Single Particle Reconstruction

by Scattering from Many

D. K. Saldin

Department of Physics

University of Wisconsin-Milwaukee

Acknowledgements

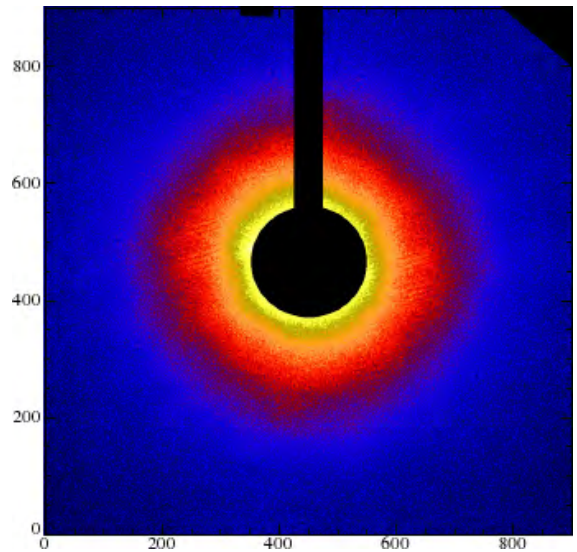
- Valentin Shneerson
- Hin-Cheuck Poon
- Kanupriya Pande
- Peter Schwander
- Sandi Wibowo
- Sung Soon Kim
- Marius Schmidt
- Miraj Uddin

EM Image of Sample for Experiment

Mainly ~ 80 nm x 20 nm randomly oriented metal rods on a SiN substrate

DKS and J C H Spence et al.. PRL 106, 115501 (2011).

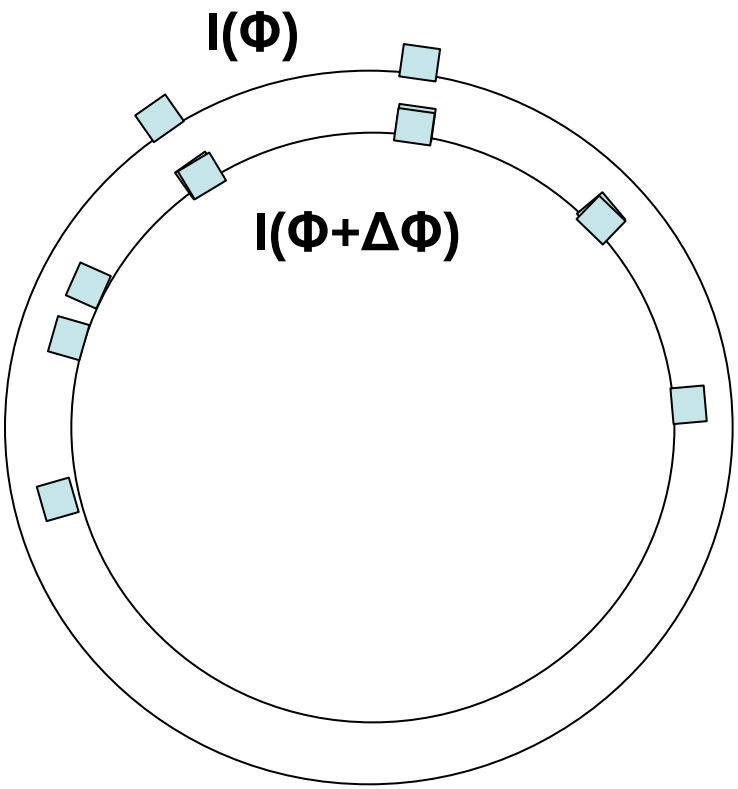
Information In Angular Correlations

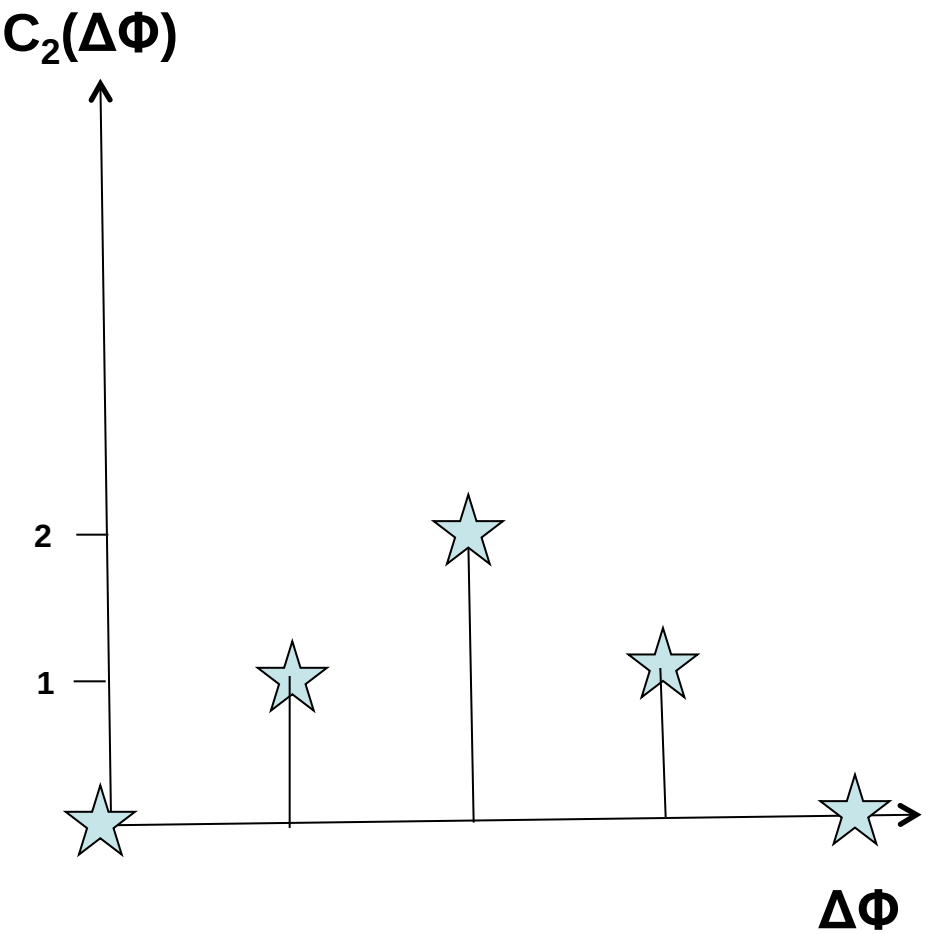


Diffraction pattern from disordered subunits Appears to have no angular structure, only radial variation, studied by SAXS. However there is untapped information in the angular correlations, revealed by evaluating

$$C_2(q;q',\Delta\varphi) = \left\langle \frac{1}{N_\varphi} \sum_j \left\{ I(q,\varphi_j) - I_{saxs}(q) \right\} \left\{ I(q',\varphi_j + \Delta\varphi) - I_{saxs}(q') \right\} \right\rangle_{DP}$$

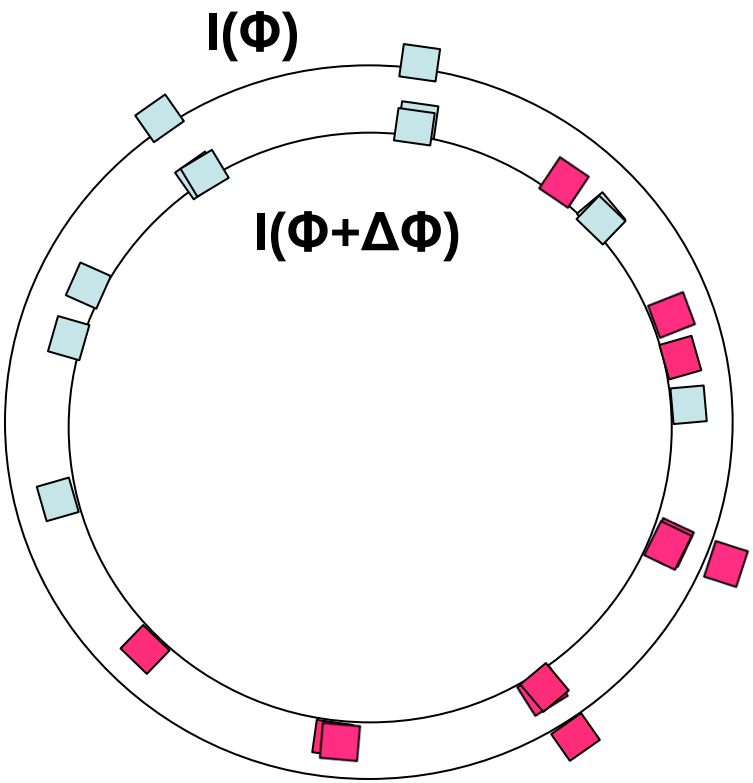
Extraction of Structure from Diffraction Patterns of Randomly Oriented Particles

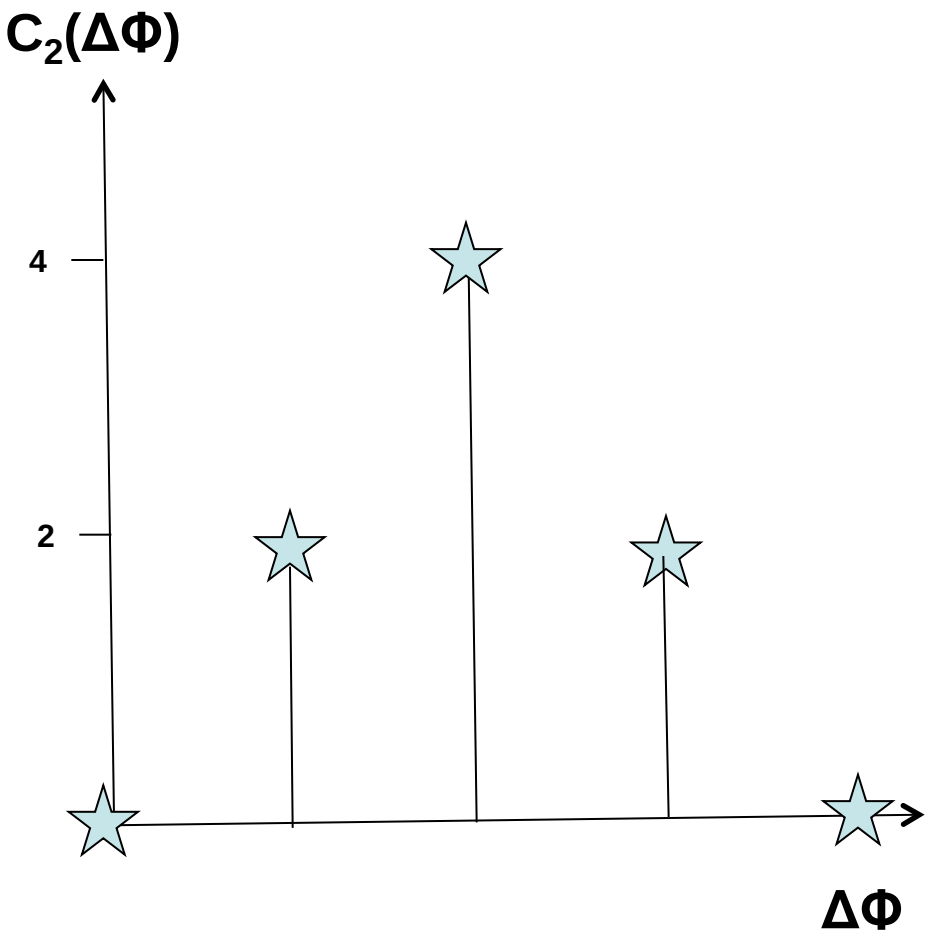




$$C_2(\Delta\Phi) = \sum_{\Phi} I(\Phi) I(\Phi + \Delta\Phi)$$

Method applicable even for Multiple Particles in Radom Orientations





$$C_2(\Delta\Phi) = \sum_{\Phi} I(\Phi) I(\Phi + \Delta\Phi)$$

$C_2 \propto N$ **but otherwise identical**

Reconstructing the Single-Particle Diffraction Pattern

Circular Harmonic Expansion

$$I(q_x, q_y) = \sum_m I_m(q) \exp(im\phi)$$

where

$$q = \sqrt{q_x^2 + q_y^2} \qquad \phi = \tan^{-1}(q_y, q_x)$$

The reality of $I(\mathbf{q})$ ensures that $I_{-m}(q)=I_m(q)$. For a flat Ewald sphere, Friedel's rule, $I(-\mathbf{q})=I(\mathbf{q})$, will be satisfied if only even m's contribute.

If the single particle diffraction pattern has a mirror line, can choose the $I_m(q)$ to be real.

Saldin et al., New J. Phys. **12**, 035014 (2010)

Relation to Scattered Amplitudes

$$A(q,\varphi_j) = \sum_k \sum_m A_m(q) \exp\{im(\varphi_j - \omega_k)\} \exp(i\mathbf{q}.\mathbf{r}_k)$$

$$C_2(q;q',\Delta\varphi) = \left\langle \frac{1}{N_\varphi} \sum_j \{I(q,\varphi_j) - I_{saxs}(q)\}\{I(q',\varphi_j + \Delta\varphi) - I_{saxs}(q')\} \right\rangle_{DP}$$

$$\begin{aligned}
&I(q,\varphi_j) - I_{saxs} \\
&= \sum_{kk'} \sum_{m \neq m'} A_m^*(q) \exp(-im\varphi_j) \exp(im\omega_k) \exp(-i\mathbf{q}.\mathbf{r}_k) A_{m'}(q) \exp(im'\varphi_j) \exp(-im'\omega_{k'}) \exp(i\mathbf{q}.\mathbf{r}_{k'}) \\
&= \sum_{kk''} \sum_{m \neq m'} A_m^*(q) A_{m'}(q) \exp\{-i(m-m')\varphi_j\} \exp(im\omega_k) \exp(-im'\omega_{k'}) \exp\{-i\mathbf{q}.(\mathbf{r}_k - \mathbf{r}_{k'})\} \\
&= \sum_{m,M \neq 0} A_m^*(q) A_{m-M}(q) \exp\{-iM\varphi_j\} \sum_{kk'} \exp(im\omega_k) \exp(-i(m-M)\omega_{k'}) \exp\{-i\mathbf{q}.(\mathbf{r}_k - \mathbf{r}_{k'})\}
\end{aligned}$$

By performing the sums over *amplitudes* we are assuming *coherent* scattering amongst the different particles

Pair Correlations in Terms of Amplitudes

$$I(q',\varphi_j+\Delta\varphi_k)-I_{saxs}$$
$$=\sum_{m',M'\neq 0}A_{m'}(q')A^*_{m'-M'}(q')\exp\{iM'(\varphi_j+\Delta\varphi_l)\}\sum_{k''k'''}\exp(-im'\omega_{k''})\exp(i(m'-M')\omega_{k'''})\exp\{i\mathbf{q}'.(\mathbf{r}_{k''}-\mathbf{r}_{k'''})\}$$

$$C_2(q;q',\Delta\varphi)=\left\langle \begin{array}{l} \sum_{m.M\neq 0}A^*_m(q)A_{m-M}(q)\ldots \\ \sum_{m',M\neq 0}A_{m'}(q')A^*_{m'-M'}(q')\exp\{iM\Delta\varphi_l\}\ldots \\ \sum_{kk'k''k'''}\exp(im\omega_k)\exp(-i(m-M)\omega_{k'})\exp\{-i\mathbf{q}.(\mathbf{r}_k-\mathbf{r}_{k'})\}\ldots \\ \exp(-im'\omega_{k''})\exp(i(m'-M)\omega_{k'''})\exp\{i\mathbf{q}'.(\mathbf{r}_{k''}-\mathbf{r}_{k'''})\} \end{array}\right\rangle_{DP}$$

Note that the average over DPs removes *all dependence on the position of the particles*, despite the fact that coherent scattering was assumed (only the diagonal terms in the sums over the particles survive)

$$C_2(q,q',\Delta\varphi)=N_p\sum_{m,M\neq 0}A_m(q)A^*_{m-M}(q)\sum_{m',M\neq 0}A^*_{m'}(q')A_{m'-M}(q')\exp\{iM\Delta\varphi_l\}$$
$$=N_p\sum_{M\neq 0}I^*_M(q)I_M(q')\exp\{iM\Delta\varphi_l\}$$

The correlations depend only on the *single-particle* quantities $I_M(q)$

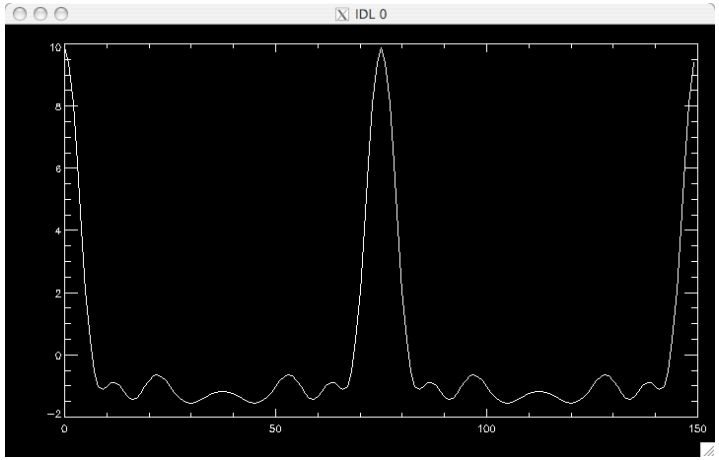
Recover $I_M(q)$ from C_2: Diffraction Pattern of Single Particle from that of Multiple Particles

- **Take FT of $C_2(q,q';\Delta\Phi)$:**

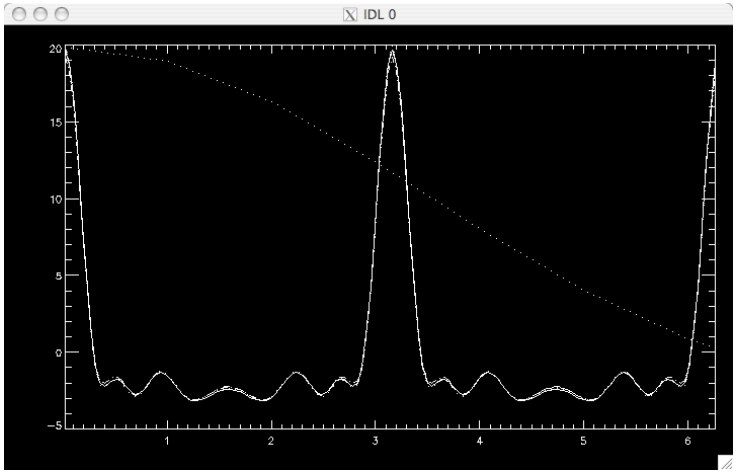
$$\int C_2(q,q';\Delta\phi)\exp(-i\Delta\phi)\,d\Delta\phi = I_M(q)I_M(q') = B_M(q,q')$$

- **Find $I_M(q)$ from $B_M(q,q')$**
- **Can be done**

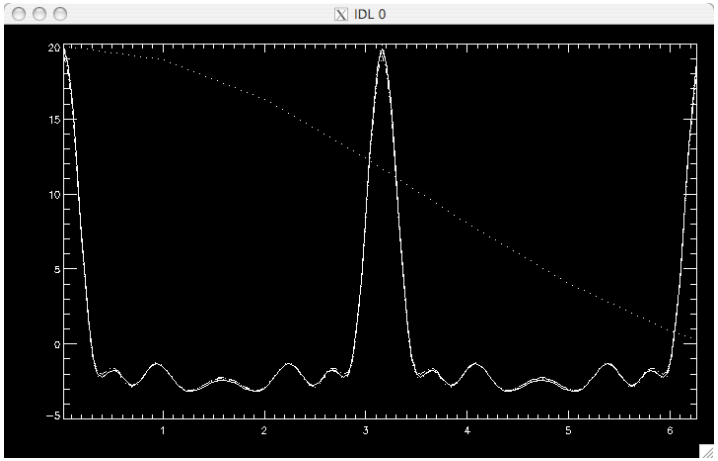
Simulated Pair Correlations



1 particle



2 particles, 1000 DPs

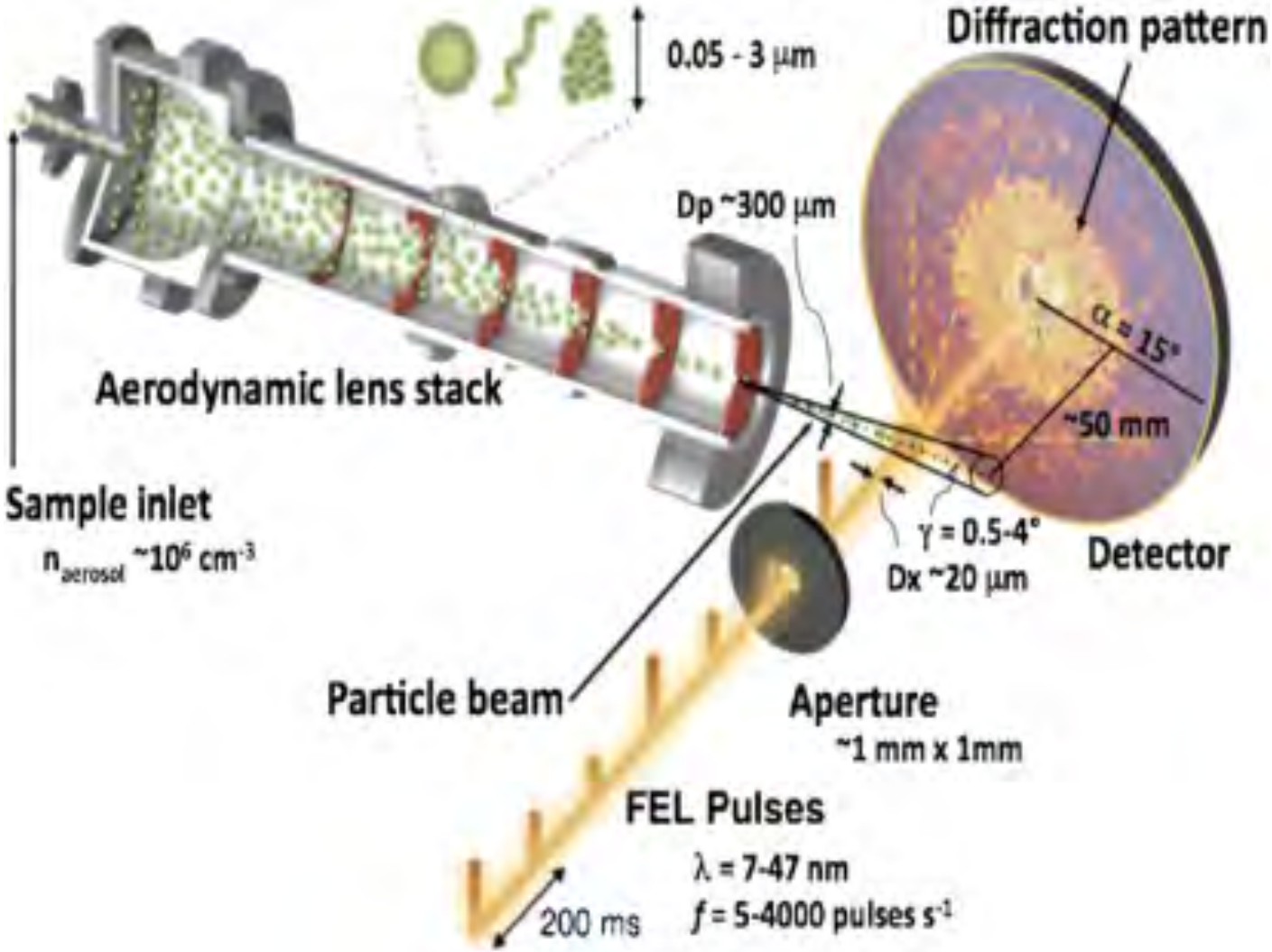


10 particles, 1000 DPs

Science of STC

- An X-Ray Free Electron Laser (XFEL) produces X-rays some 10^{10} times more brilliant that even the most brilliant synchrotron surces
- Rate limiting step of X-ray crystallography is often the ability to crystallize samples
- With some samples, e.g. membrane proteins, crystallization may not be possible at all
- Since a typical crystal used in X-ray crystallography has perhaps 10^{12} molecules, it is speculated that XFEL X-rays may be bright enough for structure determination from an individual molecules
- Molecules provided to the beam in a liquid jet to simulate their aqueous environment in nature, or else as nebulized single molecules

Aerosol Injector for XFEL



Science of STC

- Serious problem is radiation damage
- X-ray produced in pulses some tens of fs in duration
- Concept of “diffraction before destruction”
- Possibly a complete solution to radiation damage problem

Extension to General 3D Rotations

3D intensity in (X,Y,Z) coordinate system

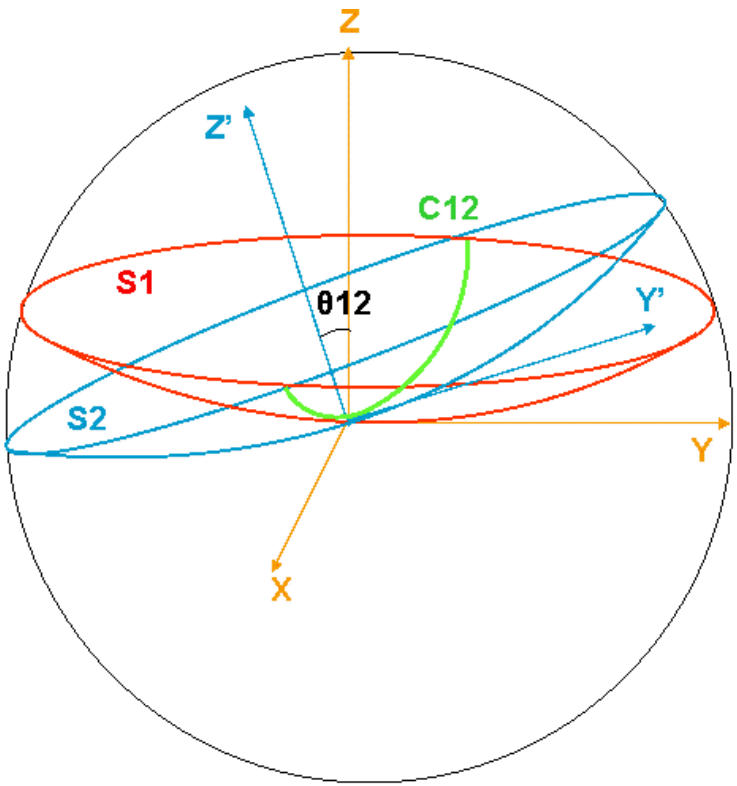
$$I^{(0)}(\mathbf{q}) = \sum_{LM} I_{LM}(q) Y_{LM}(\hat{\mathbf{q}})$$

2D intensity on red Ewald sphere S1

$$I^{(0)}(q,\phi) = \sum_{LM} I_{LM}(q) Y_{LM}\left(\pi/2 - \sin^{-1}(q/2\kappa), \phi\right)$$

2D intensity on blue Ewald sphere S2 rotated by Euler angles (Φ,θ,Ψ)≡α, say

$$I(\alpha; q,\phi) = \sum_{LMM'} D_{LMM'}(\alpha)(q) Y_{lm'}\left(\pi/2 - \sin^{-1}(q/2\kappa), \phi\right)$$



J. Phys.: Condens. Matter 21, 134014 (2009)

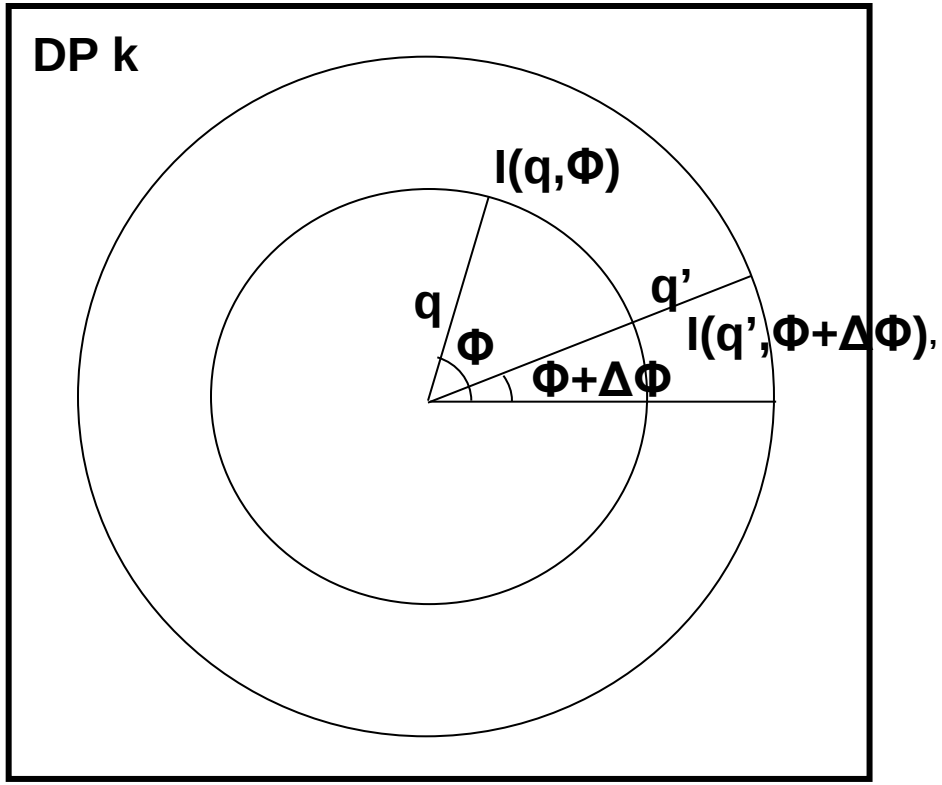
Input to algorithm

Angular correlations

$$C_2(q,q';\Delta\phi)=\left\langle \int I(q,\phi)I(q',\phi+\Delta\phi)d\phi \right\rangle_{DP}$$

between two pixels on each DP p, but summed over all N DPs.

Note neither the magnitudes nor number of such quantities grows with the number of DPs measured. The correlations just becomemore accurate as N increases.



If the scattering angle is 2ζ, q=2κsinζ

Particles Randomly Oriented in 3D

$$C_2(q,q';\Delta\phi)=\frac{1}{N_\alpha}\sum_{\alpha\phi} I(\alpha;q,\phi)I(\alpha;q',\phi+\Delta\phi)$$

$$=\frac{1}{N_\alpha}\sum_{\alpha,\phi}\sum_{LMM'} D^*_{LMM'}(\alpha)I^*_{LM}(q)Y^*_{LM'}[\theta(q),\phi]\sum_{L'M'M'''} D_{L'M''M'''}(\alpha)I_{L'M''}(q')Y_{L'M'''}[\theta'(q'),\phi+\Delta\phi]$$

$$\theta(q)=\frac{\pi}{2}-\sin^{-1}(q/2\kappa) \qquad \theta'(q')=\frac{\pi}{2}-\sin^{-1}(q'/2\kappa)$$

$$\frac{1}{N_\alpha}\sum_{\alpha} D^*_{LMM'}(\alpha)D_{L'M''M'''}(\alpha)=\frac{1}{(2L+1)}\delta_{LL'}\delta_{MM''}\delta_{M'M'''}$$

Great orthogonality theorem of representations of SO(3) group

$$C_2(q,q';\Delta\phi)=\sum_{L\neq 0} P_L(\cos(\Delta\phi))B_L(q,q')$$

$$B_L(q,q')=\sum_M I^*_{LM}(q)I_{LM}(q')$$

This is an orientation-independent quantity characteristic of the "diffraction volume" of an individual particle

Extracting Structure from Correlations

If it were possible to extract $I_{LM}(q)$ from $B_L(q,q')$, possible to construct 3D diffraction volume via

$$I(\mathbf{q}) = \sum_{LM} I_{LM}(q) Y_{LM}(\Omega)$$

However, this is difficult to do since

$$B_L(q,q') = \sum_{M} I^*_{LM}(q) I_{LM}(q')$$

Is a quadratic equation for $I_{LM}(q)$ with an extra summation over m on RHS. However, solutions exist for symmetric objects like Icosahedral or helical viruses

Diffraction Volume from the Measured $B_L(q,q')$'s

- How does one extract the $I_{LM}(q)$ coefficients from the $B_L(q,q')$ ones?
- Caspar and Klug (1962) suggested that regular viruses were probably either icosahedral or helical
- For both these symmetries it is possible to reconstruct the diffraction volume from the measured $B_L(q,q')$'s up to a certain resolution

Reconstructing the Diffraction Volume

Once both the magnitudes and signs of the real coeffs. $g_l(q)$ are determined, one can reconstruct one shell of the diffraction volume. The signs of the $g_L(q')$ of the other shells q' may be determined uniquely from an analysis of the cross pair correlations which yield the quantities

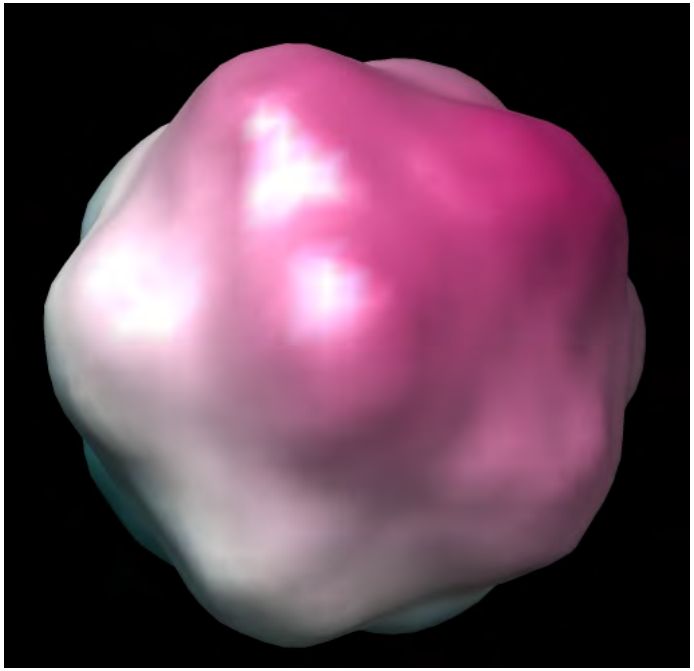
$$B_L(q,q') = g_L(q)g_L(q')$$

Thus, once the $g_l(q)$'s of a particular shell have been determined the $g_l(q')$'s of all other shells may be determined directly. Then the 3D diffraction volume may be reconstructed via

$$I(q,\Omega) = \sum_L g_L(q)\tilde{S}_L(\Omega)$$

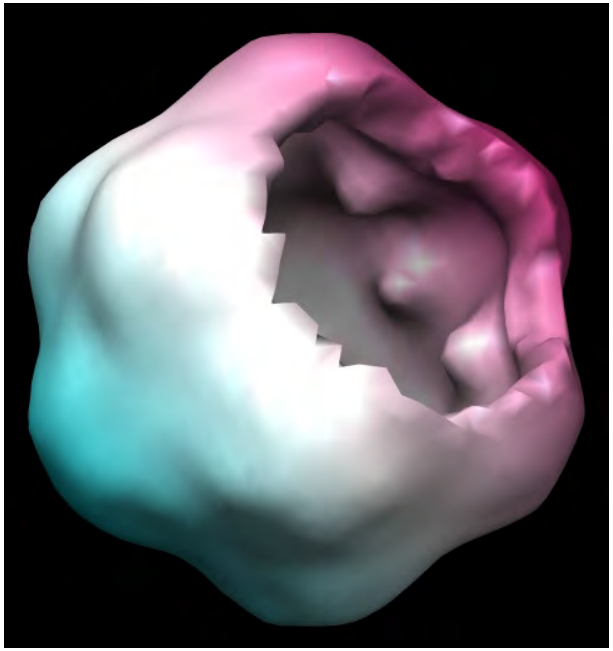
Note that since the icosahedral harmonics (by definition) have icosahedral symmetry, so must the 3D diffraction volume.

Recovery of 3D image of STNV



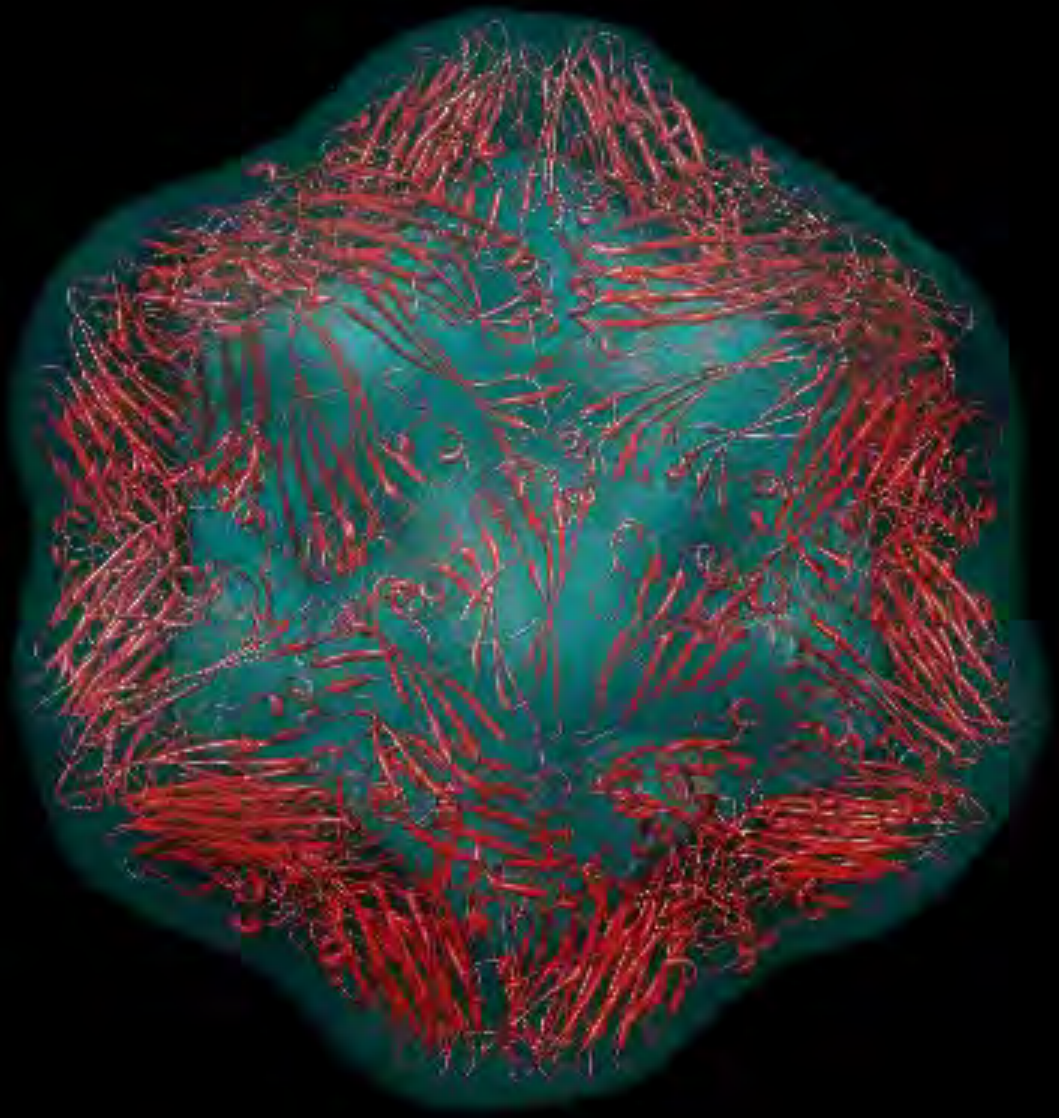
Capsid Viewed Down the 5-Fold Axis

lmax=30
13 Å resl.
for STNV

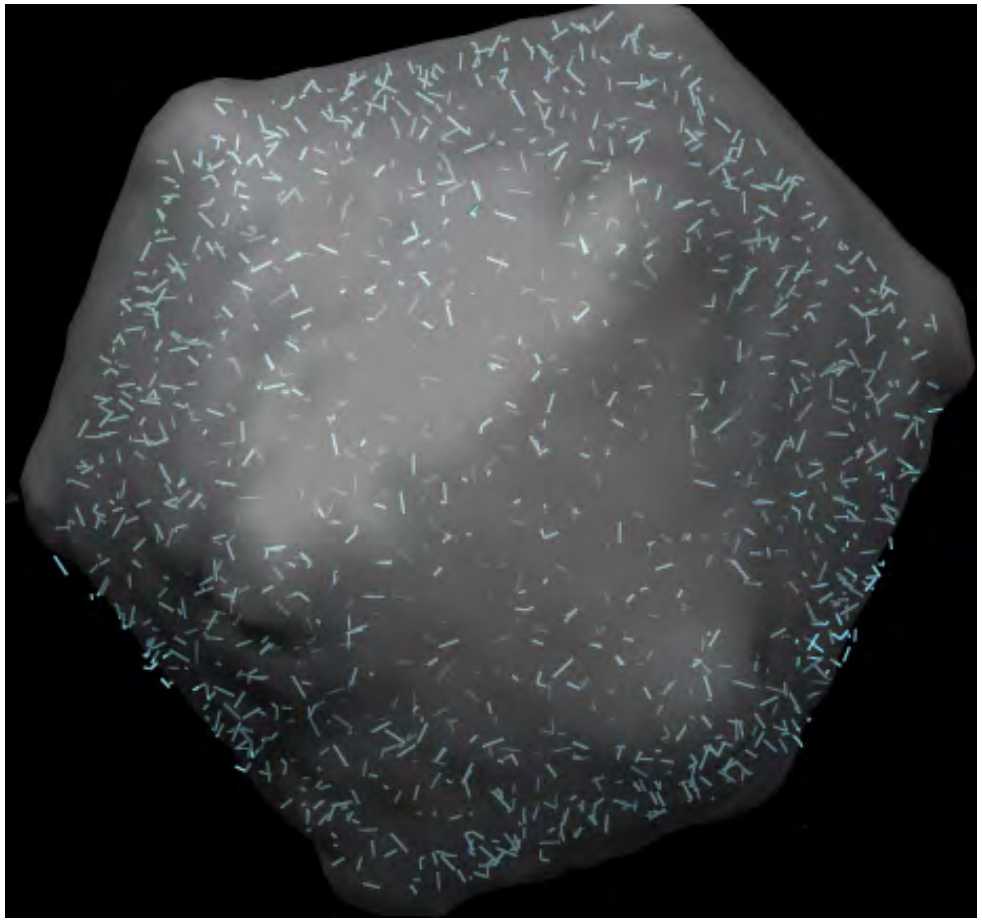


Computational slice showing hollow nature of virus capsid

DKS et al., Optics Express 19, 17318-17335 (2011)

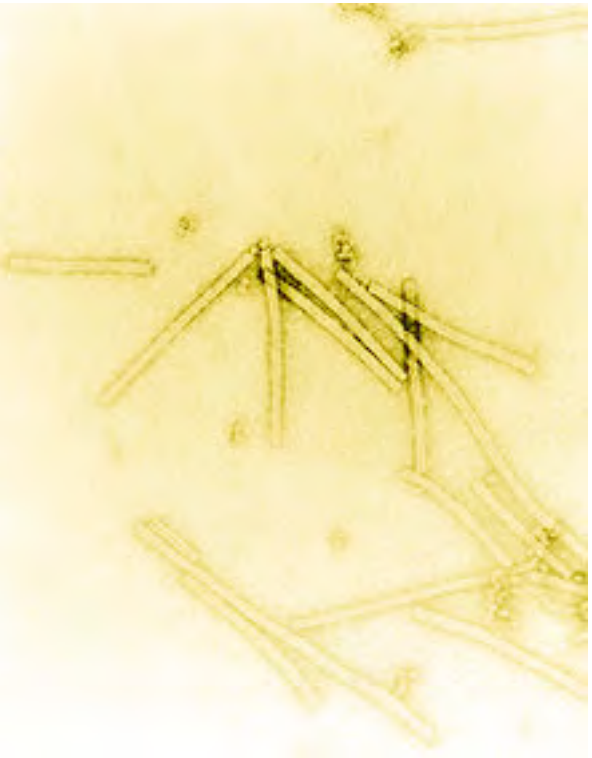


Virus Reconstruction from Chlorella Virus (Simulation)



Paramecium Bursaria Chlorella Virus
Diameter ~ 1900 A

Helical Virus: Tobacco Mosaic Virus (TMV)



First virus discovered. Electron micrograph of TMV particles stained to enhance visibility at 160,000x magnification

Finding the I's from the B's and Alignment by Computational Postprocessing

As before, structural information is contained in

$$B_L(q,q') = \sum_M I^*_{LM}(q) I_{LM}(q')$$

Extracting the $I_{LM}(q)$'s from the $B_L(q,q')$ is possible for icosahedral structures because the ratios of the coeffs. of different m's for the same l's are known. In the case of a helical structure one may exploit the fact that if the z axis is taken as the helix axis, only the m=0 components are non zero up to a resolution of about $d=2\pi/q_{max}$, where $q_{max}=48/100 \cong 0.5$ Å^{-1} (Cochran, Crick, and Vand, 1952). Thus d ≅ 12 Å. Up to this resolution $|I_{L0}(q)| = \sqrt{B_L(q,q)}$ and the signs may be determined by e.g. optimizing the triple correlations. Note that $B_L(q,q)$ is an orientation-independent quantity. The expansion coefficients $I_{LM}(q)$ require the definition of an orientation. Since any orientation is consistent with $B_l(q,q)$, can choose this axis arbitrarily. Choosing a common z-axis for all molecules is equivalent to alignment by computational postprocessing.

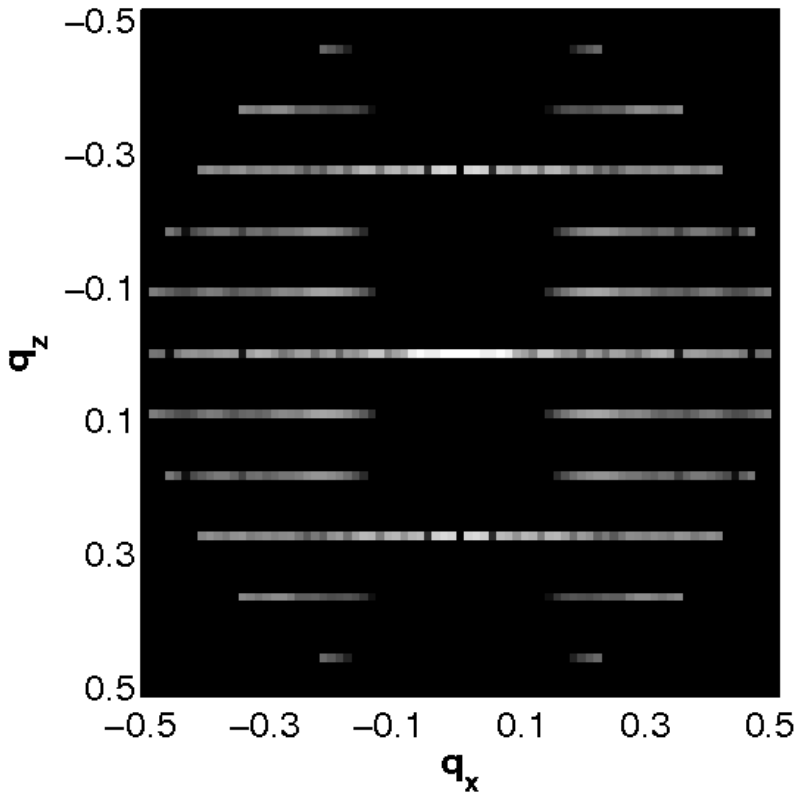
Simulations

We have tested this idea by constructing the $B_L(q,q)$ and $T_L(q,q)$ coefficients using known spherical harmonic expansion coefficients $I_{LM}(q)$ of the 3D diffraction volume of TMV. We then assumed only $B_L(q,q)$ and $T_L(q,q)$ coefficients are known (these are the quantities that may be recovered directly from many diffraction patterns of randomly oriented particles, as measurable in a "diffract-and-destroy" type experiment with an XFEL) and recovered the $I_{L0}(q)$ coefficients from them. We then reconstructed the diffraction volume via

$$I(\mathbf{q}) = \sum_L I_{L0}(q) Y_{L0}(\Omega)$$

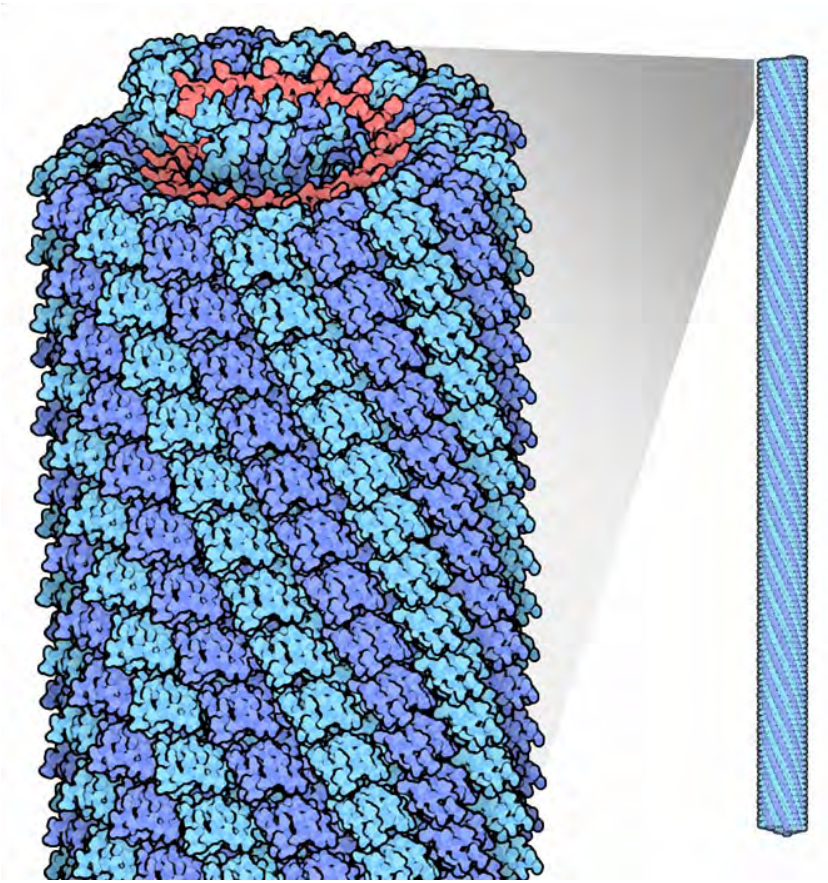
and the real-space structure via an iterative phasing algorithm. The result is shown next.

Reconstructed Helix Diffraction Pattern



Layer Lines

Reconstructed Image of TMV



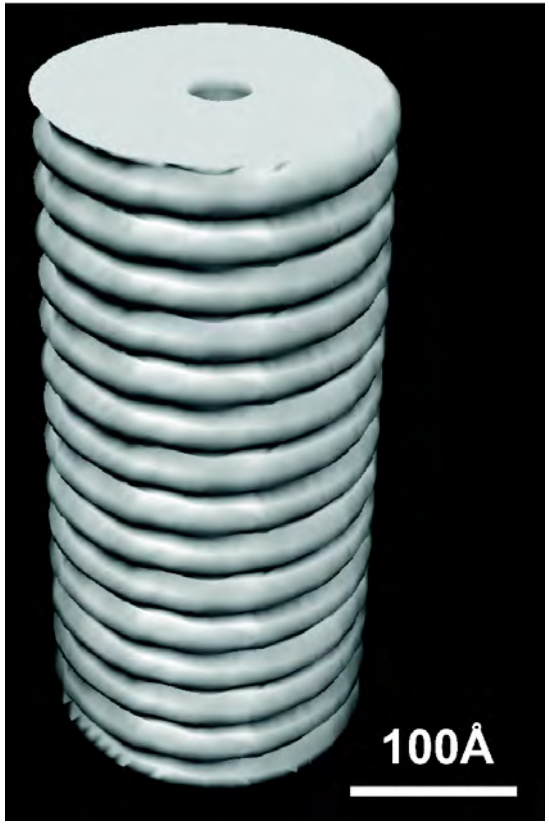


Image from reconstructed diffraction volume. Shows the helical grooves of tobacco mosaic virus (TMV) at low resolution, since m=0 (Cochran, Crick, & Vand, 1952).

Alignment by Postprocessing, “Fiber Diffraction Without Fibers”

Not necessary to perform alignment experimentally via a powerful electric field or flow alignment etc. Can be done by postprocessing DPs of randomly oriented particles. A fiber diffraction pattern is just a slice through this 3D diffraction volume parallel to the helix axis. Possible to generate a fiber diffraction pattern by postprocessing diffraction patterns of random orientations, as measured in an XFEL experiment. Note this will completely eliminate the need to perform “single-axis alignment” experimentally.

If a fiber diffraction pattern may be found by these means, possible to do “fiber diffraction without fibers”!
[H.-C. Poon, P. Schwander, M. Uddin, and D. K. Saldin, Phys. Rev. Lett. 110, 265505 (2013) (cover picture, same issue)]

See also Synopsis, “No need to line up”, M. Rini, Physics, June 27, 2013

Summary of Algorithm

- Method works on scattering data from multiple particles randomly orientated in droplet – better hit rate in experiment, due to scattering many particles, MPC can increase by orders of magnitude
- Works on DPs from multiple particles in random orientations, no need to exclude DPs from multiple particle hits – increases "hit rate"
- From the angular correlations of the data, the diffraction volume $I(q_x,q_y,q_z)$ of an individual molecule in an arbitrary orientation may be found
- This achieves molecular alignment computationally
- Electron density of virus reconstructed from oversampled intensities of a single particle by an iterative phasing algorithm
- Solutions possible for the two most common virus types of regular virus: icosahedral and helical

Poisson Noise and Mean of Correlations

Poisson distribution P(n,λ): probability of measured count of n, mean λ.

$$P(n,\lambda) = \lambda^n e^{-\lambda} / n!$$

Mean count

$$\lambda = \sum_{n=0}^{\infty} nP(n,\lambda)$$

Mean value of product of two variables with individual means λ_1 and λ_2:

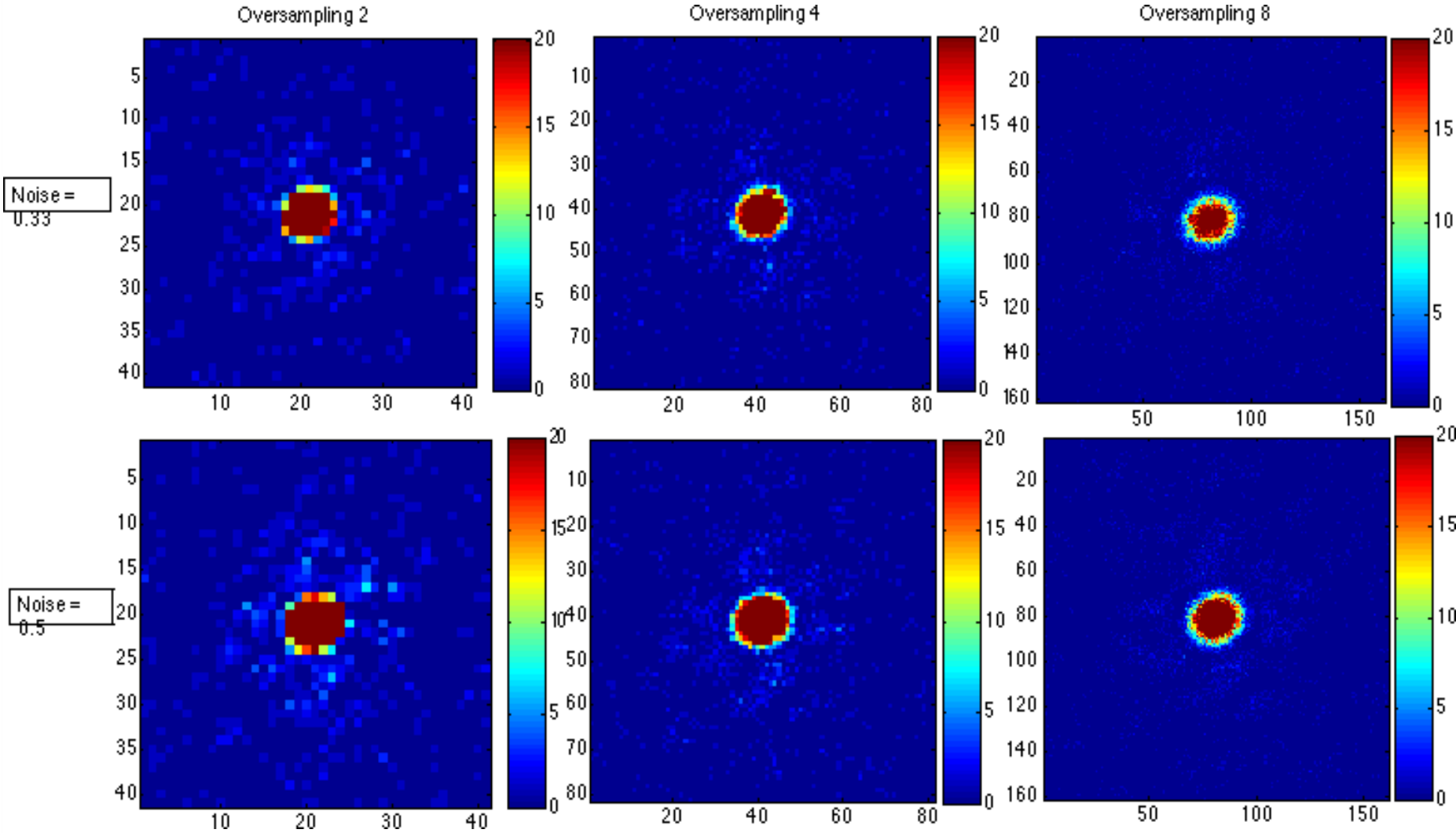
$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} nmP(n,\lambda_1)P(m,\lambda_2)$$

$$= \left\{\sum_{n=0}^{\infty} nP(n,\lambda_1)\right\}\left\{\sum_{m=0}^{\infty} mP(m,\lambda_2)\right\}$$

$$= \lambda_1\lambda_2$$

Mean of products of counts, nm = product of individual means, $\lambda_1\lambda_2$.

Noisy PYP Diffraction Patterns



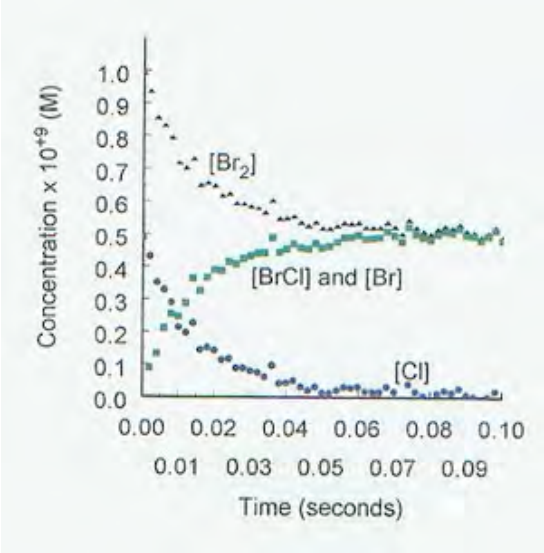
Conclusions

- One key to getting structural information from an XFEL is to realize that nothing moves in a femtosecond, so that even if a sample is destroyed by an XFEL beam diffraction patterns are measured before the destruction “diffract and destroy”
- Note that this is potentially a complete solution to the radiation damage problem
- We can also exploit the fact that the pulse repetition rate of an XFEL is great, so possible to measure perhaps 100 DPs per second
- Can measure perhaps a million DPs during a single experiment
- The question is whether it is possible to analyze such a large number of such weak DPs to retrieve 3D structural information
- This requires great computer power and the development of brand new algorithms that can extract structural information from a large number of very weak diffraction patterns
- We have described one such idea for recovering time-resolved structural changes of uncrystallized biomolecules in solution

Elasped Time: Δt
Distance Traveled by Light: ΔL

Speed of Light

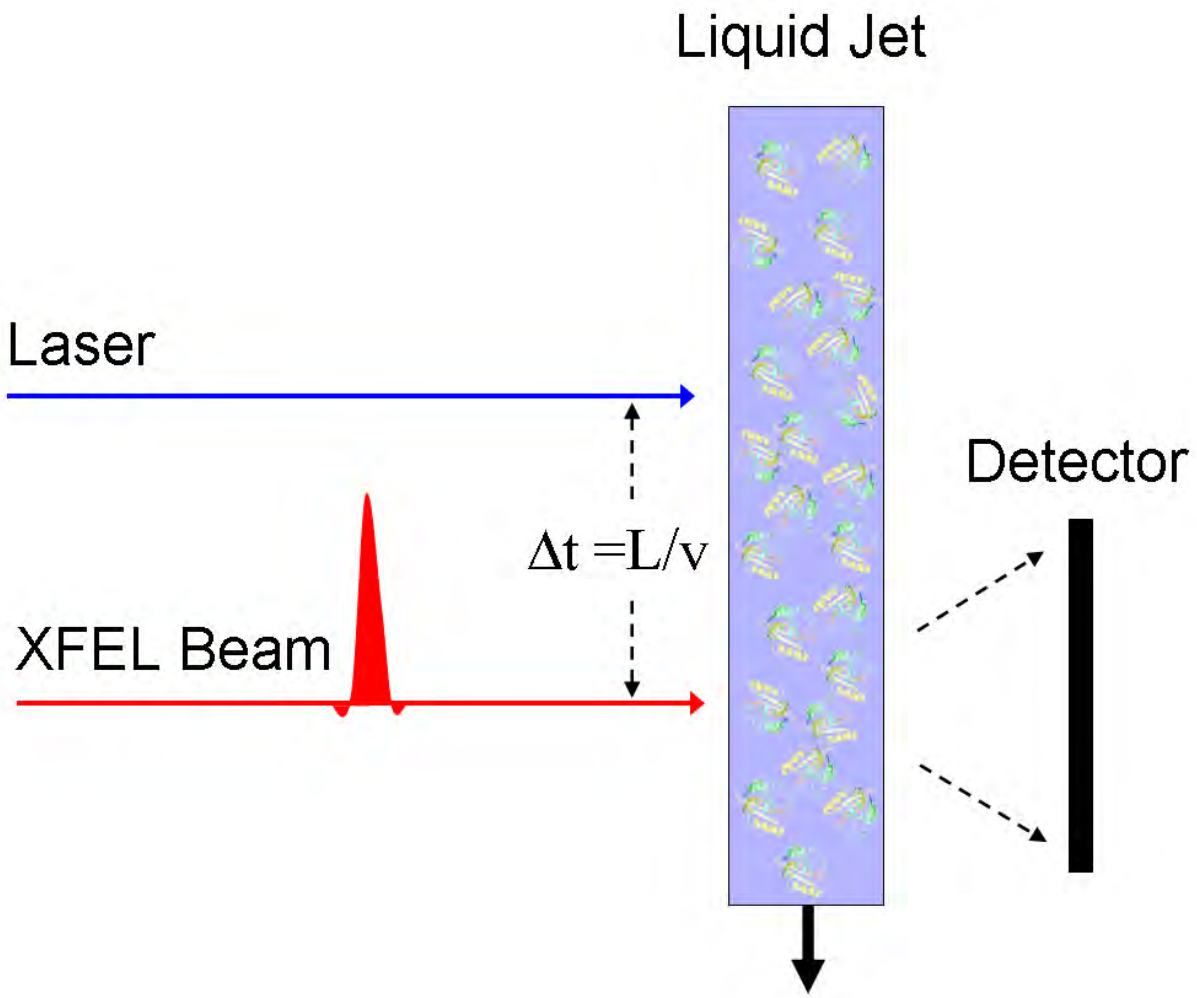
Δt=1 s
ΔL=186,000 mi
~ Distance to the moon.



Δt=1/2000 s
ΔL=100 mi.
~ Distance NYC-PHL

Pump-probe for Time Resolution

v= 10 m/s
Δt=2 ms
ΔL=2 cm



General Theory for Small Structural Changes

$$B_L(q,q') = \sum_M I^*_{LM}(q) I_{LM}(q')$$

$$\delta B_L(q,q') = \sum_M \left\{ \delta I^*_{LM}(q) I_{LM}(q') + I^*_{LM}(q) \delta I_{LM}(q') \right\}$$

$$I(\mathbf{q}) = A^*(\mathbf{q}) A(\mathbf{q})$$

$$\delta I(\mathbf{q}) = \delta A^*(\mathbf{q}) A(\mathbf{q}) + A^*(\mathbf{q}) \delta A(\mathbf{q})$$

$$= \sum_j \delta\rho(\mathbf{r}_j) \left[A^*(\mathbf{q}) \exp(i\mathbf{q}.\mathbf{r}_j) + A(\mathbf{q}) \exp(-i\mathbf{q}.\mathbf{r}_j) \right]$$

Multiply both sides by $Y^*_{LM}(\hat{\mathbf{q}})$ and integrate wrt $\hat{\mathbf{q}}$

Small Structural Changes (Cont.)

$$\delta I_{LM}(q) = \int \delta I(\mathbf{q}) Y_{LM}^{*}(\hat{\mathbf{q}}) d\hat{\mathbf{q}}$$

$$\delta I_{LM}(q) = \sum_{j} \delta\rho(\mathbf{r}_j) C_{LM}(\mathbf{r}_j, q)$$

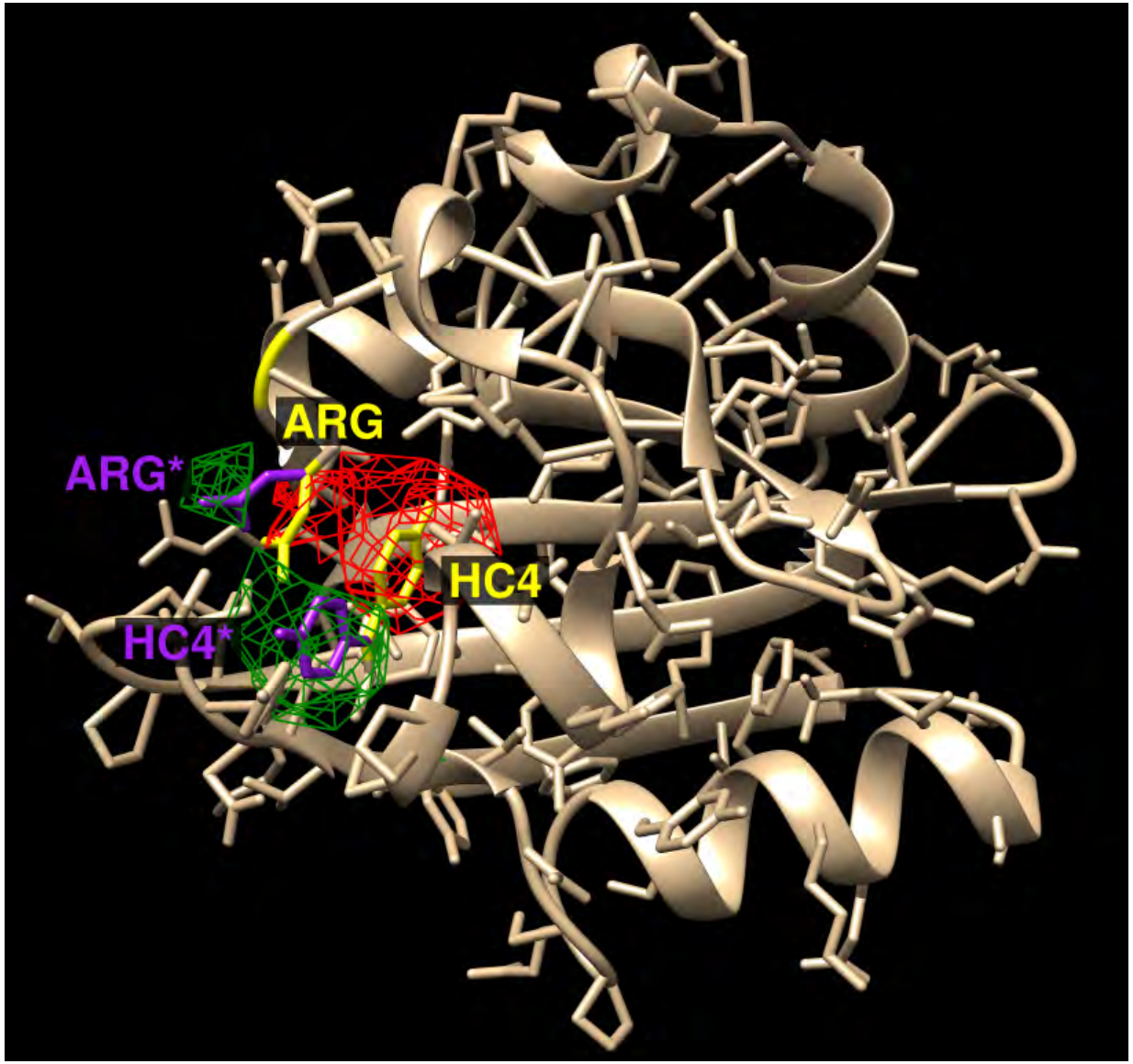
$$C_{LM}(\mathbf{r}_j, q) = \int \left[A(\mathbf{q})^{*} \exp(i\mathbf{q}.\mathbf{r}_j) + c.c. \right] Y_{LM}^{*}(\hat{\mathbf{q}}) d\hat{\mathbf{q}}$$

$$\delta B_L(q, q') = \sum_{j} M_{Lqq';j} \delta\rho(\mathbf{r}_j)$$

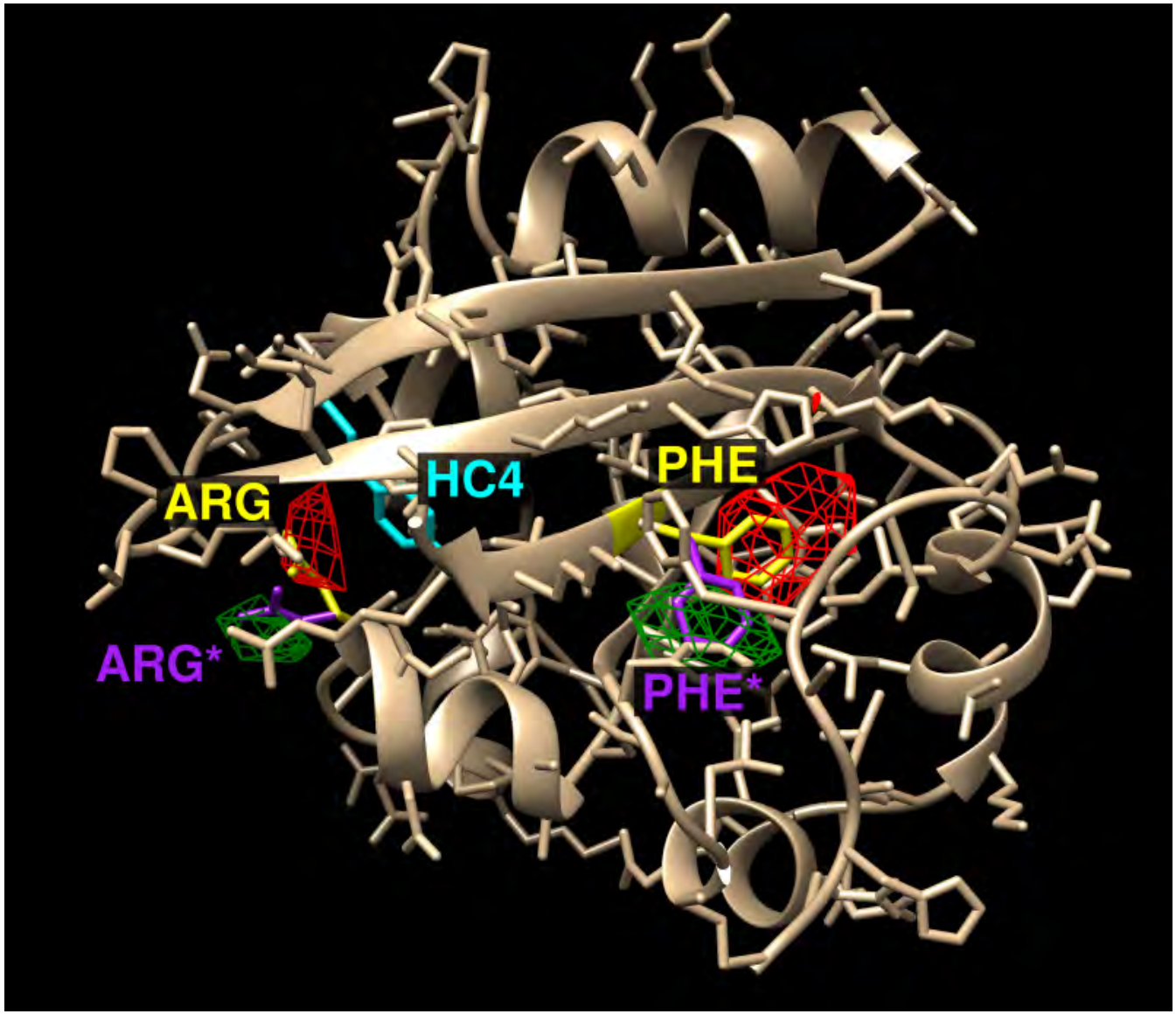
$$M_{Lqq';j} = \sum_{M} \left[C_{LM}^{*}(q, \mathbf{r}_j) I_{LM}(q') + I_{LM}(q) C_{LM}(q', \mathbf{r}_j) \right]$$

$$\delta\rho(\mathbf{r}_j) = \sum_{j} \left[M^{-1} \right]_{j,Lqq'} \delta B_L(q, q')$$

Cis-Trans Isomerization of the Chromophore (with accompanying movement of nearby ARG)



Changes in two Widely Separated Residues (ARG 52 and PHE 121)



Multiple Particles, Coherent Scattering

$$C_2(q,q';\Delta\phi)=\left\langle\left|\sum_j A(q,\phi,\Omega_j)\exp\{iqR_j\cos(\phi-\phi_j\}\right|^2\left|\sum_{j'} A(q',\phi+\Delta\phi,\Omega_{j'})\exp\{iq'R_{j'}\cos(\phi+\Delta\phi-\phi_{j'})\}\right|^2\right\rangle_{DP,\phi}$$

$$=\left\langle\sum_{j,j''} I^{(DP)}(q,\phi,\Omega_j)I^{(DP)}(q',\phi+\Delta\phi,\Omega_{j'})\right\rangle_{DP,\phi}$$

$$=\left\langle\sum_{j} I^{(DP)}(q,\phi,\Omega_j)I^{(DP)}(q',\phi+\Delta\phi,\Omega_{j})\right\rangle_{DP,\phi}+\left\langle\sum_{j\neq j'} I^{(DP)}(q,\phi,\Omega_j)I^{(DP)}(q',\phi+\Delta\phi,\Omega_{j'})\right\rangle_{DP,\phi}$$

Term with uncorrelated orientations tends to a constant. Correlated term is;

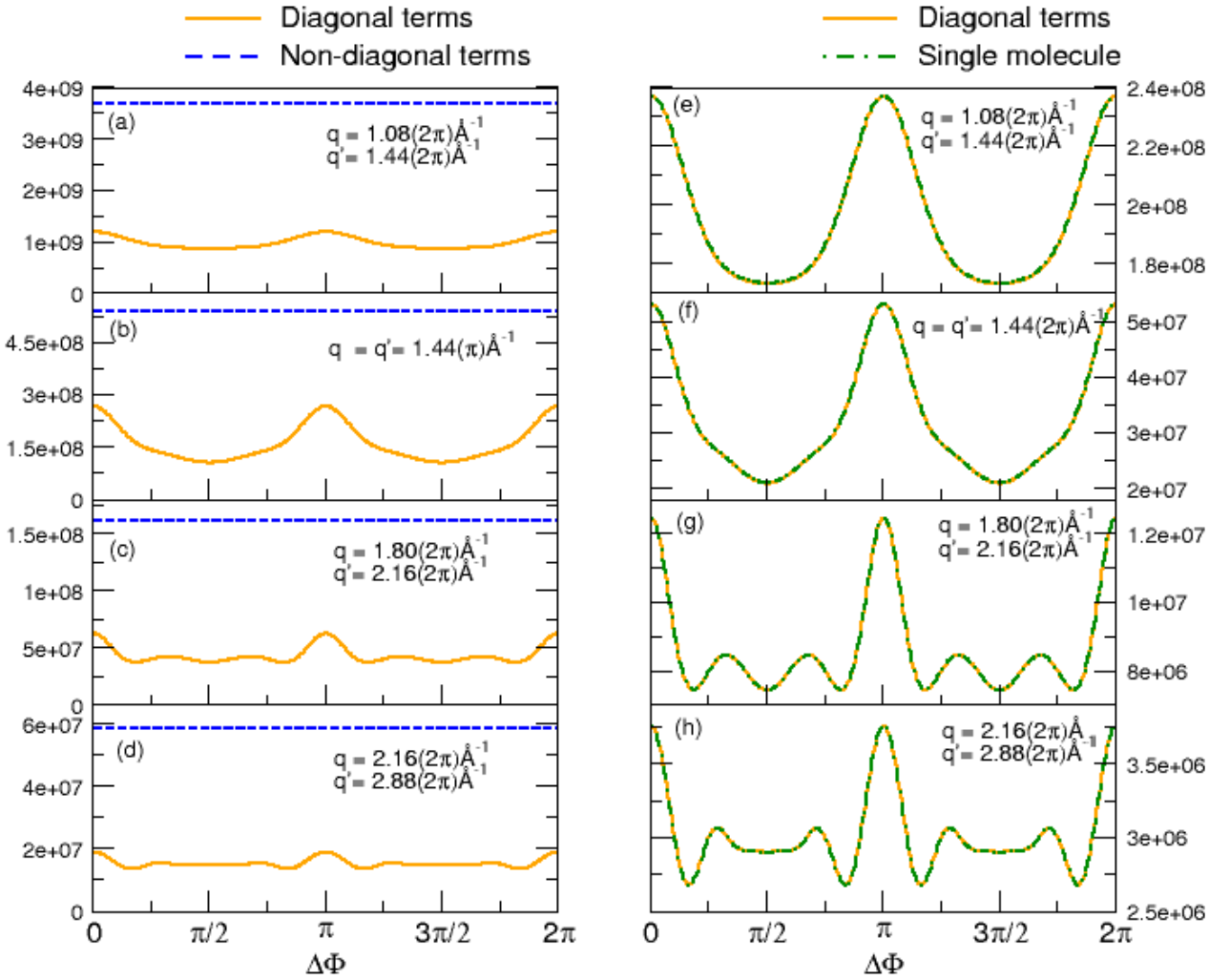
$$\sum_L P_L(\theta'')B_L(q,q')$$

θ" is the angle between $\theta(q)=\frac{\pi}{2}-\sin^{-1}(q/2\kappa)$ **and** $\theta'(q')=\frac{\pi}{2}-\sin^{-1}(q'/2\kappa)$

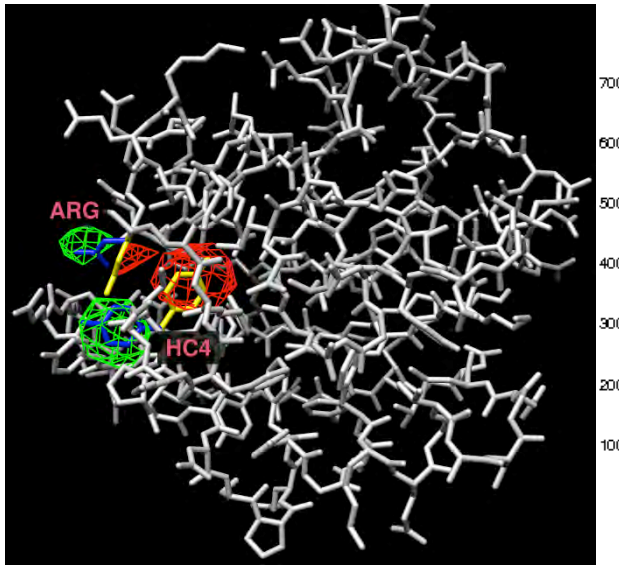
$$B_L(q,q')=\sum_M I_{LM}(q)I_{LM}(q')$$

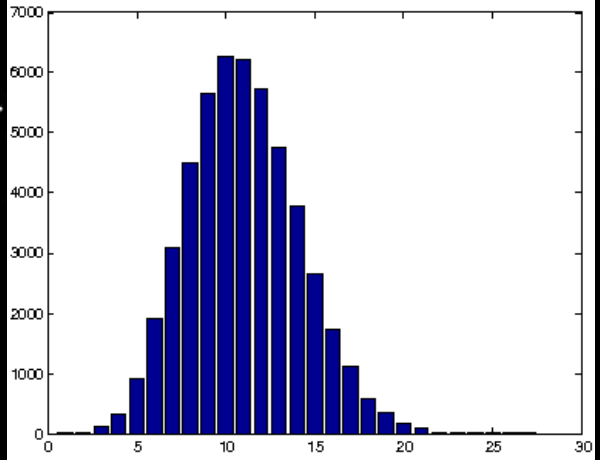
This is an orientation-independent quantity characteristic of the "diffraction volume" of an individual particle

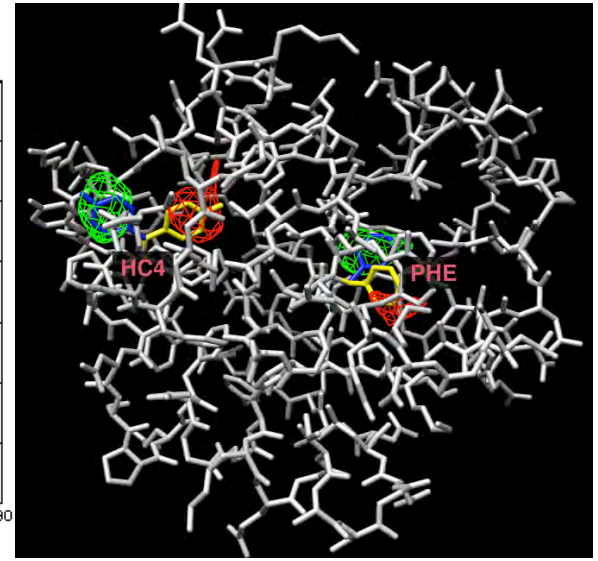
Pair correlations with Multiple Particles vs. Single Particles



Effect of Variation of Number Of Particles Illuminated





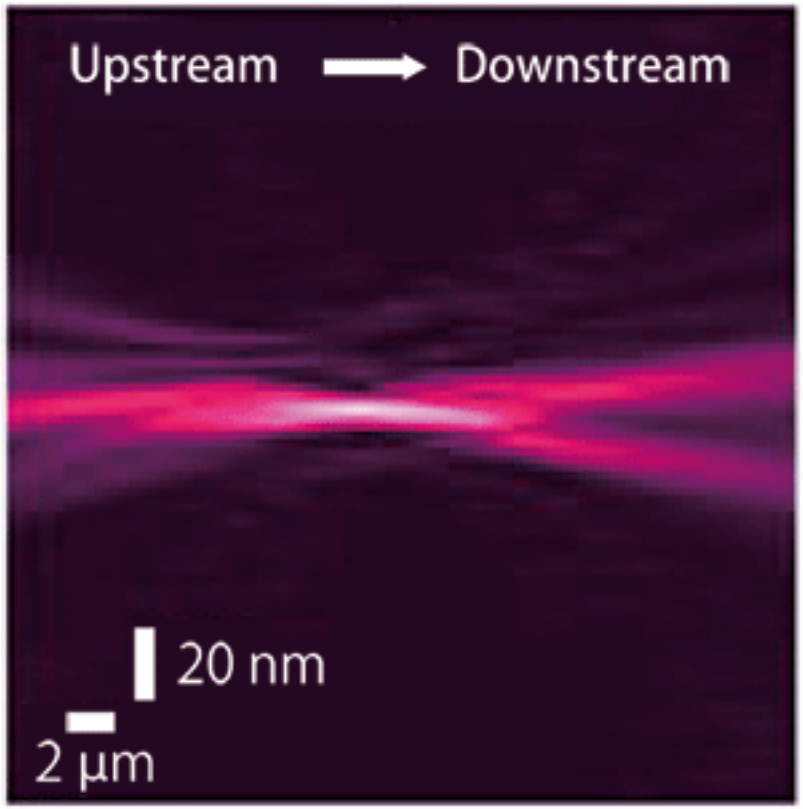


CH4+ARG52

Posson distribution of Particle number

CH4+PHE121

Feasible Beam Cross-Section (Spring-8)



X-ray nanobeam developed at Spring-8 using the X-ray light-collecting Optical system. X-ray beams of 7 nm diameter were generated (Sping-8 web site).

Poisson Noise and Mean of Correlations

Poisson distribution P(n,λ): probability of measured count of n, mean λ.

$$P(n,\lambda) = \lambda^n e^{-\lambda} / n!$$

Mean count

$$\lambda = \sum_{n=0}^{\infty} nP(n,\lambda)$$

Mean value of product of two variables with individual means λ_1 and λ_2:

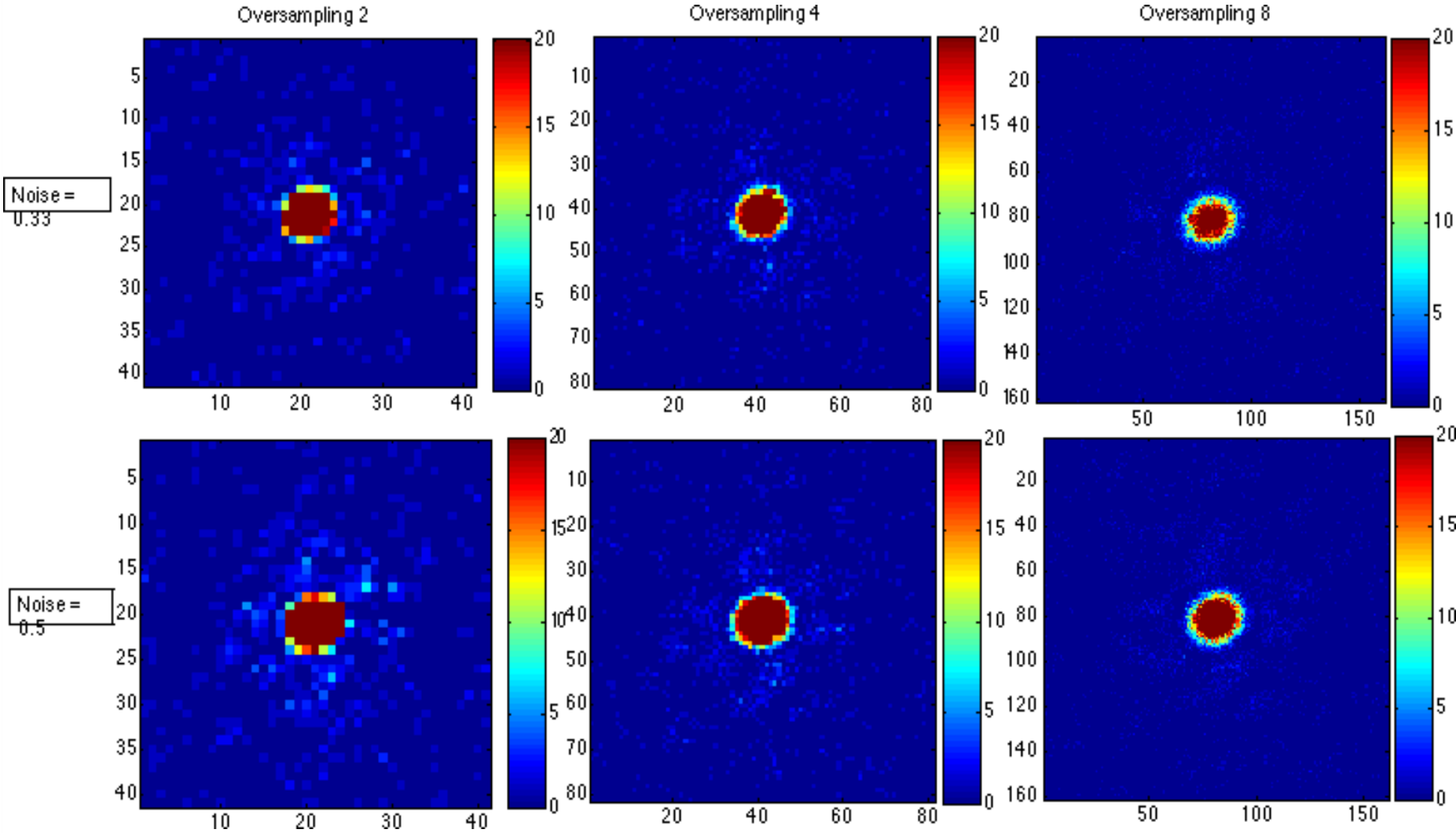
$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} nmP(n,\lambda_1)P(m,\lambda_2)$$

$$= \left\{\sum_{n=0}^{\infty} nP(n,\lambda_1)\right\}\left\{\sum_{m=0}^{\infty} mP(m,\lambda_2)\right\}$$

$$= \lambda_1\lambda_2$$

Mean of products of counts, nm = product of individual means, $\lambda_1\lambda_2$.

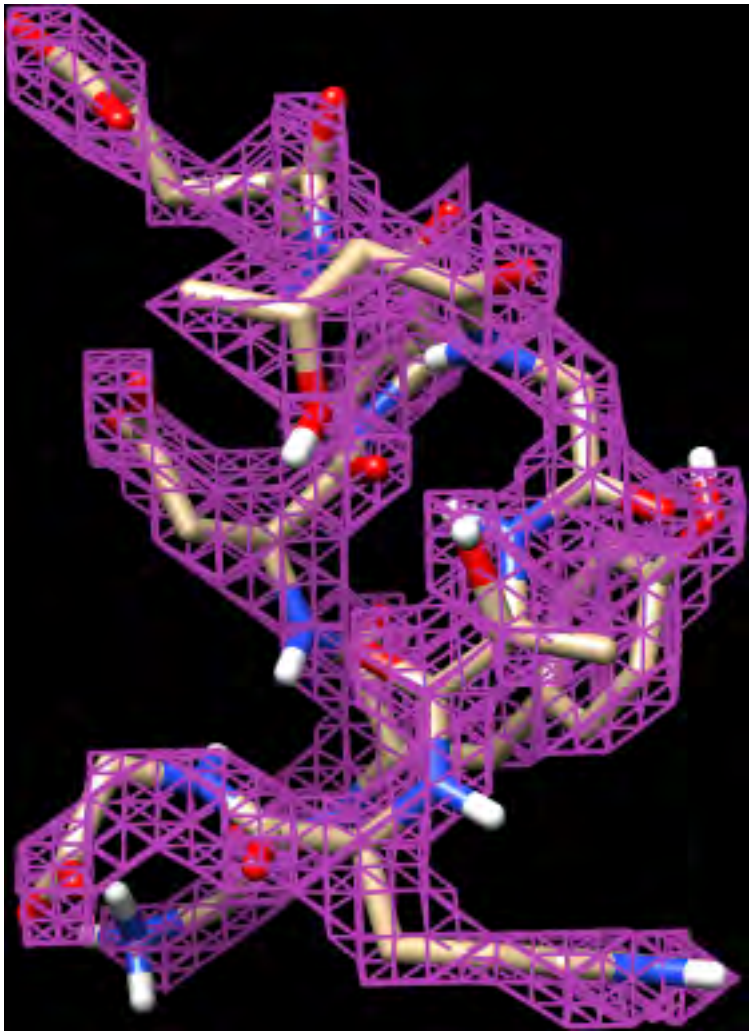
Noisy PYP Diffraction Patterns

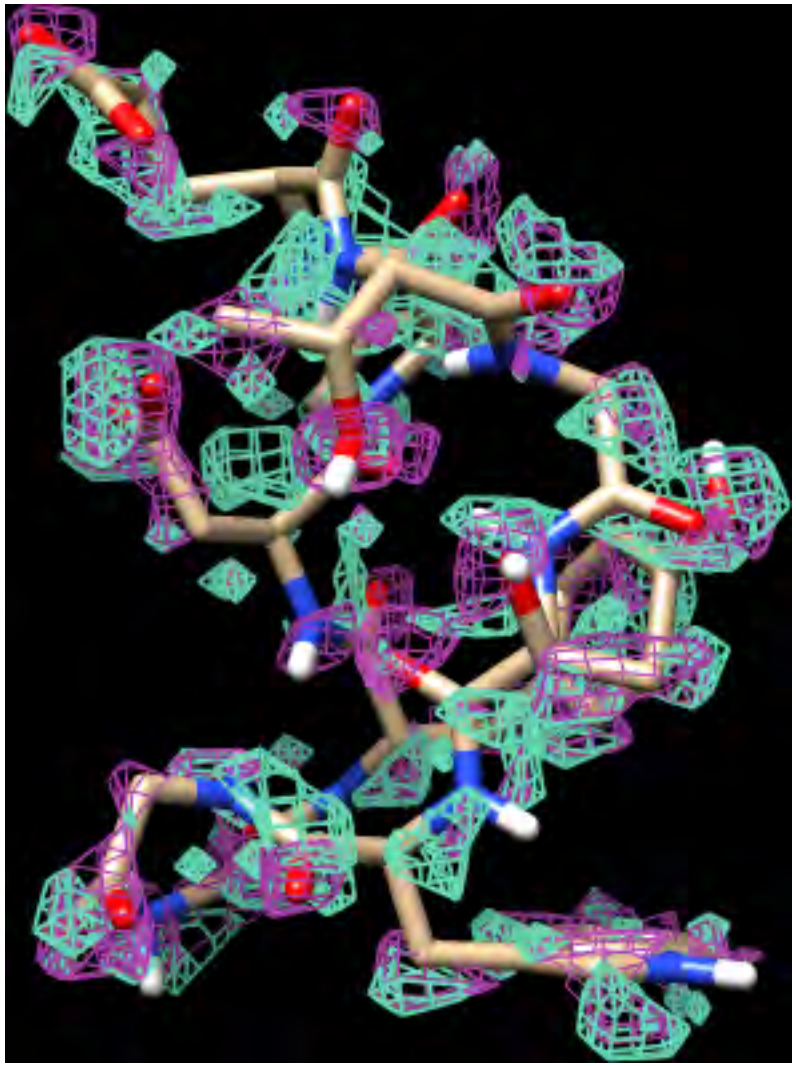


Problems with Obtaining Small Structural Differences from Multiple Particles

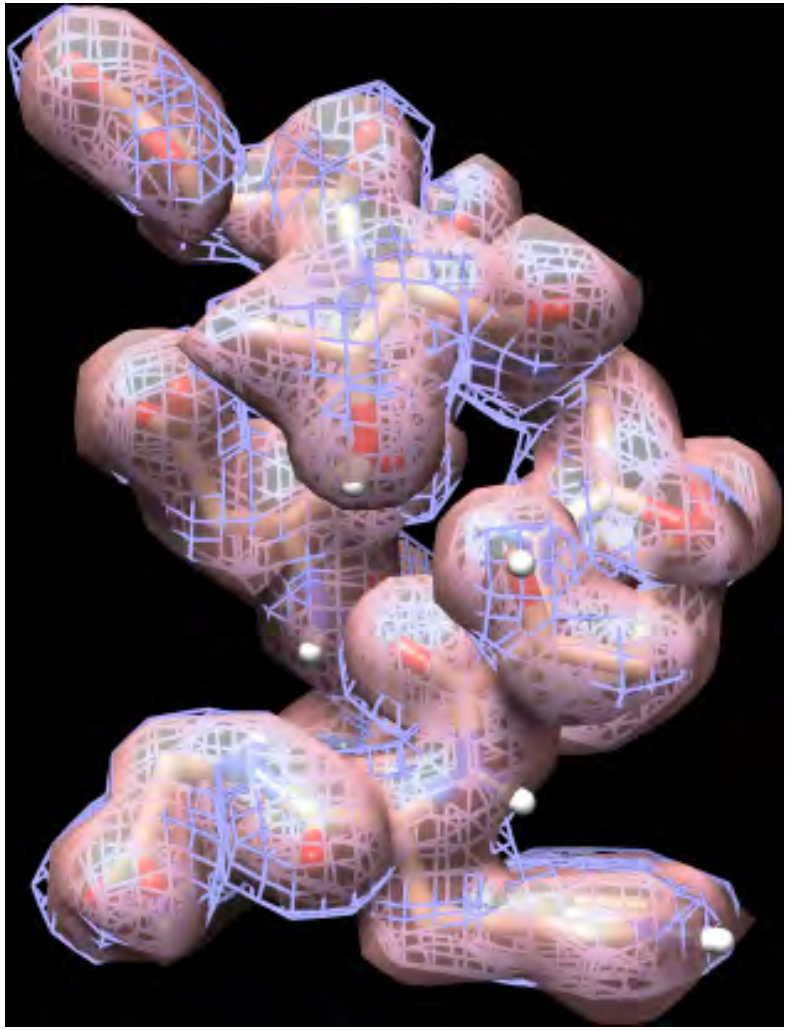
- When number of particles is large, N times as many off-diagonal terms (which do not carry structural info) as particle-diagonal terms (which do).
- Shot noise on these can be as large as large as signal
- Application to single particle overcomes this problem
- Possible to get a 1-0 mask of particle (with SA method (H. Liu et al, Acta Cryst. A 69, 365 (2013), can use this method to improve electron density of entire protein

Mask+Difference Density





Sum Density



Conclusions

- One key to getting structural information from an XFEL is to realize that nothing moves in a femtosecond, so that even if a sample is destroyed by an XFEL beam diffraction patterns are measured before the destruction: “diffract and destroy”
- Note that this is potentially a complete solution to the radiation damage problem
- We can also exploit the fact that the pulse repetition rate of an XFEL is great, so possible to measure perhaps 100 DPs per second
- Can measure perhaps a million DPs during a single experiment
- The question is whether it is possible to analyze such a large number of such weak DPs to retrieve 3D structural information
- This requires great computer power and the development of brand new algorithms that can extract structural information from a large number of very weak diffraction patterns
- We have described one such idea for recovering time-resolved structural changes of uncrystallized biomolecules in solution