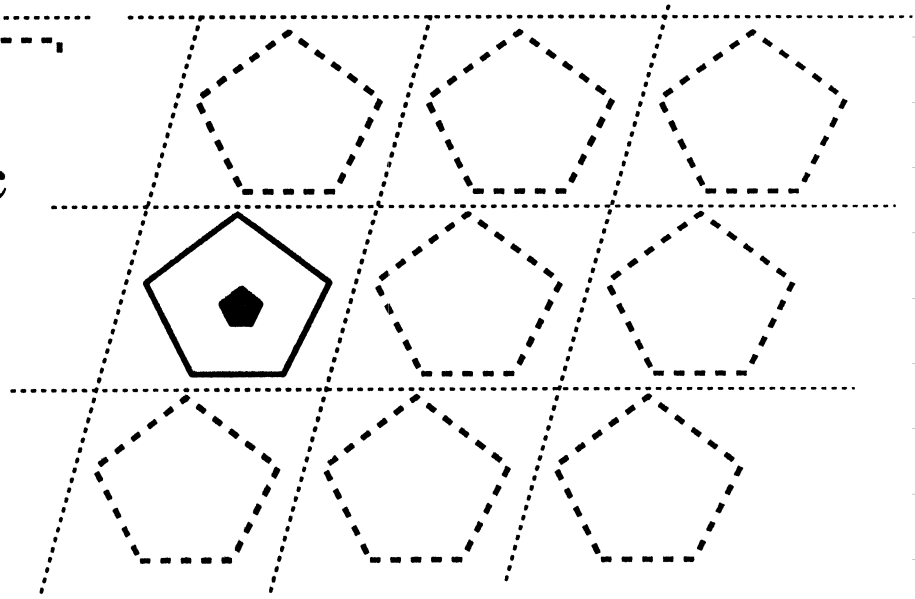
Crystallographic symmetry holds true over the entire crystal and gives relationships between the data but it gives no new phase information.

Non−crystallographic symmetry only holds locally. It is not trivially contained the structure factors, and provides phase information



4−fold crystallographic symmetry

5−fold non−crystallo−graphic symmetry



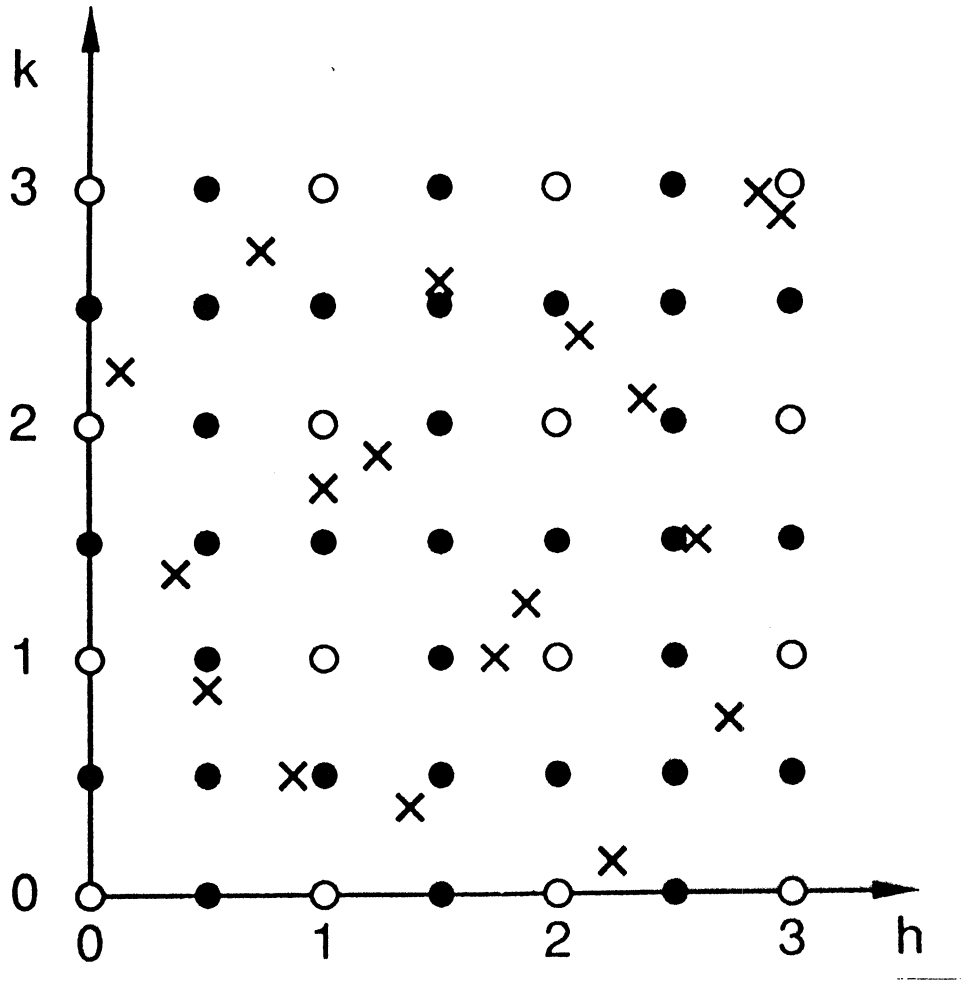


Fig. 2. One quadrant of reciprocal (Fourier) space showing the reciprocal lattice points (○), the Nyquist sample points for $|E(u,v)|^2$ (○ and ●), and the additional sample points that result from threefold noncrystallographic symmetry (×).

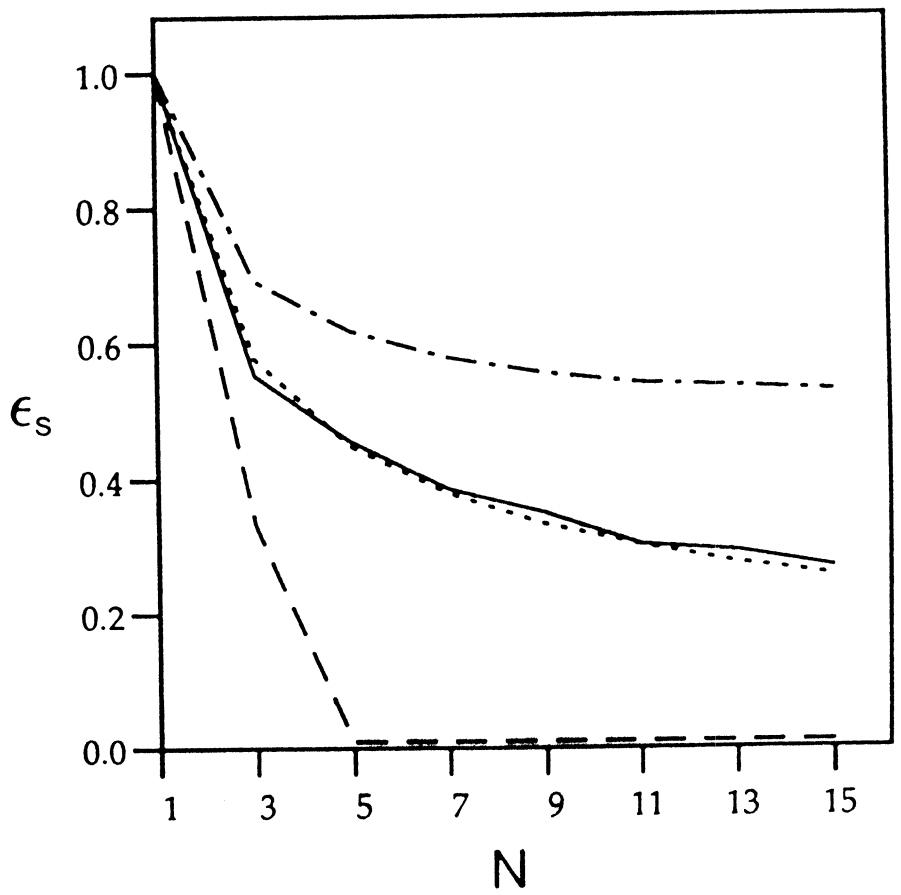
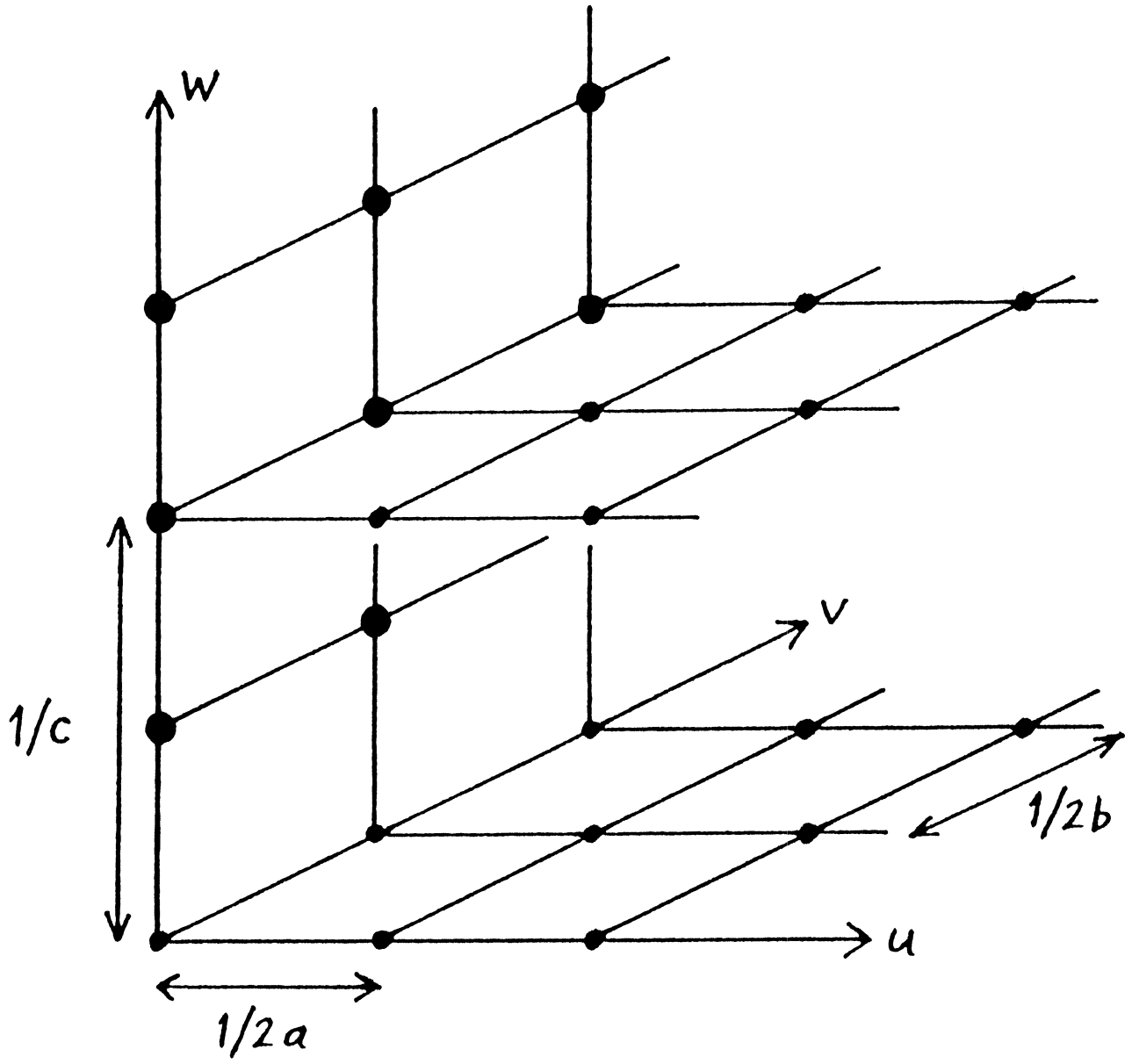


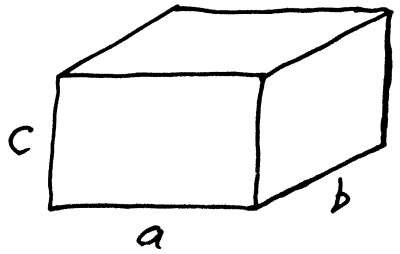
Fig. 7. Error metrics $\epsilon_s^{(1)}$ (– – –), $\epsilon_s^{(2)}$ (——), $\epsilon_s^{(3)}$ (– · – ·), and $\epsilon_s^{(5)}$ (· · ·) as a function of N.

MULTI DIMENSIONAL CASE

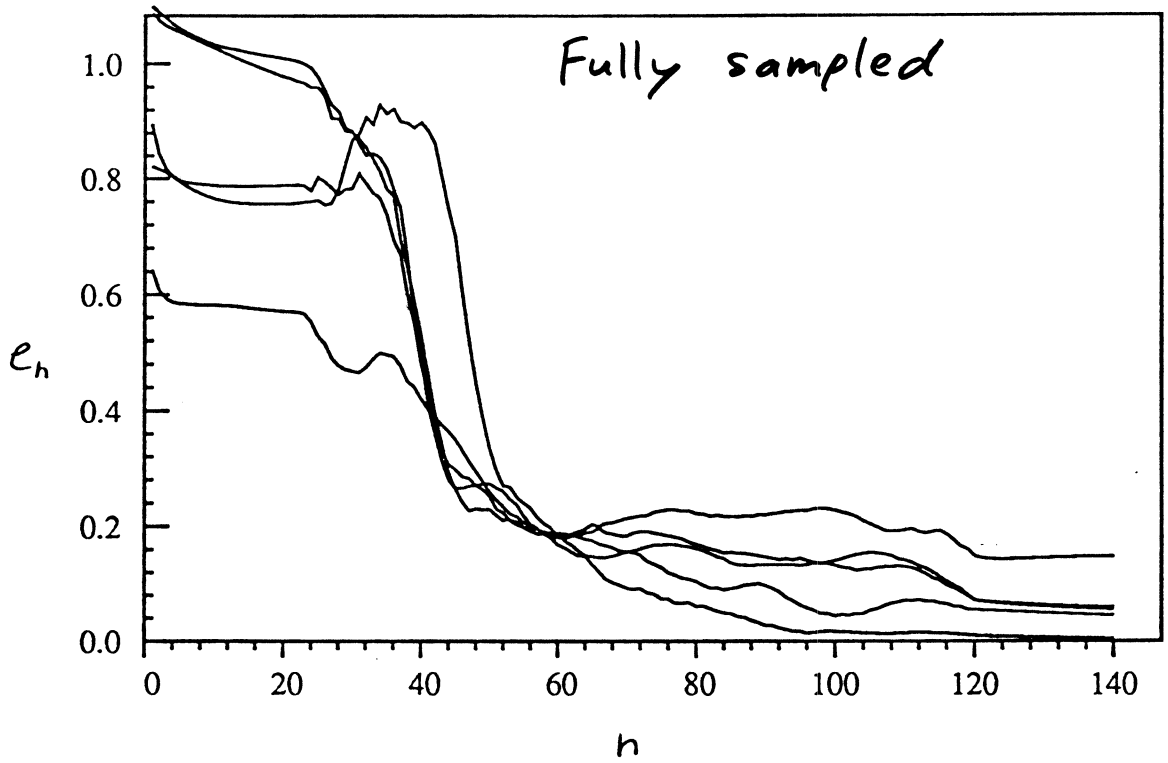
Millane, J. Opt. Soc. Am. A, 13, 725-734 (1996).

3-D case:

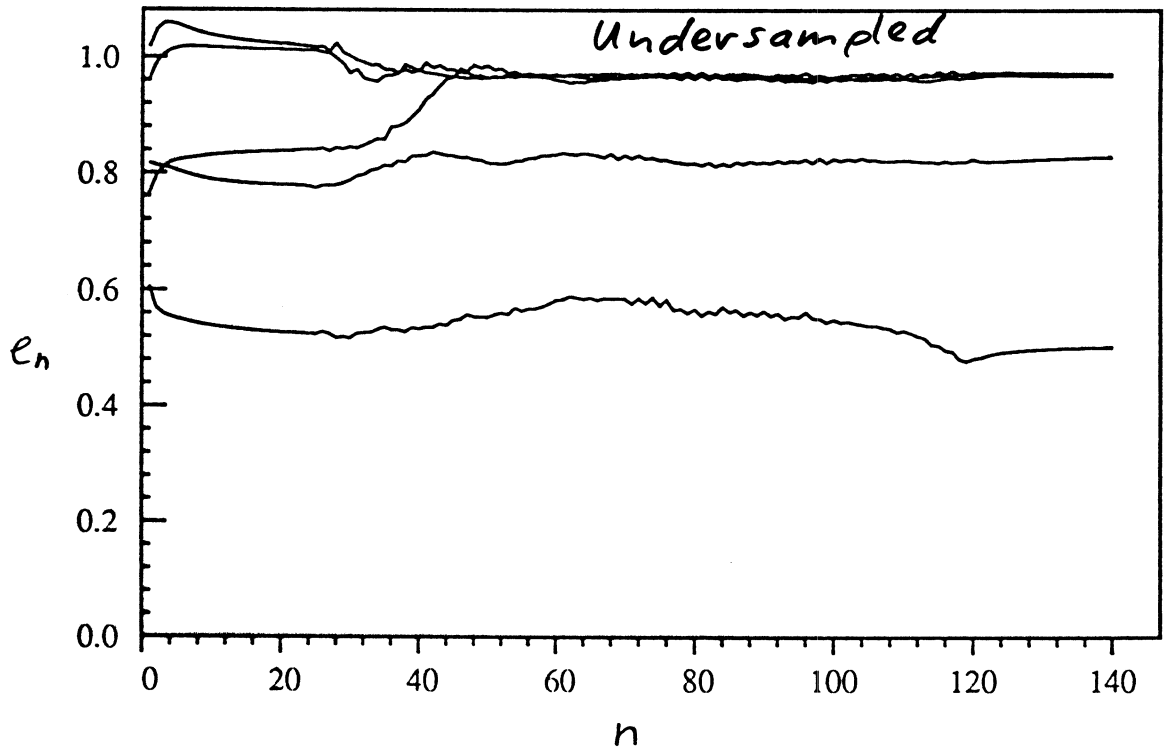




2D 8×8 image



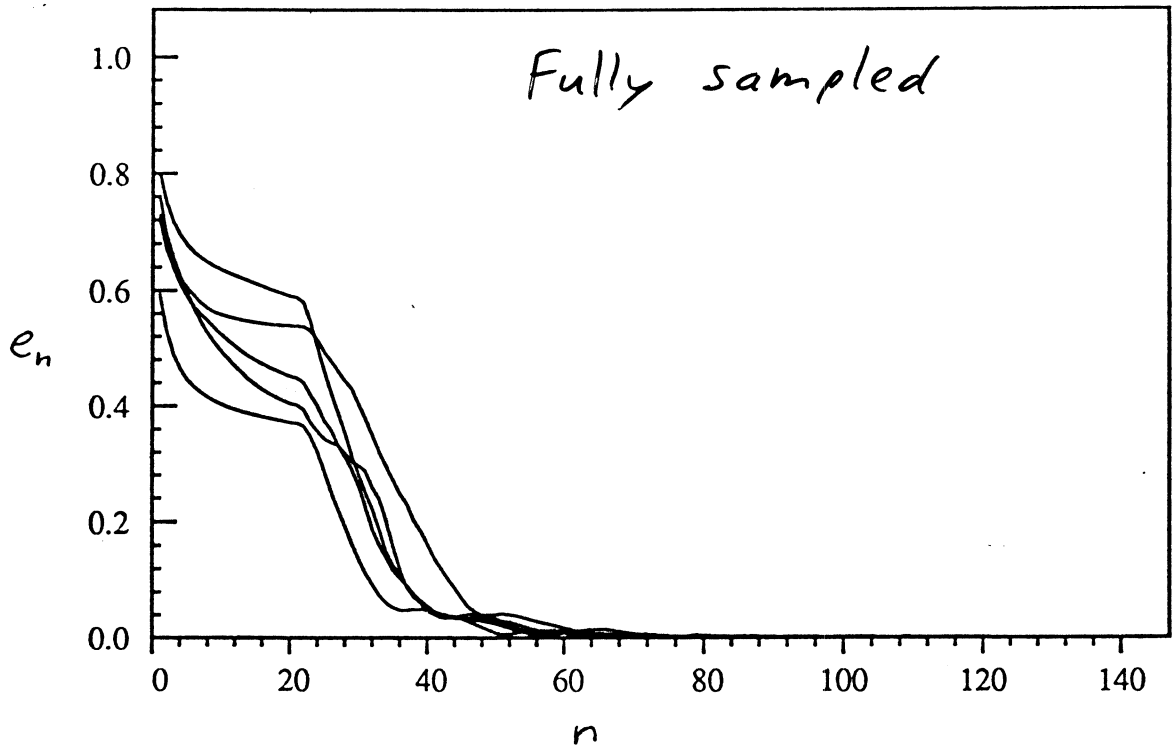
(a)



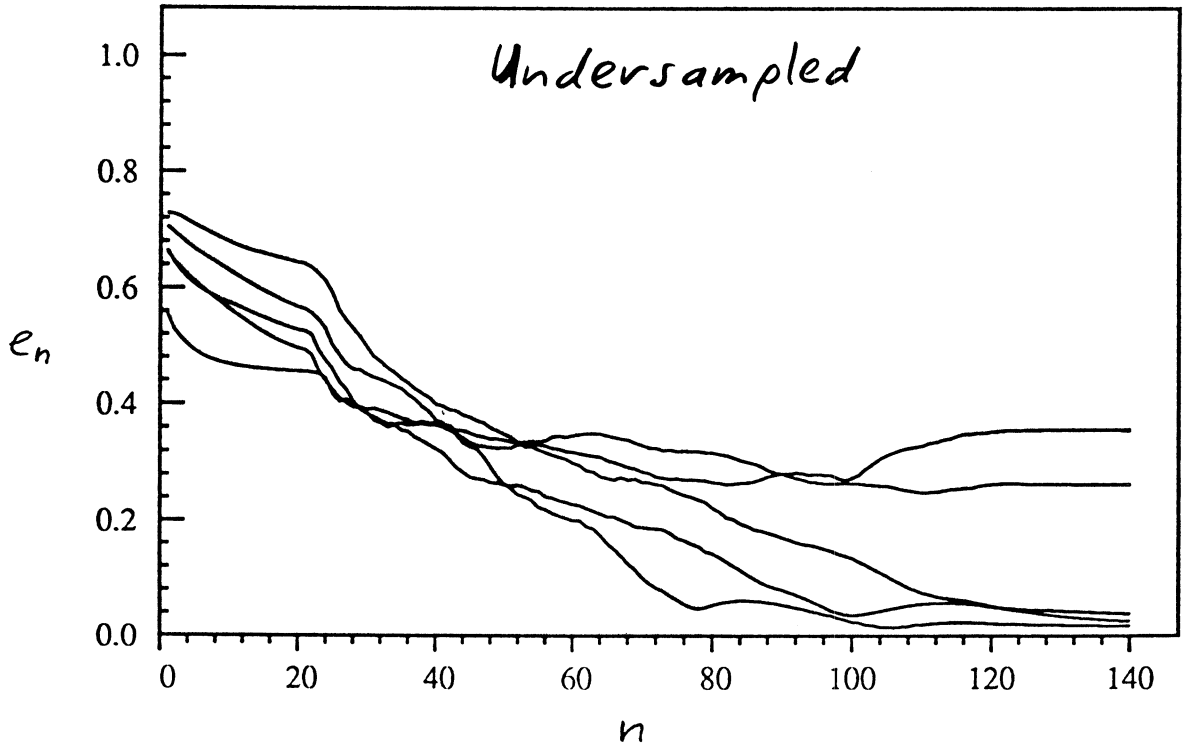
(b)

Fig. 2

3D 8×8×4 image

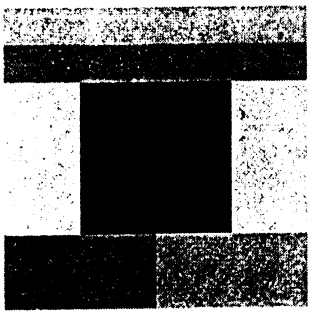


(a)

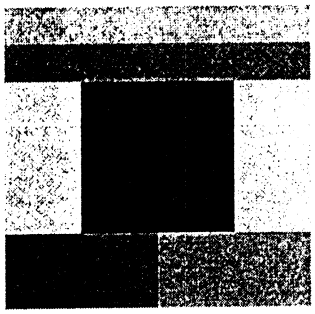


(b)

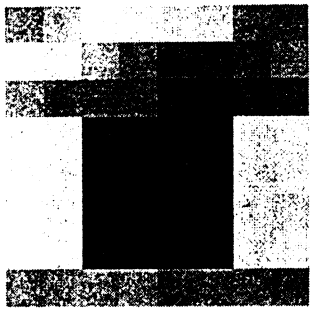
Fig. 1



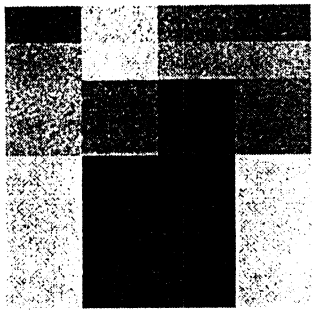
(a)



(b)

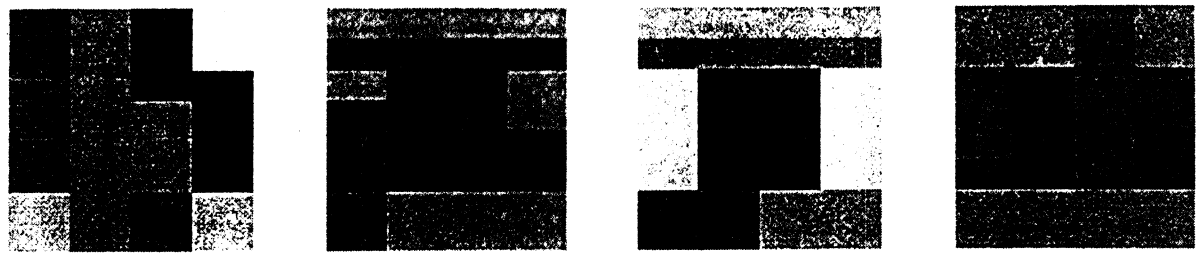


(c)

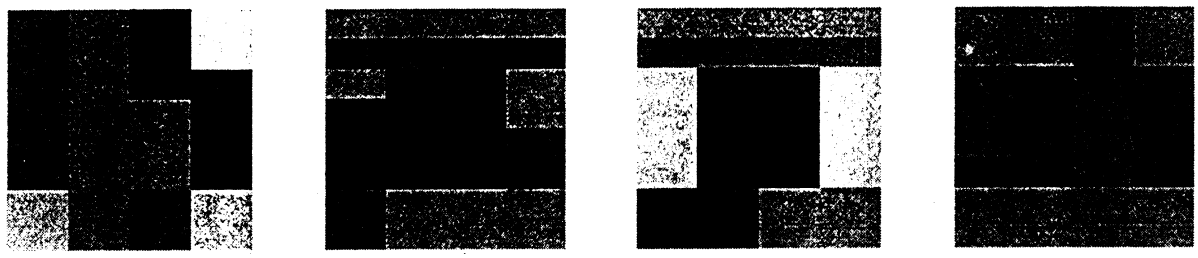


(d)

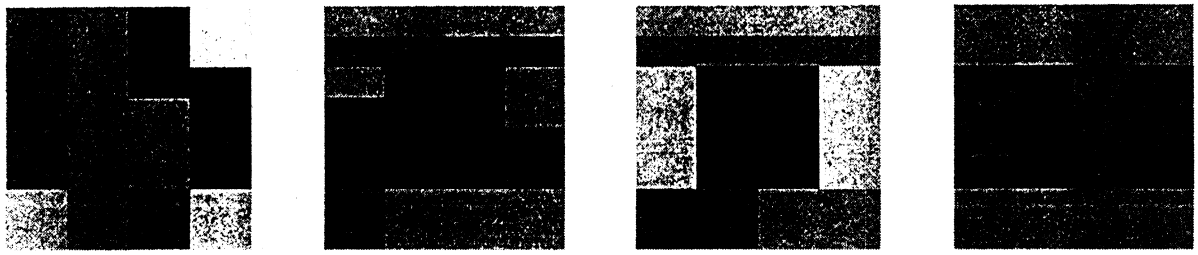
Fig. 4



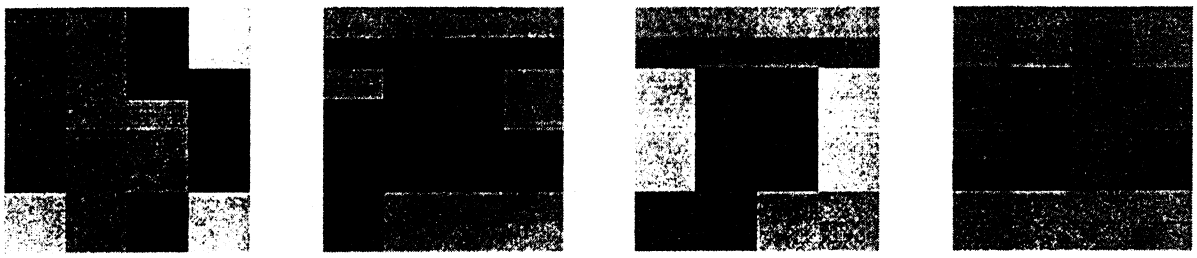
(a)



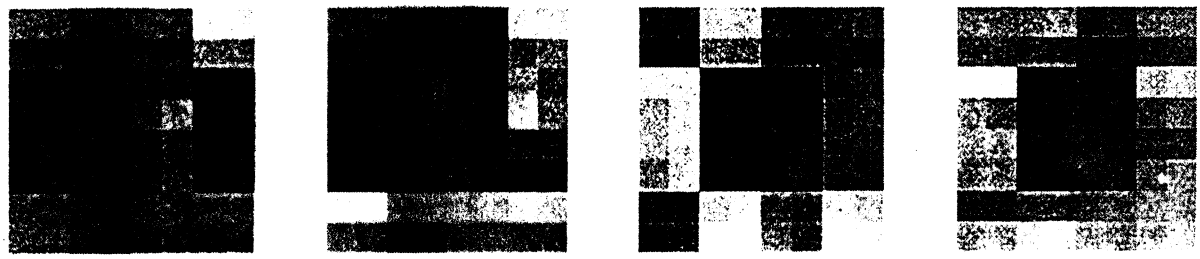
(b)



(c)



(d)



(e)

Fig. 3

THREE-DIMENSIONAL CASE

Fully sampled data

- Phase problem is overdetermined
- Phase retrieval is more stable
- Faster convergence

Undersampled data (up to a factor of two)

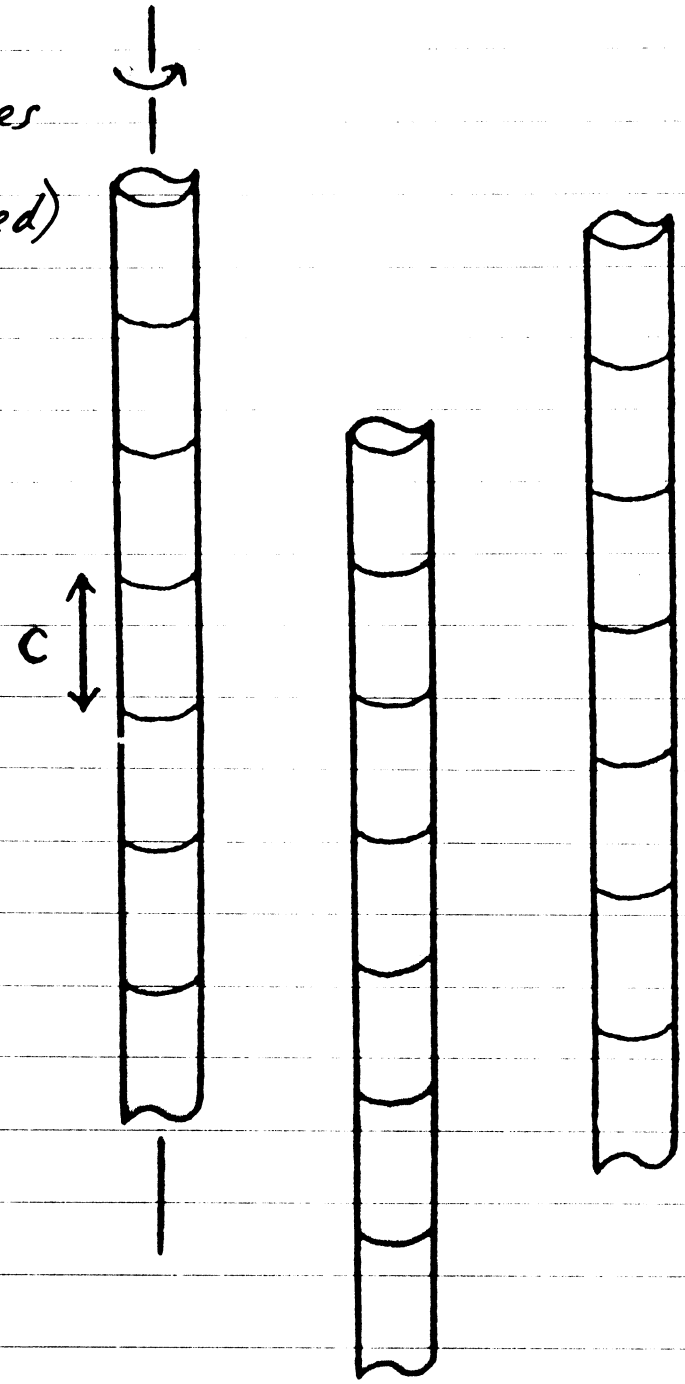
- Phase problem is still unique
- Conventional algorithms are effective

FIBER DIFFRACTION

Periodic particles

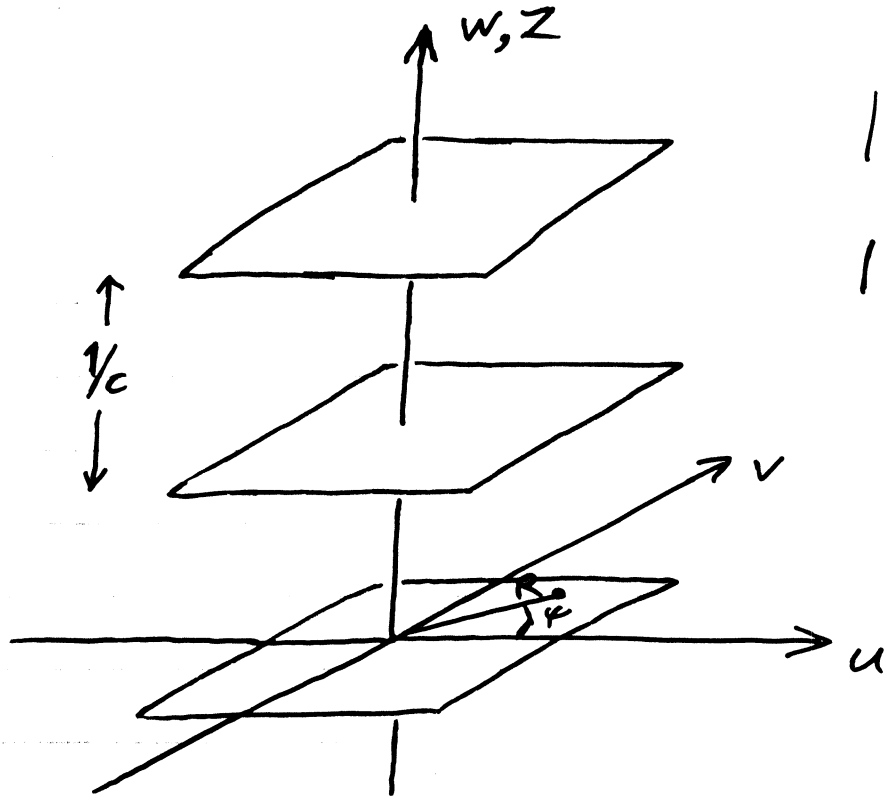
Aligned (oriented)

Cylindrically disordered



Electron density (image) $f(r, \phi, z)$

Fourier transform $F(R, \psi, z) \rightarrow F(R, \psi, \frac{\ell}{c})$

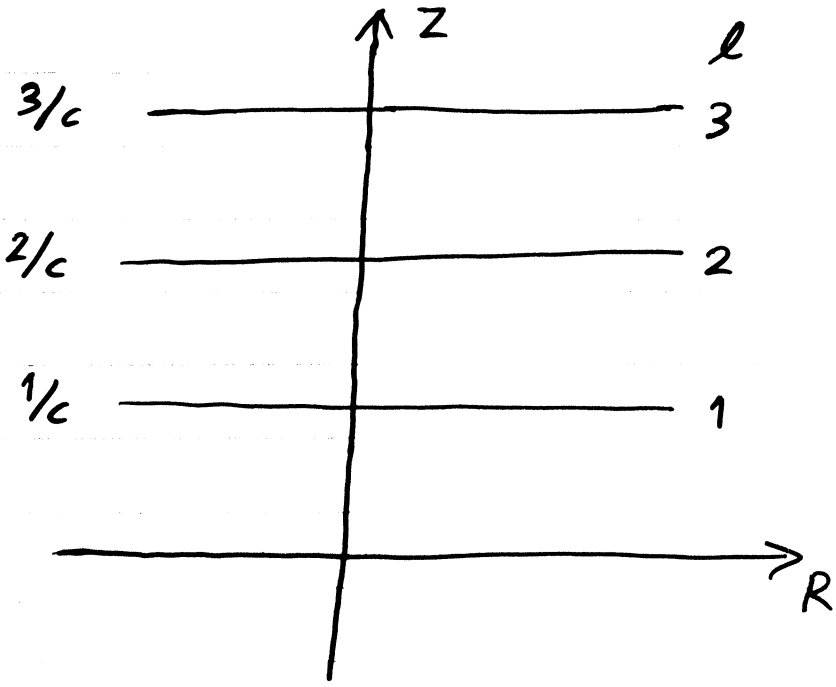


$$|F(u,v,\tfrac{\ell}{c})|^2$$

$$|F(R,\psi,\tfrac{\ell}{c})|^2$$

$$I_\ell(R) = \int_0^{2\pi} |F(R,\psi,\tfrac{\ell}{c})|^2 \, d\psi$$

$$= \sum_n |G_{n\ell}(R)|^2$$



APPLICATION OF THE HIO ALGORITHM TO PROTEIN (VIRUS) CRYSTALLOGRAPHY

Ab initio phasing

Underdetermined by a factor of four

Applied to an icosahedral virus that has five-fold noncrystallographic symmetry

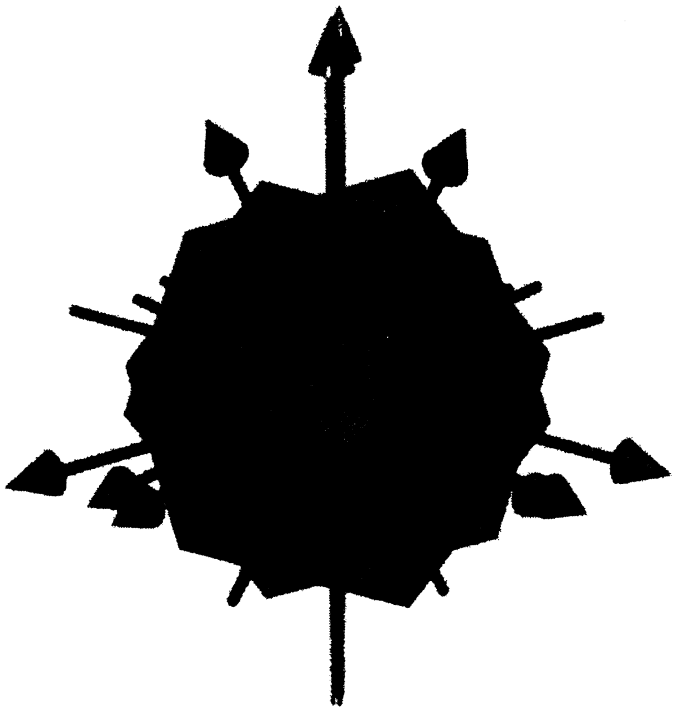
Virus diameter = 320 Å

Cubic unit cell of dimensions 350 Å

Constraints: Spherical support of 340 Å diameter

5-fold noncrystallographic symmetry

Positivity



The simulation

We used synthetic data from the cowpea mosaic virus (cpmv), an icosahedral virus with a 320 Å diameter.

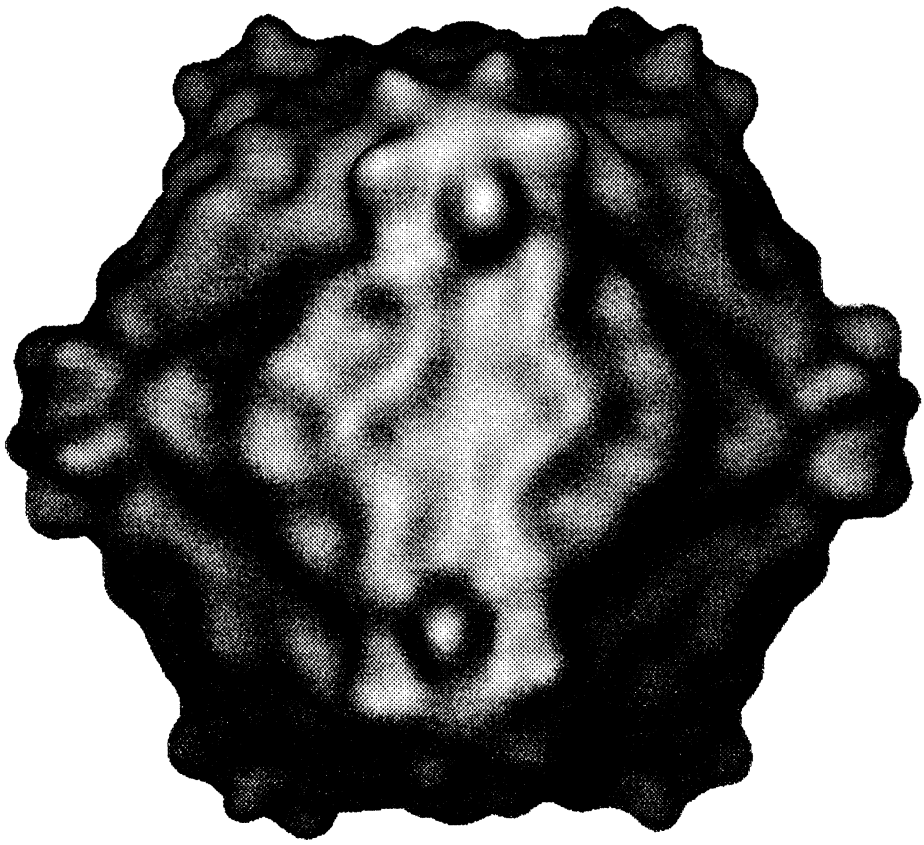
Synthetic data:

* 5% (S/N) noise
* reflections up to 100 Å missing

Two refinements:

* just ER cycles (reference)
* combination of ER and HIO

From 20A to 8Å resolution



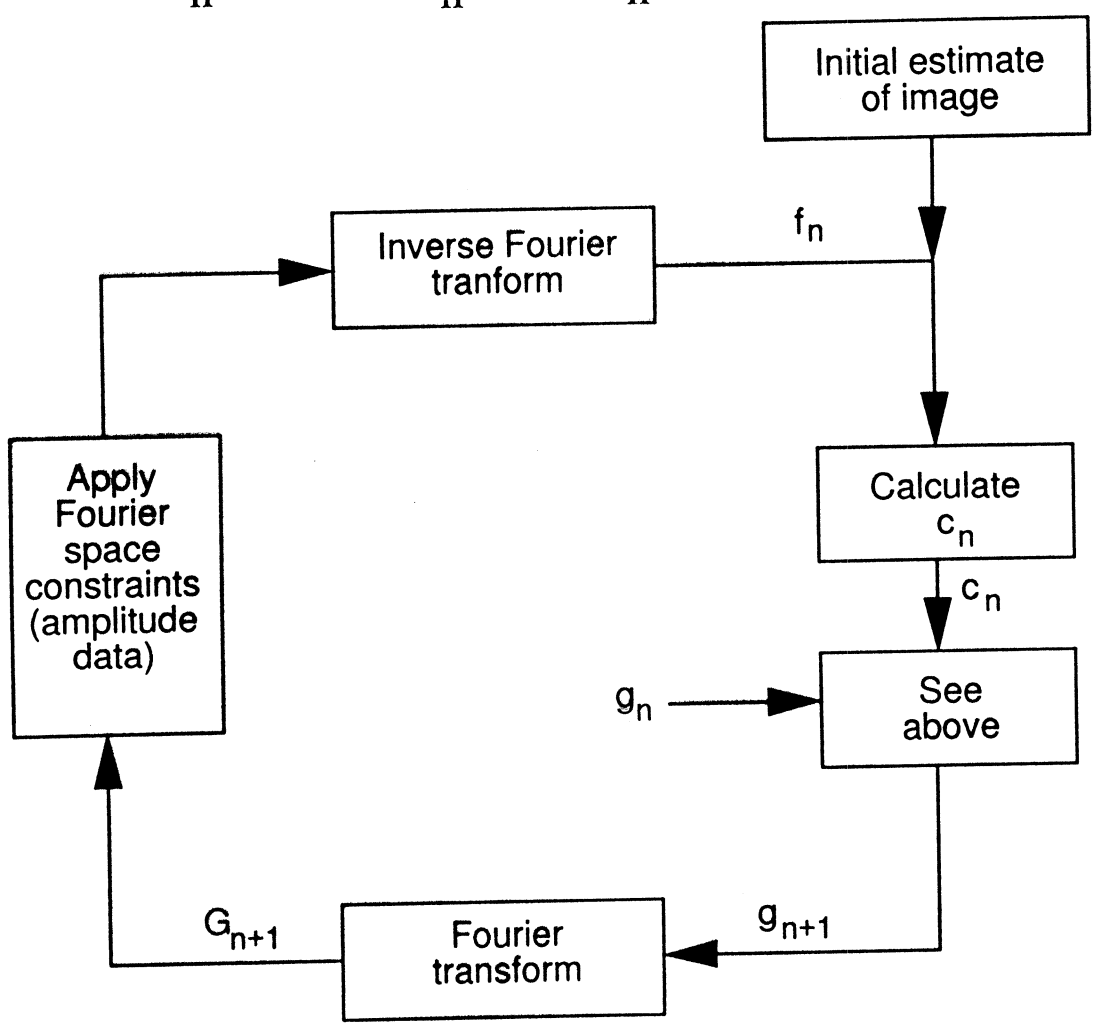
HIO Algorithm

Millane & Stroud, J. Opt. Soc. Am. A, 14, 568-579 (1997)

The hybrid input/output algorithm uses both the constraint values and the estimate from the previous iteration to improve the image

The object is to drive the reconstruction from fixed points of the er algorithm

$$g_{n+1}(\mathbf{x}) = f_n(\mathbf{x}) \qquad \textit{if } f_n(\mathbf{x}) = c_n(\mathbf{x})$$
$$= g_n(\mathbf{x}) - \beta\,(f_n(\mathbf{x}) - c_n(\mathbf{x})) \qquad \textit{otherwise}$$



Phase extension

Reconstruct image at low resolution

Include higher resolution amplitudes with random phases and reconstruct higher resolution image.

Keep 'stepping out' in resolution bins until maximum resolution is reached.

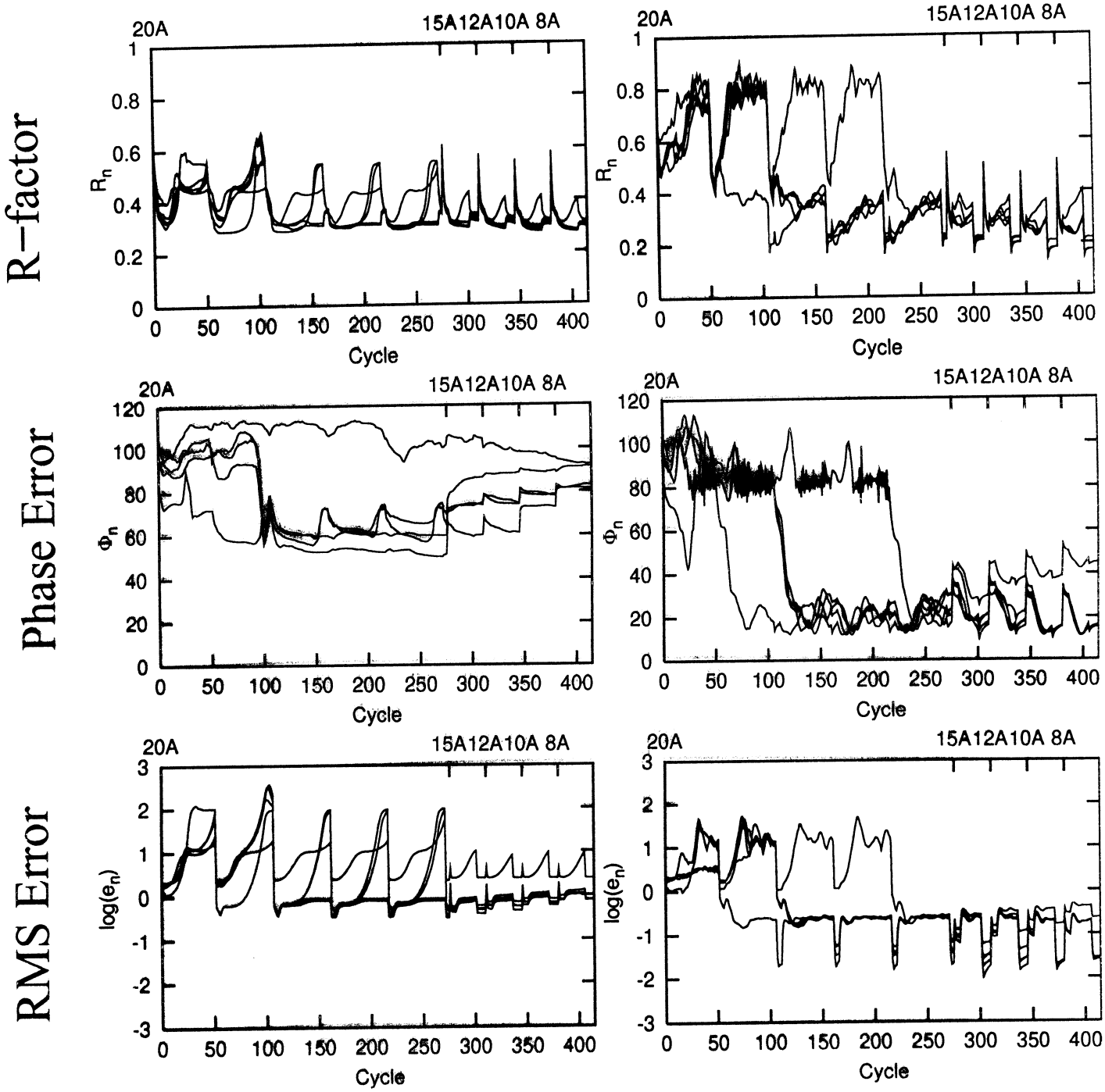
Missing reflections

Often the low order reflections are not measured

Obtain initial estimates of the amplitudes from the mask

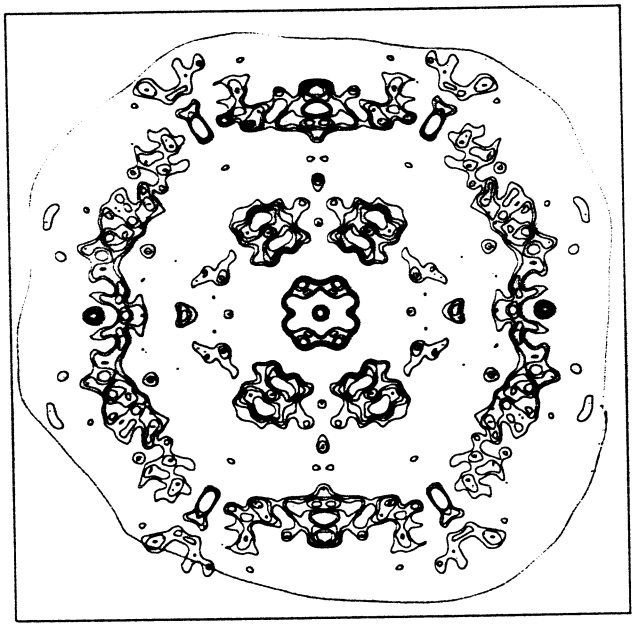
Optionally update these during refinement from (improved) electron density estimates.

Results (error metrics)

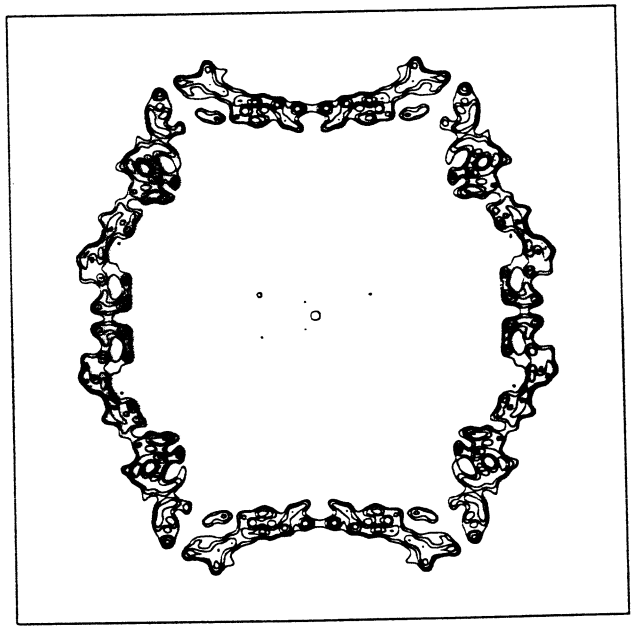


ER HIO

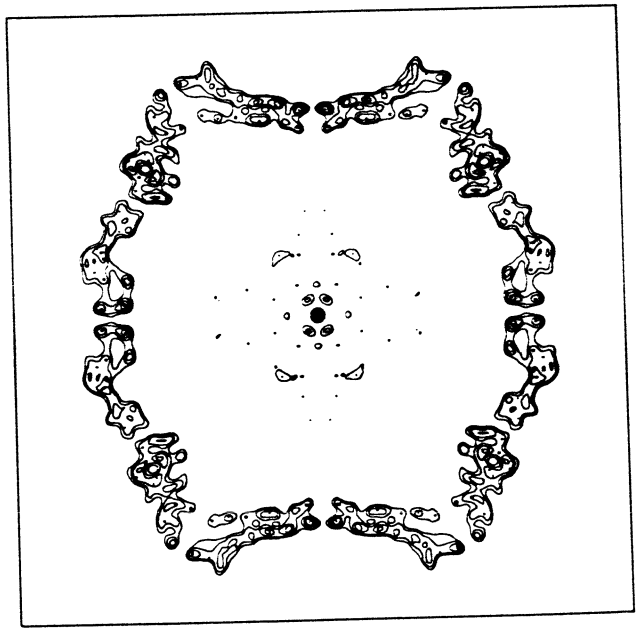
Results: contour plot



ER



True Image



HIO