Towards a Humpty Dumpty approach for solving the structure of individual nanoparticles

Structure from Fleeting Illumination of Faint Spinning Objects in Flight with Application to Single Molecules,
Authors: Russell Fung, Valentin Shneerson, Dilano K. Saldin, and Abbas Ourmazd,

*arXiv:*0806.2341v1 [physics.bio-ph] (2008)

Recommended with a commentary by Simon J. L. Billinge, Department of Applied Physics and Applied Mathematics, Columbia University, and Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory.

All the kings horses and all the king's men, couldn't put Humpty Dumpty together again (http://en.wikipedia.org/wiki/Humpty_Dumpty) after he fell off the wall and was smashed to pieces. Next generation hard x-ray free electron laser (HXFEL) experiments will take the same approach to solve the structure of individual nanoparticles and proteins. The highlighted paper presents one crucial step in the piecing together of information from smashed nanoparticles so that modern scientists can surpass the limitations of the king's horses and men.

The nanostructure problem is the fact that the powerful traditional crystallographic approaches, which we so take for granted, break down for nano-sized objects. New methods have to be developed to fill this void [1]. One of the most challenging of a wide range of nanostructure problems is to solve the structure of *individual nanoparticles*, quantitatively and in 3D. Right off the bat we bump into the problem that x-ray scattering is very weak from such an object and even modern, powerful, third generation synchrotron x-ray sources are woefully inadequate. The community is eagerly awaiting the incredible intensities from HXFELs, with fluxes of short-wavelength x-rays suitable for structure determination many orders of magnitude higher in peak brightness than current sources. Calculations indicate that the electric fields in the beam are easily enough to blow any nanoparticle apart in a single pulse, so our synchrotron-style experiments of hitting the sample with pulse-after-pulse-after-pulse to build up statistics, and where "beam-damage" refers to a gradual degradation of the sample over minutes or hours, are over. So the bad news is that the beam blows up the sample in a single pulse; what John Spence (ASU) refers to as a "diffract-and-destroy" experiment. The good news is that the inertia of the atoms means the explosion takes some time to develop and the pulse duration is so short that the scattering takes place from the nanoparticle effectively before the explosion allowing us, in principle, to put Humpty together again. That's the good news,

but the bad news is that we are still fighting the Thompson scattering cross-section and detected scattered intensities will still be very low, predicted to be $\sim 10^{-2}$ photons per pulse per pixel from a biological sample of interest, a 500kD protein for example. We are still going to have to do serial experiments, but we have to feed the beam a new, identical, particle for each HXFEL pulse.

One of the miracles of a protein crystal is that 10^{22} or so nanoparticles, the proteins, are all arrayed in an *orientationally ordered* array, so the structure factor of the crystal reveals the internal structure of the nanoparticles. How can this be replicated in a molecular beam of individual nanoparticles so that serial experiments can be summed to increase statistics? One approach is to try and align the molecules in an electric field. The other approach is just not to bother, but to find a way of reorienting the resulting diffraction patterns until they agree with each other, which is the approach taken to reconstruct images in the powerful cryo-electron microscopy method for imaging viruses. The algorithm of choice, since the 1980's, is the "common-line" algorithm [2], but it fails for faint objects with fewer than 10 counts per pixel per shot [3]. Fung *et al.* propose a completely novel and ingenious new algorithm for aligning unoriented diffraction patterns and demonstrate that it works even when the scattering is extremely weak, solving a potentially show-stopping problem in the nanostructure determination of individual nanoparticles from HXFELs

The method is based on concepts from linear factor analysis. Each diffraction "snapshot" is represented as an N_{pixel}-dimensional vector, where N_{pixel} is the number of pixels in the detector, whose components are the intensities in each pixel. Subsequent "snapshots" will yield new vectors. However, since all the snapshots originate from identical objects that are randomly oriented in 3-D space, the tips of the vectors must all lie on a 3-dimensional manifold in the N_{pixel}-dimensional vector space (called a "manifest", or measurement, space). What then remains to be done is to determine the transformation that maps the points on the 3-D manifold in the vector space, back on to a unique orientation in the physical diffraction space. This is an inverse problem, and the approach taken by the authors is to embed a 3-D manifold through the measured vector points using a non-linear factor analytical approach, taken from the data projection and visualization literature, called generative topographic mapping (GTM). The key aspect of this approach is that no averaging is done on individual vectors, but a constrained maximum-likelihood fit is made to the whole collection of measured vectors. This appears to be a better way to use all the information available, even though individual vectors are quite poorly determined, and is more tolerant to the poor statistics of the data than approaches that try and align the diffraction patterns to each other one-by-one. The authors demonstrate that it works in a realistic simulation of an HXFEL experiment of the protein chignolin with 10^3 pixels and 72,000 diffraction patterns of unknown orientation.

Reconstructing a structure from unoriented frames of low statistics data has been demonstrated. If the algorithm proves to be robust against inevitable

aberrations, backgrounds and systematic errors of the real measurement then one of the key unsolved problems on the critical path for nanostructure solution from HXFEL data has been solved. However, almost more interesting is the application to a concrete problem in the experimental physical sciences of the concepts of "manifest" and "latent" spaces, and the powerful tools developed for interpolating between the two from the world of image projection and analysis. Potential applications of this generic approach would seem to extend into other areas of experimental physical sciences.

References

1. S. J. L. Billinge and I. Levin, Science 316, 561 (2007).
2. M. van Heel, Ultramicroscopy 21, 111 (1987).
3. V. Shneerson, A. Ourmazd, and D. Saldin, Acta Crystallogr. A 64, 303 (2008).