Phasing Methods for Structure Completion in Crystallography

D. K. Saldin

Department of Physics

University of Wisconsin-Milwaukee

Collaborators

- R. J. Harder, V. L. Shneerson
  (University of Wisconsin-Milwaukee)
- W. Moritz, H. Vogler
  (University of Munich)
- D. L. Wild
  (Keck Graduate Institute)
- I. K. Robinson
  (University of Illinois at Urbana-Champaign)

Structure Completion

Measured Intensities $I_{\mathbf{g}} = |A_{\mathbf{g}}|^2$

where $A_{\mathbf{g}} = R_{\mathbf{g}} + O_{\mathbf{g}}$

$$R_{\mathbf{g}} = \sum_j n_j \exp(i\mathbf{g}.\mathbf{r}_j)$$

$$O_{\mathbf{g}} = \sum_j u_j \exp(i\mathbf{g}.\mathbf{r}_j)$$

Problem

Find $\{u_j\}$ (or equivalently $O_{\mathbf{g}}$)

given $\{I_{\mathbf{g}}\}$ and $\{R_{\mathbf{g}}\}$

Analogous to *holography*

A. Szoke, Phys. Rev. B 47, 14044 (1993)

Direct Solution for Unknown Electron Density

$$I_{\mathbf{g}} = \left|F_{\mathbf{g}}\right|^2 = \left|R_{\mathbf{g}} + O_{\mathbf{g}}\right|^2$$

Since

$$O_{\mathbf{g}} = \sum_j u_j \exp\left(i\mathbf{g}.\mathbf{r}_j\right)$$

$$I_{\mathbf{g}} - \left|R_{\mathbf{g}}\right|^2 = 2\sum_j u_j \mathrm{Re}\left[R_{\mathbf{g}}^* \exp\{i\mathbf{g}.\mathbf{r}_j\}\right] + \sum_{j,l} u_j u_l \exp\{i\mathbf{g}.(\mathbf{r}_j - \mathbf{r}_l)\} \qquad \forall \mathbf{g}$$

Methods for solving these Eqns. directly have been given by e.g.:

A. Szöke, *Phys. Rev. B* **47**, 14044 (1993)

D. K. Saldin *et al.*, *Phys. Rev. Lett.* **70**, 1112 (1993)

A. Szöke, *Acta Cryst.* A**53**, 291 (1997)

Solution by Phasing Methods

$$u_j = \frac{1}{N} \sum_{\mathbf{g}} \left[\left| F_{\mathbf{g}} \right| \exp\{i\,\phi_{\mathbf{g}}\} - R_{\mathbf{g}} \right] \exp\left(-i\,\mathbf{g} \cdot \mathbf{r}_j \right)$$

Problem is that $\phi_{\mathbf{g}}$ is not known from experiment (phase problem)

Unweighted Difference Fourier

W. Cochran, *Acta Cryst.* **4**, 408 (1951)

$$u_j = \frac{1}{N} \sum_{\mathbf{g}} \left[\left| F_{\mathbf{g}} \right| \exp\{ i\, \phi_{\mathbf{g}}^{(R)} \} - R_{\mathbf{g}} \right] \exp(-i\, \mathbf{g} \cdot \mathbf{r}_j)$$

$$\phi_{\mathbf{g}}^{(R)} = \arg(R_{\mathbf{g}})$$

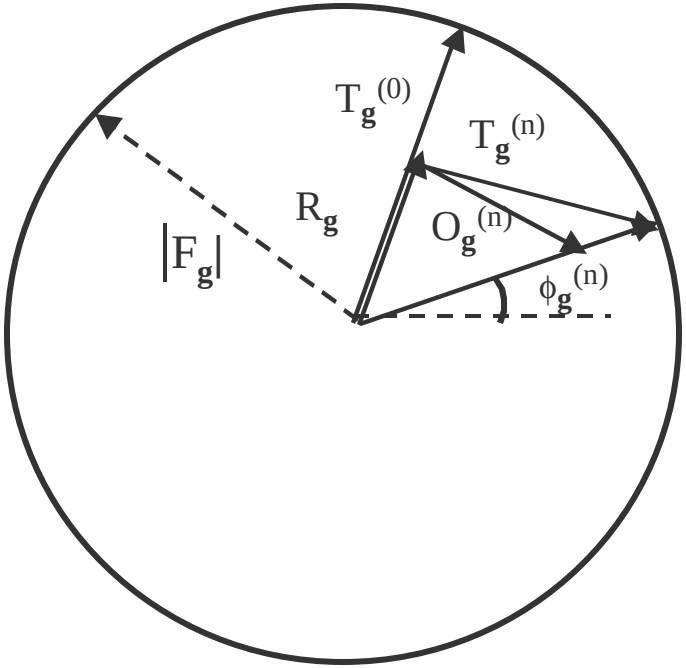
Weighted Difference Fourier

G. Sim, *Acta Cryst.* **12**, 813 (1959); *ibid.* **13**, 511 (1960)

$$u_j = \frac{1}{N} \sum_{\mathbf{g}} \left[w_{\mathbf{g}} \left| F_{\mathbf{g}} \right| \exp\{ i\, \phi_{\mathbf{g}}^{(R)} \} - R_{\mathbf{g}} \right] \exp(-i\, \mathbf{g} \cdot \mathbf{r}_j)$$

$$w_{\mathbf{g}} = f(X_{\mathbf{g}}) \qquad X_{\mathbf{g}} = 2 \left| F_{\mathbf{g}} \right| \left| R_{\mathbf{g}} \right| \Big/ \sum_j f_j^2$$

Fienup Algorithm for Structure Completion



$\{u_j^{(n)}\}$

$O_g^{(n)}=FT_g\{u_j^{(n)}\}$

$\phi_g^{(n)}=\arg[R_g+O_g^{(n)}]$

$T_g^{(n)}=|F_g|\exp[i\ \phi_g^{(n)}]-R_g$

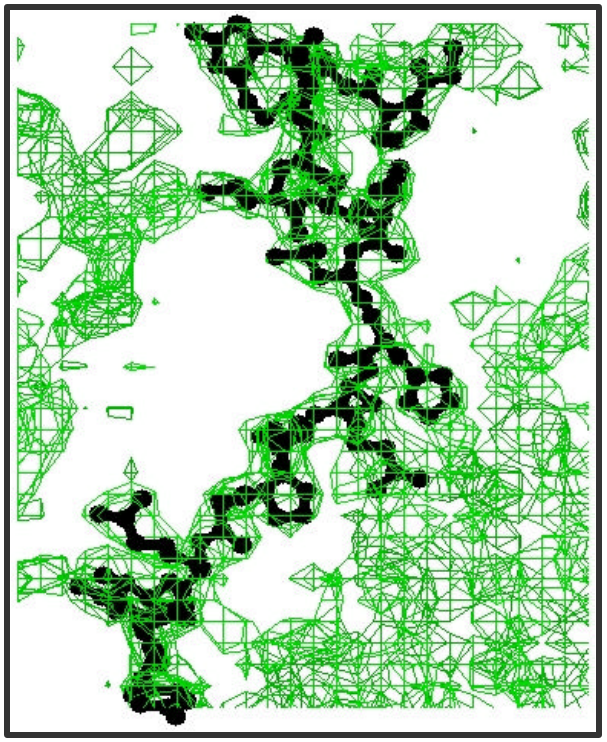
$u'^{(n)}_j=FT^{-1}_j\{T_g^{(n)}\}$

$\{u'^{(n)}_j\}$

$u_j^{(n+1)}=f(u_j^{(n)},t_j^{(n)})$

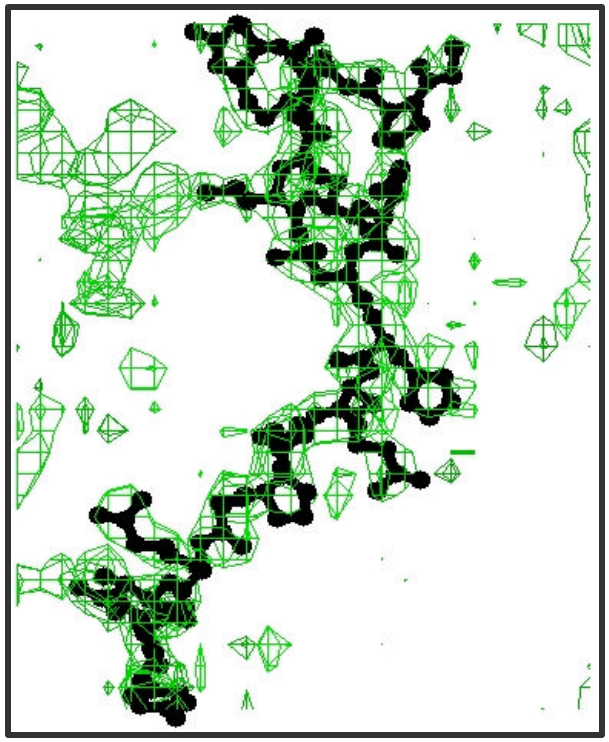
$n \Rightarrow n+1$

TOXD: Recovery of residues 1-18



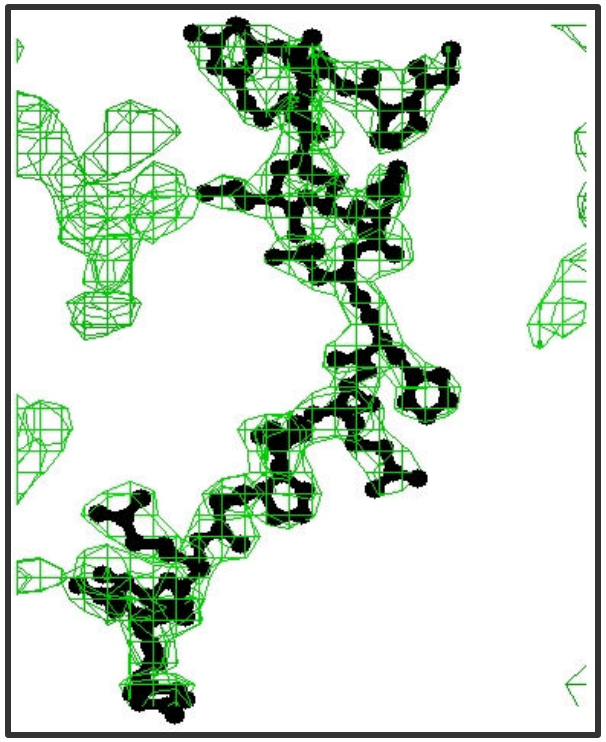
DF

1.3σ+mean



Fienup

2.2σ+mean

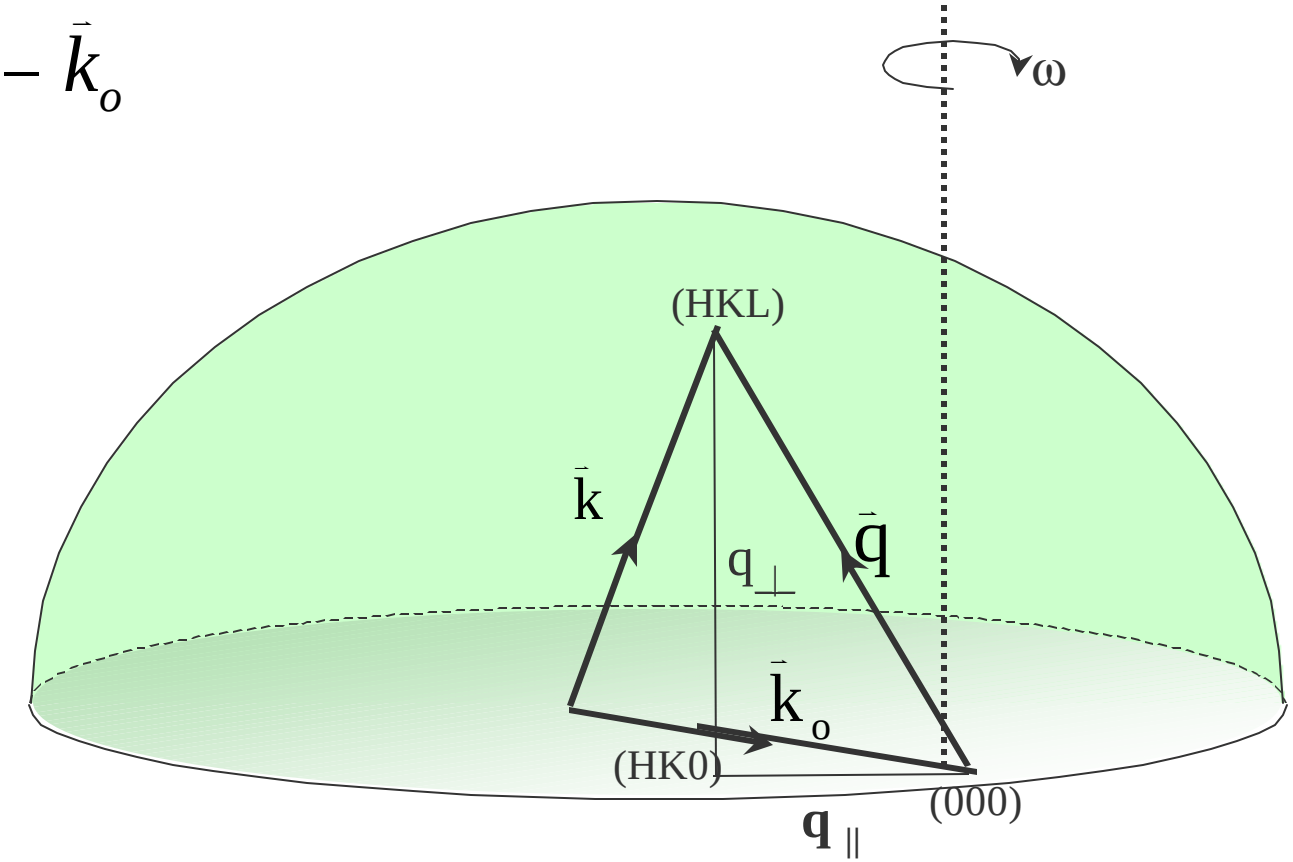


Exact

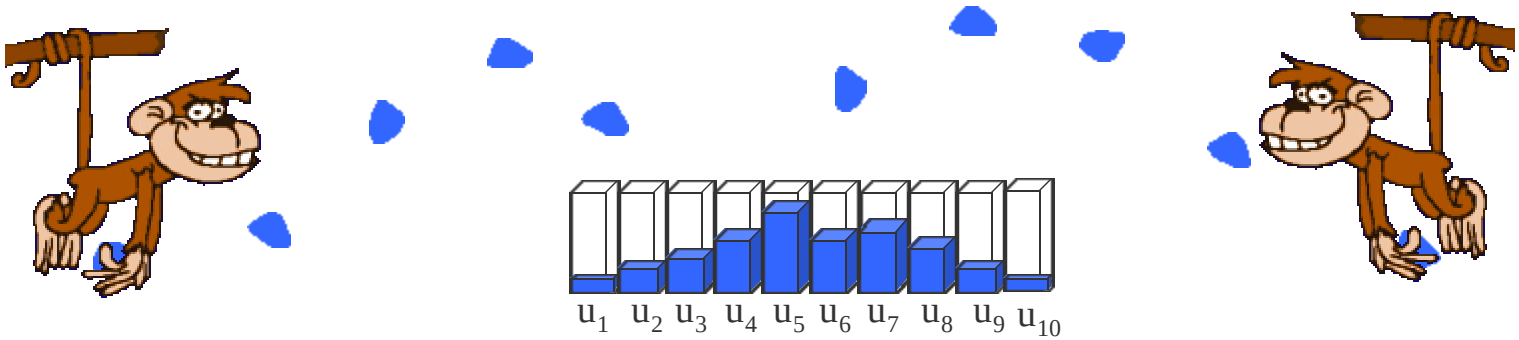
2.9σ+mean

Ewald sphere construction for surface x-ray diffraction

$$\vec{q} = \vec{k} - \vec{k}_o$$



Prior Probability of a Distribution and its Entropy



Consider a team of monkeys (unbiased individuals) throwing objects at random into boxes 1,2,3,... of capacities m_1, m_2 m_3, respectively.

The probability of particular distribution $\{u_1, u_2, u_3, ..\}$ is proportional to number of ways it can come about:

$$P(\{u_j\}) = Ù = \prod_j m_j^{u_j} \Big/ \prod_j u_j!$$

But

$$S = k \log Ù$$

Therefore

$$P(\{u_j\}) \propto Ù = \exp(S/k)$$

Hence most probable distribution is that of the maximum entropy

Boltzmann and Gibbs Forms of the Entropy

$$S = k\ln(\Omega)$$ Boltzmann

$$\grave{U} = \prod_j m_j^{u_j} \Big/ \prod_j u_j!$$

Stirling's Approximation $$\ln(u_j!) = u_j \ln(u_j) - u_j$$

$$\ln(\Omega) = \sum_j u_j \ln(m_j) - \sum_j u_j \ln(u_j) + \sum_j u_j \ln(e)$$

$$S = -k\sum_j u_j \ln(u_j / e m_j)$$ Gibbs

Maximization of the Constrained Entropy

$$Q = -\sum_{l=1}^{N} u_l^{(n+1)} \ln[u_l^{(n+1)}/e u_l^{(n)}] - \frac{\ddot{e}}{2} \sum_{\mathbf{q}} \left| |F_{\mathbf{q}}| e^{i\ddot{o}_{\mathbf{q}}^{(n)}} - [R_{\mathbf{q}} + FT_{\mathbf{q}}\{u_l^{(n+1)}\}] \right|^2$$

$$\partial Q[\{u_l^{(n+1)}\}]/\partial u_j^{(n+1)} = -\ln[u_j^{(n+1)}/u_j^{(n)}] - \ddot{e}[u'^{(n)}_j - u_j^{(n+1)}] = 0$$

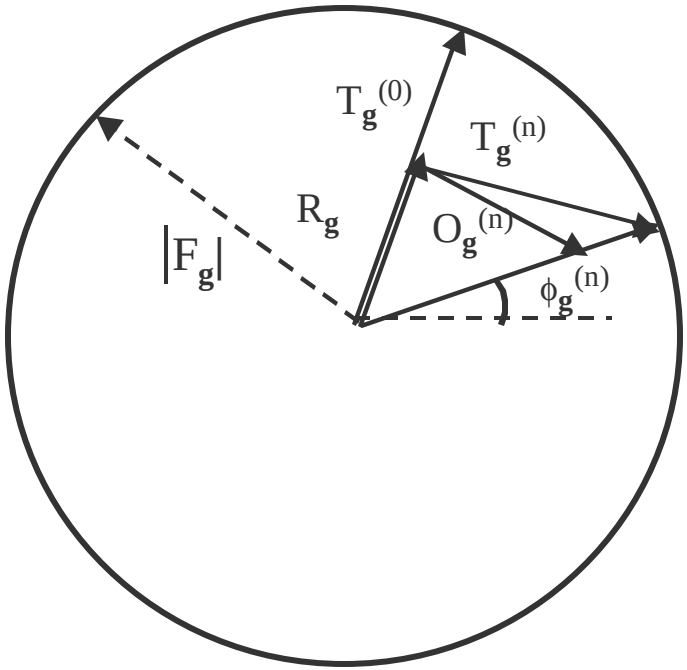
$$u_j^{(n+1)} = u_j^{(n)} e^{-\ddot{e}\{u_j^{(n)} - u'^{(n)}_j\}}$$

Exponential Modeling Algorithm: Collins *Nature* **298**, 49 (1982)

$$u'^{(n)}_j = FT_j^{-1}[|F_{\mathbf{q}}| e^{i\ddot{o}_{\mathbf{q}}^{(n)}} - R_{\mathbf{q}}]$$

$$\ddot{o}_{\mathbf{q}}^{(n)} = \arg\{R_{\mathbf{q}} + FT_{\mathbf{q}}(u_j^{(n)})\}$$

Exponential Modeling Algorithm for Structure Completion



$\{u_j^{(n)}\}$ → $O_g^{(n)}=FT_g\{u_j^{(n)}\}$

$\phi_g^{(n)}=\arg[R_g+O_g^{(n)}]$

$T_g^{(n)}=|F_g|\exp[i\ \phi_g^{(n)}]-R_g$

$u'^{(n)}_j=FT^{-1}_j\{T_g^{(n)}\}$ → $\{u'^{(n)}_j\}$

$u_j^{(n+1)}=u_j^{(n)}\exp[-\lambda\{u_j^{(n)}-u'^{(n)}_j\}]$

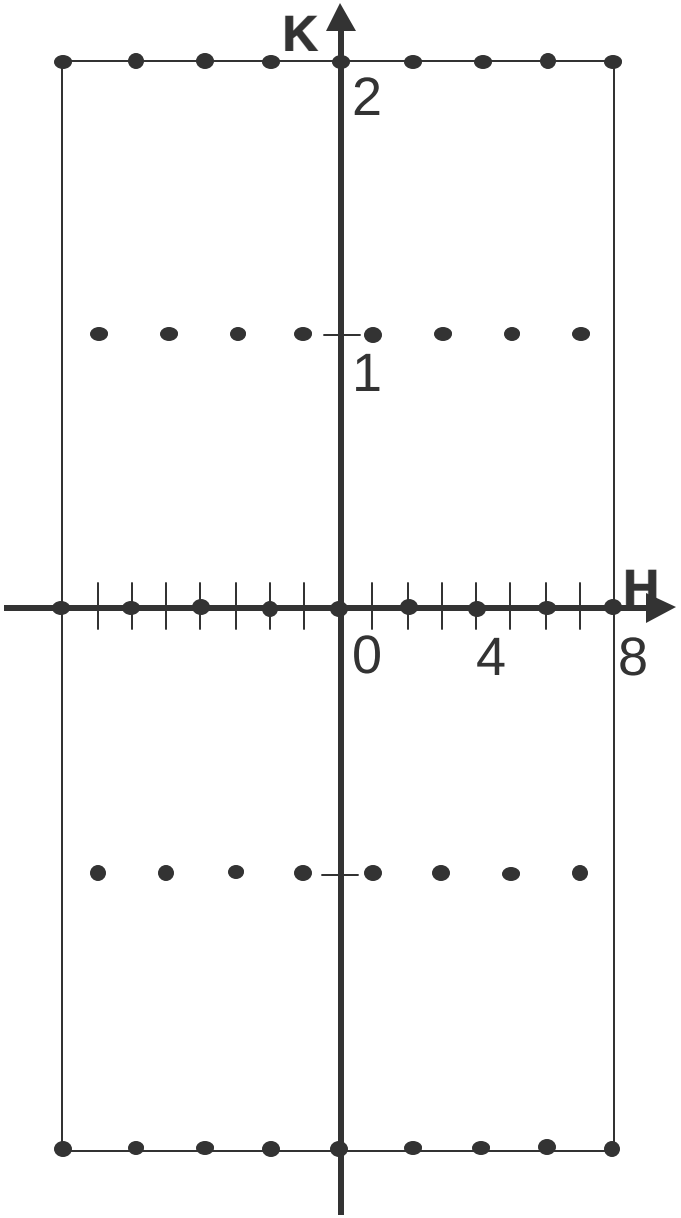
$n \Rightarrow n+1$

Proteins: Shneerson, Wild, Saldin *Acta Cryst.* A**57** 163 (2001)

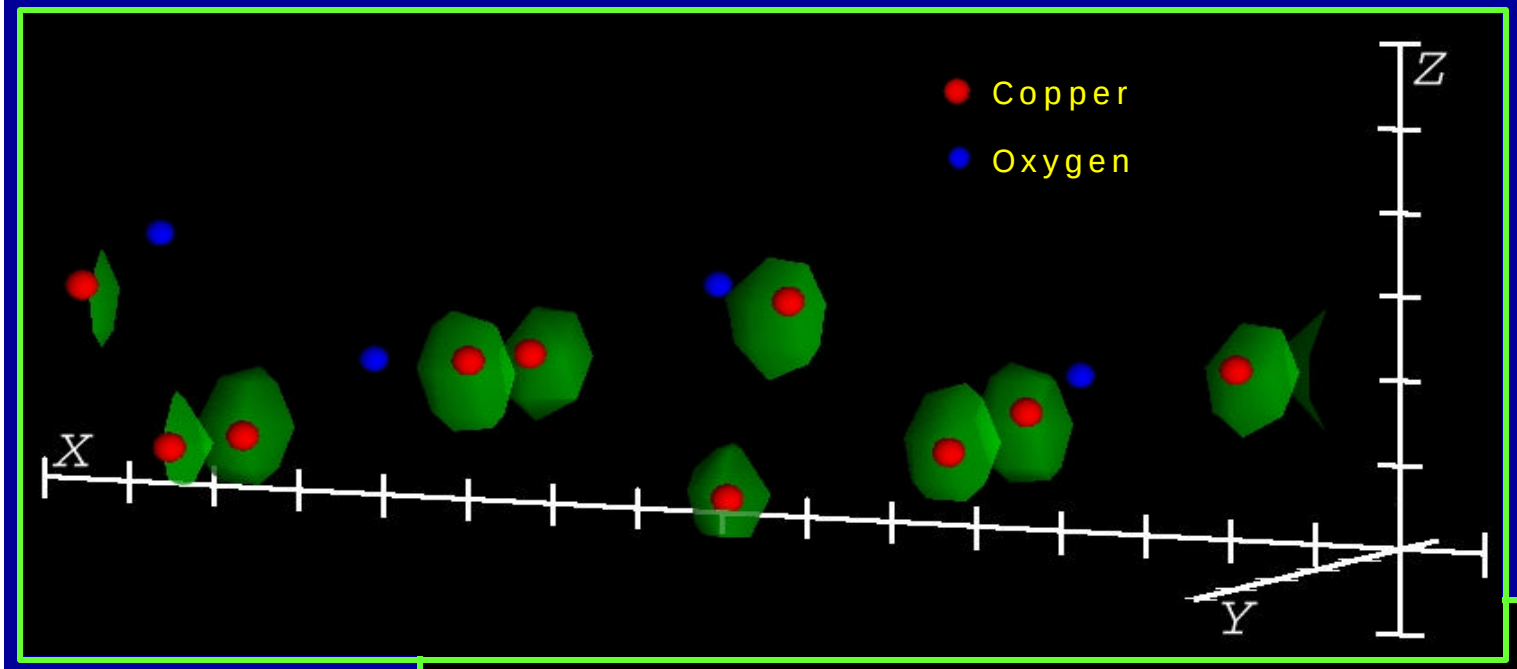
Surfaces: Saldin, Harder, Vogler, Moritz, Robinson *Comput. Phys. Commun.* **137** 12 (2001)

Diffraction pattern

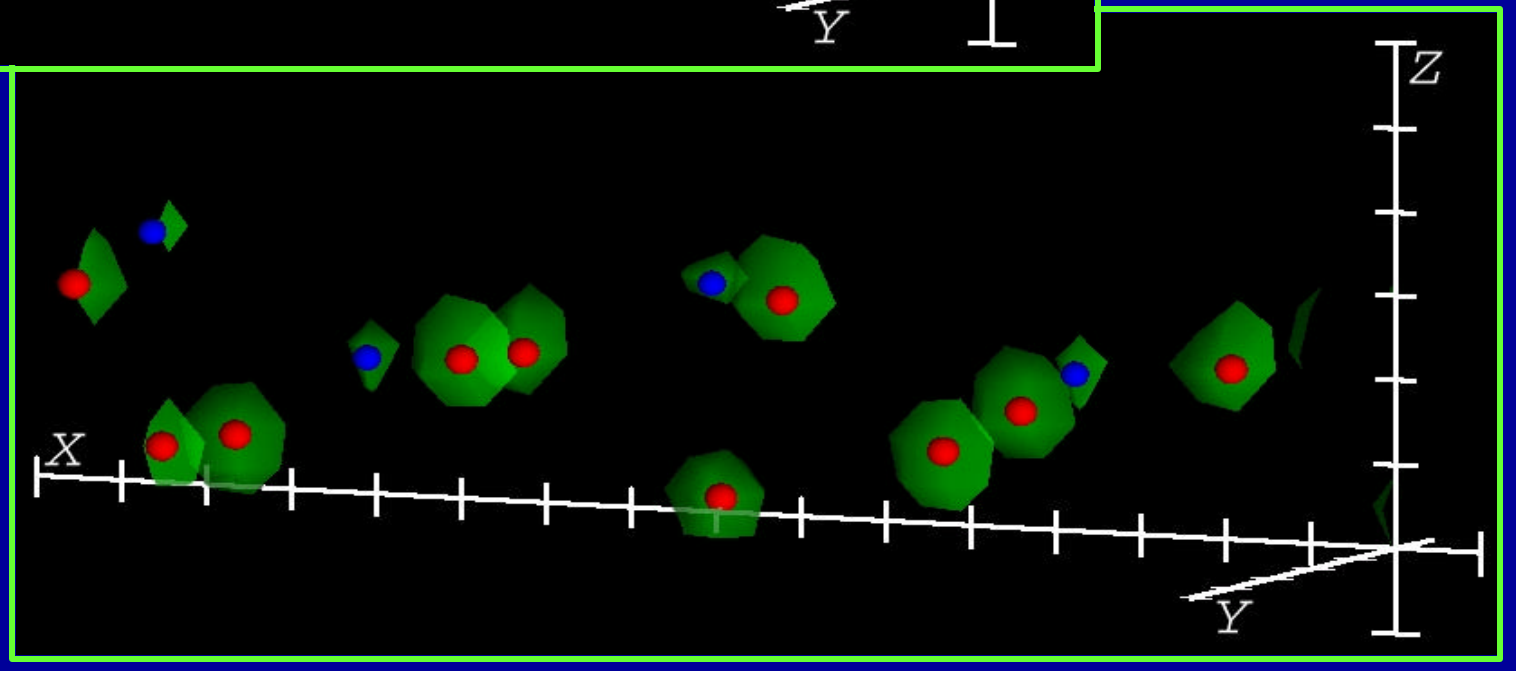
O/Cu(104)



O/Cu(104)



Difference Fourier

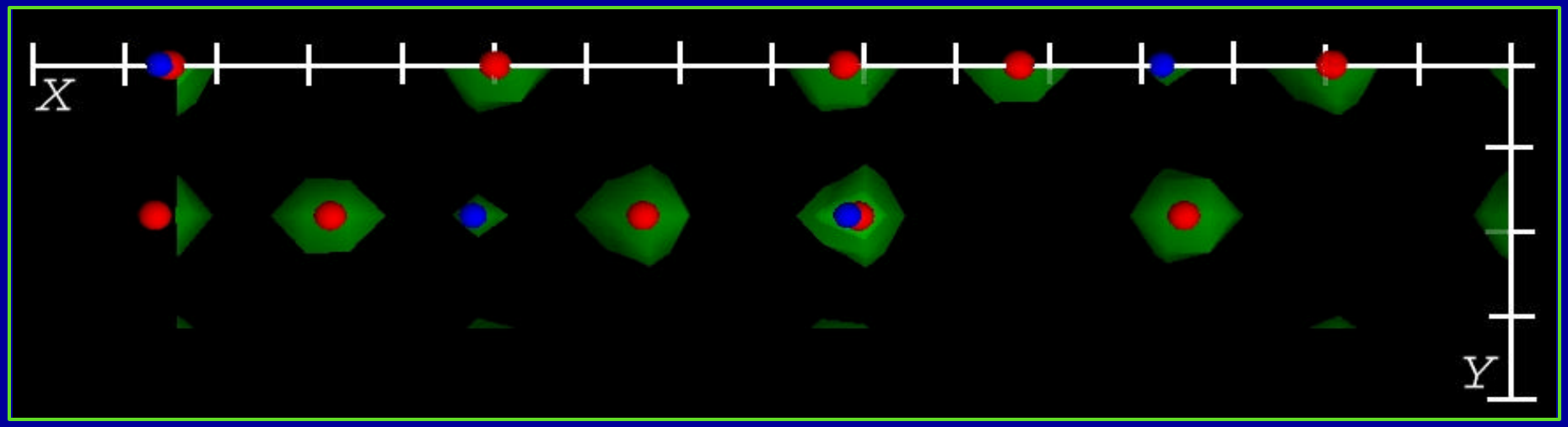


Exponential Modeling

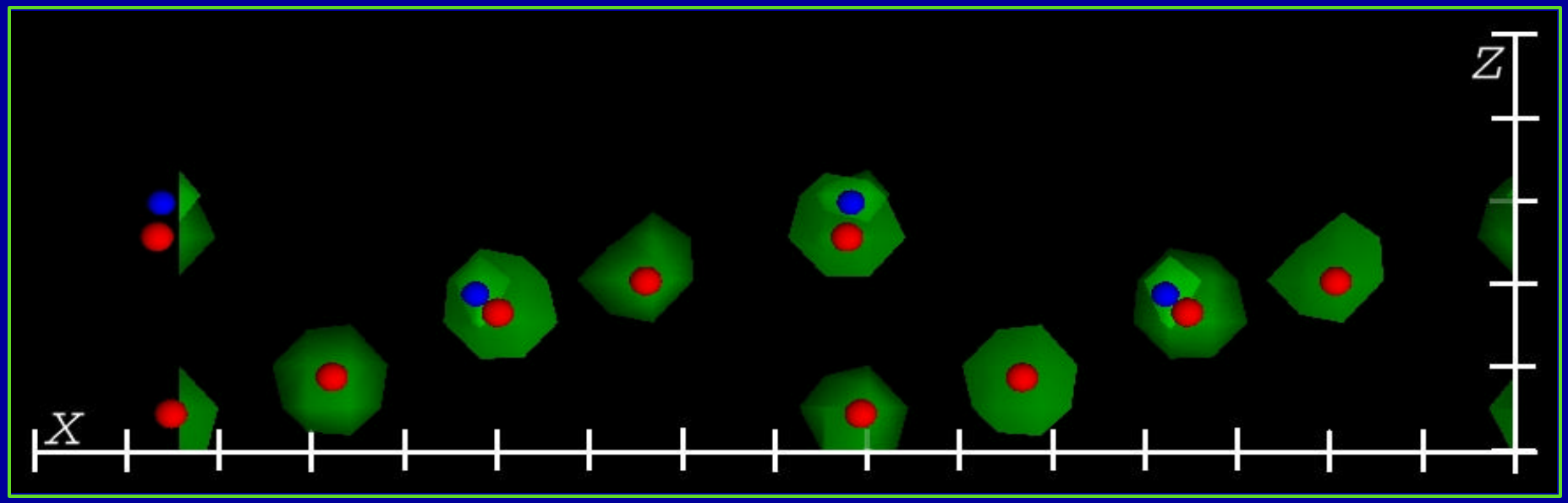
O/Cu(104)

Exponential Modeling

Top View

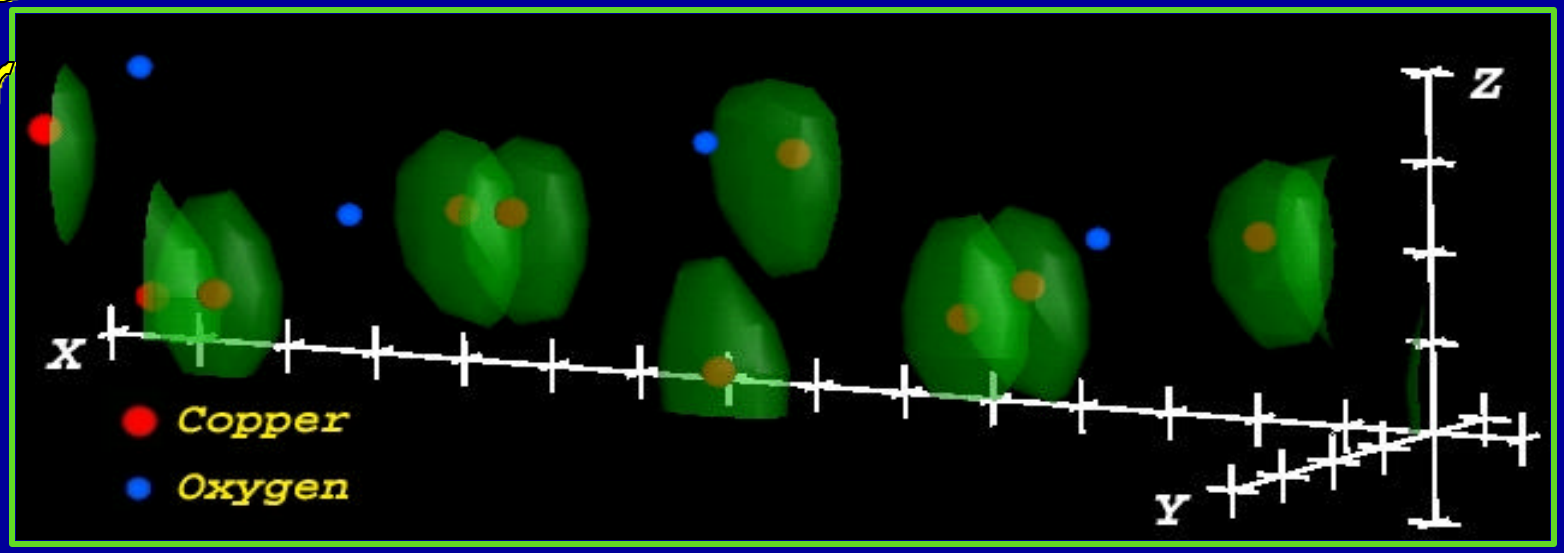


Side View

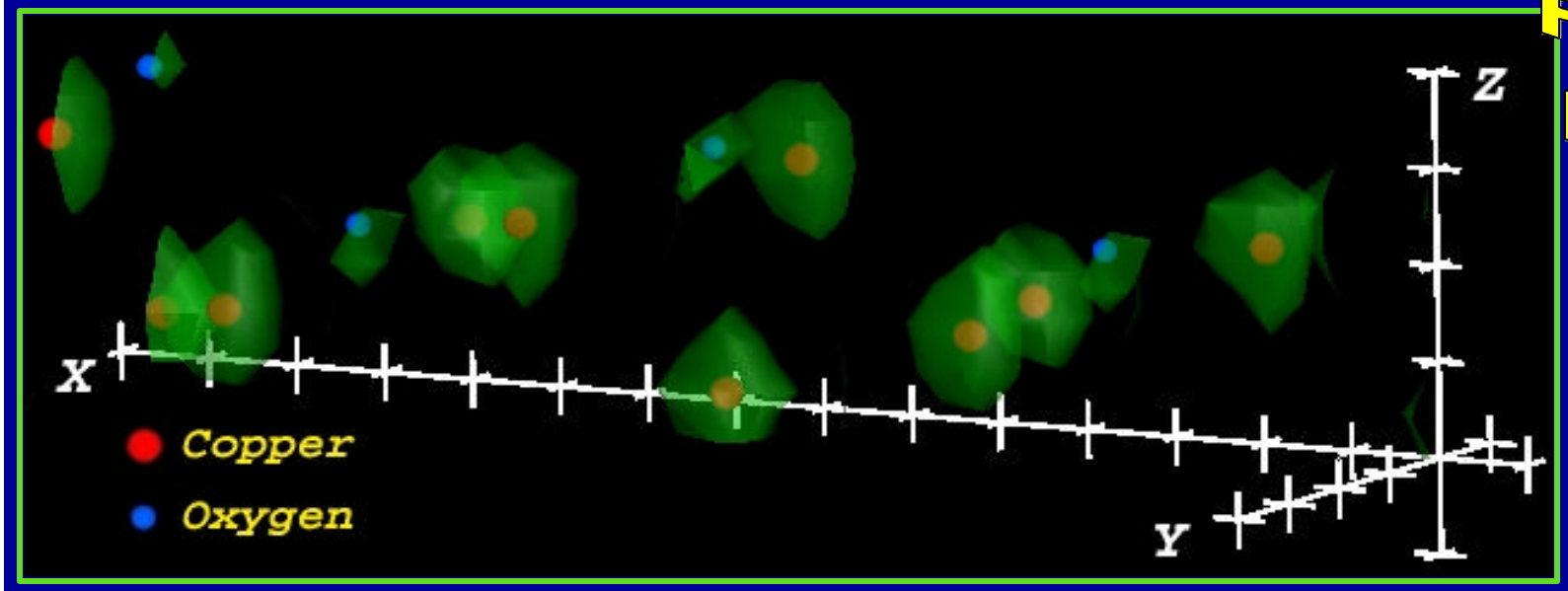


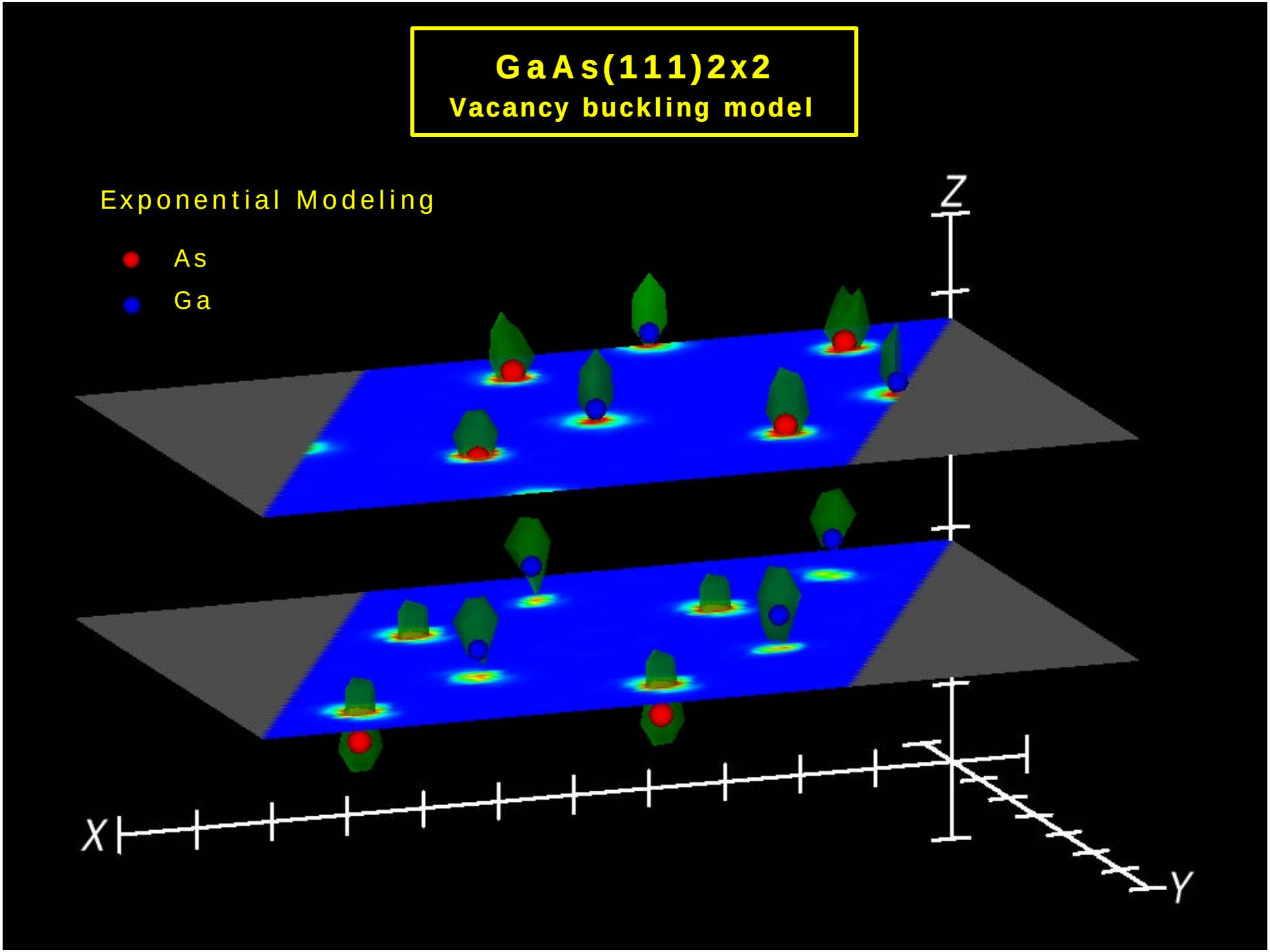
O / Cu(104)

Difference Fourier



Fienup Result




GaAs(111)2x2
Vacancy buckling model
Exponential Modeling
As
Ga
Z
X
Y

Conclusions

1. The Fienup algorithm may be adapted very effectively to the *structure completion* problem in protein and surface crystallography.
2. Collins' exponential modeling algorithm may be re-cast as a Fienup-type input-output feedback loop between real and reciprocal space.
3. Surface crystallography is by its nature a structure completion problem.
4. Either algorithm seems very effective in recovering the surface electron density from surface x-ray diffraction (SXRD) measurements given a knowledge of the bulk structure.