

# DIFFRACTION IMAGING OF THE GENERAL PARTICLE

**David Sayre**  
**Department of Physics**  
**State University of New York at Stony Brook**

At most synchrotrons today, something like 10 percent of the photons generated are employed in experiments in which a crystal is illuminated and its far-field coherently scattered x-ray diffraction pattern is measured. From the measurements crystallographers have for years prepared the atomic-, or near-atomic-, resolution 3-dimensional images of matter which have made possible much of 20th century science. X-ray crystallography is thriving today, but it could do much more if it were not required to work with crystals. Crystals call for very large numbers of exact copies of the structure of interest, and for many of the most interesting structures numerous exact copies cannot or do not exist. Amorphous materials, imperfect crystals, and biological cells are examples of this kind. In other cases exact copies may exist, but may be hard to persuade to form crystals. Many biological macromolecules fall in this category.

The main role of the crystal in this type of work is that in the Bragg directions it acts as an extremely powerful amplifier of the coherent scattering signal. (Between the Bragg spots, the crystal greatly diminishes the scattered intensity.) The amplification is of the order of the square of the number of unit cells in the crystal specimen, and can easily amount to a factor of  $10^{12}$  or more. Prior to the synchrotron, it was not practical, without that amplification, to measure the coherent signal, and that is why x-ray crystallography grew up around the crystal.

Work on exploring whether the synchrotron, with its orders of magnitude of greater illuminating power over earlier sources, could take the place of the crystal amplifying power, had its inception in the early 1980s at the Stony Brook physics department and the Brookhaven synchrotron. Principal participants, in addition to Janos Kirz, were WenBing Yun, Henry Chapman, and John Miao. Briefly, by 1990 it was established that the synchrotron can do the job, and by 1995 it was also clear that with the loss of crystallinity, and the gaining of

information between Bragg directions, the phase problem of diffractive imaging is greatly reduced in difficulty. This topic was referred to quite extensively by Ian Robinson and John Spence yesterday. Also by 1995 some progress had been made in quantifying the degree to which the intense illumination, through radiation damage of the specimen, may interfere with high-resolution imaging. Finally, in 1999, John Miao, then a graduate student working with Janos and myself, successfully demonstrated the complete procedure of pattern recording, phasing, and imaging, on a 2-dimensional man-made radiation-resistant specimen. Following this, other groups, especially here in the West of the U.S., began to take up the subject, and a cooperative working arrangement in the subject, largely centered in fact on the Advanced Light Source here at Berkeley, has now recently been brought into existence. Arizona State, Lawrence Berkeley, Lawrence Livermore, SSRL, as well as ourselves in the East, are among the groups entering into this arrangement. Thus really major research strength, in terms of excellence as well as numbers, has been added to the work since the meeting which was held here two years ago, and there seems a real possibility now that imaging by the phasing of diffraction patterns may in the fairly near future come into practical use in the imaging of small general specimens as well as the imaging of crystal unit cells.

I brought up above the existence of two categories of structures which call for treatment non-crystallographically. In one case all the 3-dimensional work, in which the structure must be presented to the photons in many orientations, must be done with a single copy of the structure. In the other case, where although there is no crystal there are numerous exact structural copies, the work can be divided over many copies, these copies being presented to the photons in say random orientations, each copy retaining its structural integrity at least long enough in the beam to yield a good diffraction pattern. The latter case has recently been studied in detail by Janos Hajdu and his group in Uppsala, and calls for extremely intense, extremely short-duration x-ray pulses, and that case today will receive a fresh discussion by Stefan Hau-Riege in the fourth talk of the session. The one-copy-only case is obviously less favorable, but is likely to be the case faced by those wishing to use the biological cell as specimen, and this is the subject of the important third paper of Malcolm Howells et al. in the session. I would like to note, happily, that although Malcolm uses a much more sophisticated approach to the problem of resolution and dose than that

which Henry Chapman and I used in 1995, his resulting estimates of achievable resolutions are actually somewhat more optimistic than ours, which is good. But for now the overall message to the designers is that in the need for future sources there can be a need for both flash and quasi-continuous sources. Something similar I think can be said concerning wavelength. Flash-source users, who can get to atomic resolution, will undoubtedly want Angstrom wavelength x-rays, while quasi-continuous source users, who will often not be able to work at or near atomic resolution, will want to supplement their imaging with absorption edge information, and thus will want also x-rays running down to below the carbon edge, say to 250 eV.

I spoke of the transition toward practical useability, and would like now to talk about the project of imaging the yeast cell which we are now particularly engaged in at StonyBrook/Brookhaven. Our central team consists of 5 persons, Janos Kirz and Chris Jacobsen (team leaders), David Shapiro and Enju Lima (graduate students), and myself. Our first goal is to image the cell in the freeze-dried state using photons at about 1200 eV energy. The freeze-dried state will obviate the need for cryo technique (although that capability is present in our apparatus), and the energy chosen will cause the principal constituents (C,N,O) to act nearly as real-valued scatterers, reducing absorption effects and allowing us, at least in our initial imaging work, to treat the specimen as only mildly complex-valued. On the phasing front we don't yet have actual 3D experimental data to work with, but Enju has successfully run 3-dimensional problems on large simulated data sets which do include data error (though not yet data incompleteness), and which are not much smaller than the actual ones we expect to be dealing with in the future. Enju's phasing software is based on Elser's difference-map iterative technique, about which I will say a little more later. Also, in the experiment it will be necessary to convert measurements made in an Ewald sphere geometry to data on a rectangular lattice geometry, and preliminary software for doing that has been written by Janos. Meanwhile the hardware, a year in design, has been built and most but not all of its design capabilities successfully tested. It is dangerous to predict, but there seems no particular reason why we could not have a data set at some point early in 2004 and a first image not much later. After that we would want to start on a biologist-driven program of imaging cells in various stages of their life-cycle, various environmental situations (including frozen-hydrated), and at various

x-ray wavelengths; in short, starting to make an assessment of the usefulness of diffractive imaging to the cell biologist. Thus in biology, part of the promise of single-particle diffractive imaging may lie in the study of large biomolecules as discussed in papers 2 and 4, while a second valuable part could lie in what it may be able to do in the study of the whole biological cell.

Let me offer a few thoughts on other topics.

First, with regard to gathering a data set, there is the desirability of eliminating or reducing certain dataset absences which we currently have. In both the multi-copy and one-copy-only cases this could be helped by manufacturers supplying detectors with a missing central pixel; this would assist in reducing data lost in a beam stop. In the one-copy-only case data loss would be further lessened if the specimen could be held on a carbon nanofiber instead of lying on an EM grid which interferes with edge-on presentation of the specimen to the beam; a second goniometer arc is also desirable to avoid the missing cone which is generated when only a single axis of rotation is available. In the absence of these experimental refinements the phasing algorithm can to a certain degree be asked to fill in missing amplitudes as well as phases, and some experimentation has been done along these lines, but there is some loss of reconstruction quality, and we are not yet entirely sure of how things in that area will work out in practice.

A related matter concerns deciding on the dosage per exposure to be delivered to the specimen. This is most noticeable in the one-copy-only situation, where a careful design of the experiment as a whole will need to be done, taking into account the best set of orientations to employ, plus the total exposure tolerance of the specimen, and thus the exposure per orientation; also, automation of the actual orientation settings and the exposures will be needed. We plan to precede the actual experiment with an experiment which we call the exposure series experiment, in which a specimen like the specimen to be used will -- in the simplest form -- be kept at a fixed orientation and pattern recorded after say .1 second, 1 second, 10 seconds, 100 seconds, ... up to definite specimen deterioration in the beam. From those patterns best total exposure and best achievable imaging resolution will be estimated, and a plan involving total exposure and needed sampling fineness and number of orientations can be drawn up. Exposure series data will also resolve the

mildly ongoing debate about what power-law relates exposure and resolution. Finally, the exposure series will be a simple and low-cost way of evaluating different cell types, different preparation modes (frozen-hydrated, freeze-dried, critical-point dried, etc.), different operating wavelengths, etc. To supplement this, during the actual data taking, a scheme of periodic re-visiting of standard points in diffraction space, together with an on-the-fly statistical analysis to check on specimen condition, is being developed by Janos. Somewhat different considerations will later apply in the multi-copy case, largely having to do with how many photons the pulsed-source hardware designers can squeeze into how many femtoseconds, etc., and I think we will hear about that in Stefan's paper.

Those are some of the gathering-of-diffraction-data considerations. Another consideration is that of obtaining a good support envelope of the specimen. I have not thought too much about this question, though I think it is an important one, but Chris Jacobsen has included a small zone-plate microscope in our apparatus, with which a good enough image of the specimen in each of its successive orientations can be obtained, to allow getting a reasonably good envelope with which to start the phasing algorithm, after which the algorithm itself will be able to supply further information.

This brings me to the phasing algorithm itself. Here I think we are pretty close now to being sure of success even with very large 3-dimensional problems, provided that the situation is one of ideal or near-ideal data quality and good envelope quality. We have the computing power, and it becomes continually clearer that in the Fienup, and the newer Elser, and I think in the Szoke (but I know the least about that) iterated projection settings we also have the algorithmic power under those favorable assumptions. What remains uncertain today is the situation with real-life data and envelope quality. There I think we can say that no one to date who has used real data has as yet had to report a failure case. I should say, however, that our simulated studies on 2D data incompleteness have warned us that our present yeast-cell experimental plans -- which do not yet allow us to present the specimen edge-on to the beam -- may produce data that could bring us close to data-insufficiency problems, in which case we may find ourselves perhaps next year temporarily having to report a phasing failure, and scurrying around having to improve the experiment. But that success will come with

time I do not doubt. Bear in mind that in crystallography it was not so many years ago that 50-atom structures represented the limit of phasability. Today structures containing more than a million atoms have been solved. I should add, incidentally, that in Enju's very large 3D phasings using simulated data, she regularly works successfully with complex-valued -- not just real-valued -- functions, and as long as data absences are not present, the convergence is reliable though somewhat slower.

Let me now emphasize, as I come to the end of this talk, that once the technique is brought into real working order, the demand for photons may rise very rapidly. In crystallography, although things are speeding up, producing a full 3D image can still be a 1-year project, counting the growing of crystals and the often complex set of phasing and refinement steps. But without the necessity for crystal growth, single-particle imaging could conceivably come down to a 3-day technique (a day for exposure series and experiment planning, a day for data collection, and a day for phasing and imaging). That, combined with the high photon density in the beam, could rapidly lead to a greatly elevated consumption of photons.

Thank you for your attention.